Statistica Sinica **11**(2001), 961-979

CROSS-RELATED STRUCTURAL TIME SERIES MODELS

Wensheng Guo and Morton B. Brown

University of Pennsylvania and University of Michigan

Abstract: This paper introduces a new class of bivariate time series models, Cross-Related Structural Models (CRSMs). In this class of models, each time series is modeled by a structural time series model and the structural parameters are modeled as functions of the latent history of the other series. These models preserve certain conditional independence structures over time and can incorporate all the features of the univariate state space models. With minimum modifications, existing forecasting and filtering algorithms for the univariate models can be applied to these models. By modeling the cross relationships through the structural parameters, these models allow flexible relationships to be modeled parsimoniously and include parameters with clear interpretations. An application to a bivariate hormone time series with pulses is used as an illustration.

Key words and phrases: Conditionally Gaussian model, multivariate time series model, multiprocess dynamic linear model, pulsatile time series, state space model.

1. Introduction

In many situations one time series may be a precursor to a second series which in turn provides feedback to the first series. Examples exist in many fields such as biology, economics and meteorology. In biological systems, different hormones regulate each other through feedback mechanisms; in economics, sales of similar products may compete with each other for market share; in meteorology, temperature and precipitation may be closely related. By modeling these relationships, we gain a better understanding of the underlying process, obtain more efficient simultaneous estimates and produce more accurate forecasts.

In this article, we introduce a new class of multivariate time series models, Cross-Related Structural Models (CRSMs), which allow flexible relationships between two or more time series. For the purpose of clarity, we focus on bivariate time series models. In this class of models, each time series is modeled by a structural dynamic model and the structural parameters are modeled as functions of the history of the other series. Discussion of the methodological and technical ideas underlying structural time series models can be found in Harvey (1989) and Harvey and Shephard (1992). A Bayesian view of structural models can be found in West and Harrison (1997). The basic idea of structural models

is to use state space models with the state of the system representing the various unobserved components and the parameters (structural parameters) having clear interpretations. The use of a structural time series model allows a better understanding of the system and provides clear interpretations for the parameters. By modeling the feedback mechanisms through the structural parameters, we not only introduce a class of flexible models which can be used to describe many potential relationships, possibly between two time series with totally different model structures, but also preserve the conditional independence structures in the univariate dynamic models, which enable us to adapt existing tools developed in the state space models for estimation and forecasting.

This research is primarily motivated by our interest in modeling the relationship among hormones. Hormones play important roles in the regulation of many biological processes, including the secretion of other hormones. Hormones are known to be secreted in pulsatile patterns (e.g., Weiss, Jameson, Burrin and Crowley (1990)). It is believed that there is a feedback mechanism such that the release of one hormone may trigger the release of the second hormone, and when the second hormone attains sufficient concentration in the blood, the secretion of the first hormone will be turned off. Since these trigger-shutdown relationships usually involve many intermediate hormones and are not necessarily one-to-one, probabilistic relationships instead of deterministic relationships are desired. To illustrate these ideas, we use the gonadotropin releasing hormone (GnRH) and luteinizing hormone (LH) data studied by Midgley et al. (1997), in which blood samples were drawn from the portal system and the jugular vein of each of the six ewes every 5 minutes for 6 hours ($N = 72$), and assayed for the concentrations of GnRH and LH. The primary interest is to study the feedback relationship between GnRH and LH. LH is released from the pituitary in a pulsatile pattern and has a central role in regulating the reproductive cycle. It is known that the release of LH is regulated by GnRH which is also released in a pulsatile pattern by the hypothalamus. It is also believed that LH and other hormones may provide feedback to GnRH in a very complex way. Figure 1 shows the concentrations and fitted values of GnRH and LH from ewe 1. Using a CRSM, we explore the trigger-shutdown relationship between the two series.

In univariate time series, several authors (e.g., Bolstad (1988), Guo, Wang and Brown (1999)) used multiprocess dynamic linear models (MDLMs) (Harrison and Steven (1976)) to model pulsatile hormone time series, where the estimates of the posterior probabilities of the process indicators can be used to identify potential locations of the pulses. In the bivariate settings, we incorporate the MDLMs in the CRSM framework. The trigger-shutdown mechanisms are modeled by regressing the prior probability of being a pulse over the history of the other series. Under this structure, the posterior probability of a pulse is a product

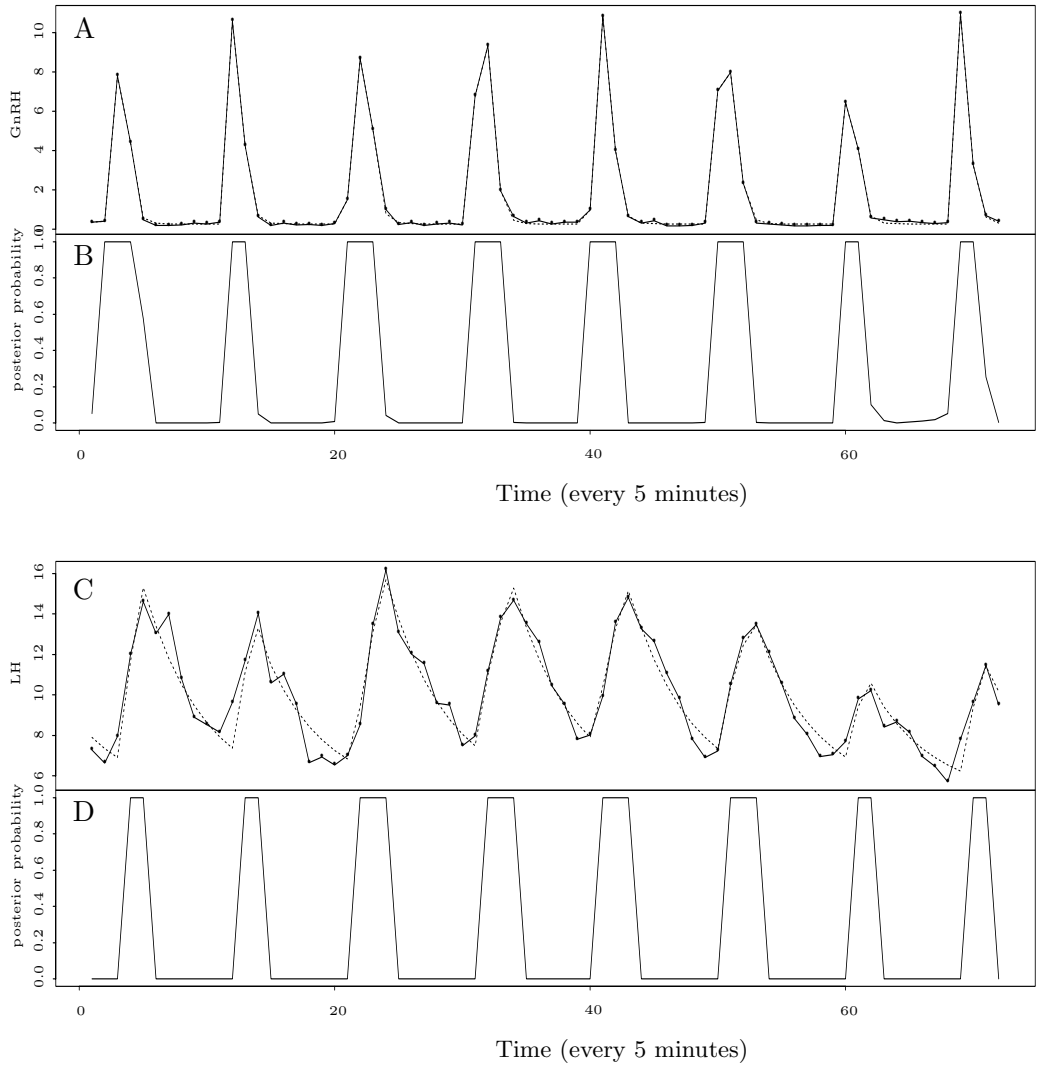

Figure 1. GnRH and LH, and their fitted values using a CRSM. A:observed and fitted values of GnRH. The dots with solid lines are observed concentrations and the dotted lines are the smoothed values. B: posterior probabilities of the pulse indicators. C:observed and fitted values of LH. D: posterior probabilities of the pulse indicators.

of the prior, which is a function of the external series, and the likelihood, which is a function of the internal history. This is a very robust probabilistic relationship because at a certain time point, even the prior probability of being a

pulse given the information from the external series may be very strong, if the internal series does not have any rise in the concentration, which contributes to a small likelihood, the posterior probability of being a pulse may be still very small. In such a situation, a deterministic feedback system may crash, while the probability relationship can still hold and the overall strength of the relationship is measured by the feedback parameter and its confidence interval. The lower panels of Figure 1 show the posterior probabilities of the pulses, which are reflections of the feedback from the other series and the internal jumps. Further details will be explained in Section 6.

One advantage of the CRSMs is that these models are conditionally Gaussian when the underlying univariate models are Gaussian or conditionally Gaussian. The basic idea behind these models is that, conditioned on a vector that consists of the entire series of values of the mixture variables at all time points, the conditional model is a Gaussian state space model and the Kalman filter can be applied. Therefore under the Markov Chain Monte Carlo framework, one only needs to sequentially draw the entire vector of the mixture variable and then apply the Kalman filter within the draw. Efficient simulation techniques for univariate conditionally Gaussian state space models (Carter and Kohn (1994, 1996), Fruhwirth-Schnatter (1994), Shephard (1994) and De Jong and Shephard (1995)) can all be extended to CRSMs. Because of the conditionally Gaussian structure, the approximate method by Harrison and Stevens (1976) can also be extended to these models. The basic idea of the approximation method is to approximate the filtered distribution by a normal distribution with the same first two moments. The approximation method is computationally efficient and is shown to produce accurate estimates in our limited simulation. Further details of the estimation procedures are described in Sections 3 and 4.

This paper is structured as follows. In Section 2, we focus on models in which each time series is characterized by a Gaussian state space model. Estimation procedures are described in Section 3. In Section 4 we generalize our models to the case where the univariate time series is characterized by a conditionally Gaussian model. Up to Section 4 we assume that the parameters are known for the purpose of clarity. In Section 5 we describe some methods to estimate the parameters. In Section 6, an application is used as an illustration. Discussion and remarks are in Section 7.

2. CRSMs with Gaussian Univariate Models

For simplicity we only consider bivariate time series. Our method can be easily extended to multivariate time series. The model is decomposed into three

levels: the observation level, the system level and the cross-regression level. The system level defines the transition of the latent signals and the observation level defines how we observe the latent signals. The first two levels are essentially state space models with their structures set up so that the structural parameters all have clear interpretations. The cross-regression level models the structural parameters of interest as functions of the history of the other series. These functions can be either linear or nonlinear. Because the only connection between the two time series is through the cross-regression, the two structures of the univariate models can be very different. In terms of using the observed history or the latent history in the cross-regression, the models can be classified as Observational History Model (OHM) and Latent History Model (LHM). Since the OHM is straight-forward, we will only focus on the LHM.

The model can be written as

(A) The observation equations:

$$\begin{aligned} y_1(t) &= F_1(t)\boldsymbol{x}_1(t) + v_1(t),\ v_1(t) \sim N(0, \sigma^2_{e1}(t)) \\ y_2(t) &= F_2(t)\boldsymbol{x}_2(t) + v_2(t),\ v_2(t) \sim N(0, \sigma^2_{e2}(t)); \end{aligned} \tag{1}$$

(B) The system level equations:

$$\begin{aligned} \boldsymbol{x}_1(t) &= H_1(t)\boldsymbol{x}_1(t-1) + \boldsymbol{w}_1(t),\ \boldsymbol{w}_1(t) \sim N(\mathbf{0}, W_1(t)) \\ \boldsymbol{x}_2(t) &= H_2(t)\boldsymbol{x}_2(t-1) + \boldsymbol{w}_2(t),\ \boldsymbol{w}_2(t) \sim N(\mathbf{0}, W_2(t)). \end{aligned} \tag{2}$$

The $y_1(t)$ and $y_2(t)$ $(t = 1, 2, \ldots, N)$ are observed time series. The $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ are state vectors, which can include dummy variables to allow more general model structures. The $F_1(t)$, $F_2(t)$, $H_1(t)$ and $H_2(t)$ are observation matrices and transition matrices and are of corresponding dimensions. All the matrices $F_1(t)$, $F_2(t)$, $H_1(t)$, $H_2(t)$, $W_1(t)$ and $W_2(t)$ can contain unknown parameters. These parameters, along with $\sigma^2_{e1}(t)$ and $\sigma^2_{e2}(t)$, are called structural parameters and are denoted as $\theta_1(t)$ and $\theta_2(t)$.

The cross-regressions model $\theta_1(t)$ and $\theta_2(t)$ as functions of the latent history. We denote the latent signals as $z_1(t) = F_1(t)\boldsymbol{x}_1(t)$ and $z_2(t) = F_2(t)\boldsymbol{x}_2(t)$. Since we are usually interested in a lagged relationship and the cross-regression usually only involves a subset of the structural parameters, we assume, without loss of generality, that $\theta_1(t) = \{\gamma_1(t), \phi_1(t)\}$ and $\theta_2(t) = \{\gamma_2(t), \phi_2(t)\}$, where $\gamma_1(t)$ and $\gamma_2(t)$ are scalars and the structural parameters of interest, and $\phi_1(t)$ and $\phi_2(t)$ are the rest of the parameters. Then we have

(C) The cross-regressed equations:

$$\begin{aligned} g_1\{\gamma_1(t)\} &= \alpha_1 + \beta_1 z_2(t-\tau_1) \\ g_2\{\gamma_2(t)\} &= \alpha_2 + \beta_2 z_1(t-\tau_2), \end{aligned} \tag{3}$$

where τ_1 and τ_2 are lags and can take values in $\{1, \ldots, k\}$ with $k < N$; $g_1\{\cdot\}$ and $g_2\{\cdot\}$ are link functions and can be viewed as transformation functions to get rid of the domain restrictions for the structural parameters $\gamma_1(t)$ and $\gamma_2(t)$. The choice of link functions is arbitrary as long as they are one-to-one from the original domain to $(-\infty, +\infty)$.

The number of terms on the right side of the cross-regressions can be increased to allow more general feedback mechanisms. Multiple feedback mechanisms can be incorporated simultaneously, we present it this way for the purpose of simplicity. The new unknown parameters are now $\Theta = \{\phi_1(t), \phi_2(t), \alpha_1, \beta_1, \alpha_2, \beta_2\}$. For the first several observations, the latent history needed in the cross-regressions may not be available. Therefore in order to fully specify the model, the initial values $\{z_1(-\tau_2+1), \ldots, z_1(-1); z_2(-\tau_1+1), \ldots, z_2(-1); \boldsymbol{x}_1(0); \boldsymbol{x}_2(0)\}$ are needed. They can be either included as unknown parameters or given diffuse priors. More generally the lags $\{\tau_1, \tau_2\}$ should also be considered as unknown. For the purpose of simplicity, we treat the parameters as known in discussing the iterative procedures. The estimation of the parameters will be deferred until Section 5.

Let capitals denote the collection of the whole history, i.e., for $i = 1, 2$, $\Gamma_i(N) = \{\gamma_i(1), \ldots, \gamma_i(N)\}$; $Y_i(N) = \{y_i(1), \ldots, y_i(N)\}$; $X_i(N) = \{\boldsymbol{x}_i(1), \ldots, \boldsymbol{x}_i(N)\}$; $Z_i(N) = \{z_i(1), \ldots, z_i(N)\}$.

3. Estimation Procedures

In state space models, estimation and forecasting are characterized by one-step-ahead prediction, filtering and smoothing (signal extraction) algorithms. The one-step-ahead prediction can be extended to m-step forecasting. These algorithms can be extended to CRSMs. If we use the observed histories in the cross-regressions, the estimation is straight-forward. The model can be viewed as two separate state space models with time changing parameters that are calculated by the cross-regressions. The algorithms only involve one additional step to the regular recursive algorithms in state space models: calculating the structural parameters as functions of the history of the other series. The basic Kalman filter and smoothing algorithms can be found in many time series books, such as Anderson and Moore (1979). Modifications to the classical smoothing algorithm that improve the computational efficiency are given by De Jong (1989) and Koopman (1993).

When latent histories are used, parallel recursive algorithms do not exist because the latent signals are estimated by their posterior distribution conditioned on the entire series, which requires smoothing. An iterative scheme between the two time series is needed. The difficulty in the iteration is that the estimates of the latent signals are characterized by their posterior distributions and it is desired to incorporate the entire distributions of the latent signals in the cross-regressions. In the Markov Chain Monte Carlo (MCMC) framework, the iterative scheme naturally leads to an efficient simulation based procedure, the Block Gibbs Sampler, which exploits the advantage of the conditionally Gaussian structure. Because of the conditionally Gaussian structure, a non-simulation based method is also possible; we extend the Harrison-Stevens approximation which approximates the filtered distribution at each step by a normal distribution with the same first two moments.

3.1. Block Gibbs sampler

The first method we propose is to extend the "Block Gibbs sampler" method first proposed by Carter and Kohn (1994) and Fruhwirth-Schnatter (1994). Some modifications of this "forward-filtering, backward-sampling" algorithm are later given by several other authors (e.g., Carter and Kohn (1996), Shephard (1994) and De Jong and Shephard (1995)). The basic idea is to exploit the conditionally Gaussian structures. That is, conditioned on the series of the external latent signals $Z_1(N)$ and $Z_2(N)$, the conditional model is reduced to two Gaussian state space models with time changing parameters which are functions of $Z_1(N)$ and $Z_2(N)$. Under the MCMC framework, we only need to sequentially draw $Z_1(N)$ and $Z_2(N)$, then use the Kalman filter to generate the states $X_1(N)$ and $X_2(N)$. For simplicity, we only outline the general scheme here. Readers are referred to the papers above for details on how to draw the states using the Kalman filter.

Block Gibbs Sampler:

(1.) Calculate $(\Gamma_2(N)|Y_1(N))$; use the filtering and smoothing methods to draw $X_2^*(N)$ from $(X_2(N)|Y_2(N), \Gamma_2)$; calculate $Z_2^*(N)$ from $X_2^*(N)$.

(2.) Calculate $(\Gamma_1(N)|Z_2^*(N))$; use the filtering and smoothing methods to draw $X_1^*(N)$ from $(X_1(N)|Y_1(N), \Gamma_1)$; calculate $Z_1^*(N)$ from $X_1^*(N)$.

(3.) Calculate $(\Gamma_2(N)|Z_1^*(N))$; use the filtering and smoothing methods to draw $X_2^*(N)$ from $(X_2(N)|Y_2(N), \Gamma_2)$; calculate $Z_2^*(N)$ from $X_2^*(N)$.

(4.) Go to (2.)

Under very weak conditions (Tierney (1994)), the block Gibbs sampling is guaranteed to converge.

3.2. Normal approximation to the filtered distribution

Although the block Gibbs sampler produces an almost exact solution, and is much more computationally efficient than regular MCMC procedures that draw one observation at a time, it still requires advanced computing resources especially when the dimension of the problem is large. In this section, we introduce an efficient $O(N)$ approximate method, which can be viewed as an extension of the Harrison and Stevens' (1976) method. Harrison and Stevens proposed to approximate the filtered distribution at each step by a normal distribution with the same first two moments for multiprocess dynamic linear models (MDLMs). Because of the conditionally Gaussian structure of the CRSMs, this method can be extended to them. The mean and variance of the filtered distribution at time t can be calculated using

$$E(\boldsymbol{x}_1(t)|Y_1(t)) = E(E[\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)]) \tag{4}$$

$$V(\boldsymbol{x}_1(t)|Y_1(t)) = V(E[\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)]) + E(V[\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)]). \tag{5}$$

The external signal $z_2(t-\tau_1)$ serves as the mixture variable. Because we approximate the filtered distribution by the marginal distribution at each step of the Kalman filter, $(\boldsymbol{x}_1(t)|Y_1(t))$ implicitly depends on $\{z_2(t-\tau_1-1), \ldots, z_2(-\tau_1)\}$ through $(\boldsymbol{x}_1(t-1)|Y_1(t-1))$, (*i.e.,* conditional on $(\boldsymbol{x}_1(t-1)|Y_1(t-1))$, $(\boldsymbol{x}_1(t)|Y_1(t))$ is independent of $\{z_2(t-\tau_1-1), \ldots, z_2(-\tau_1)\}$. The formulas for $\boldsymbol{x}_2(t)$ are similar. The smoothing is then run backward based on the normal approximations from the forward filtered results.

The original Harrison-Stevens approximation was proposed for MDLMs where the mixture variable is discrete. The computation of (4) and (5) in MDLMs is straight-forward because we only need to calculate the conditional means and conditional variances for all the possible values of the mixture variable and then combine them. (See Harrison and Stevens (1976) or Bolstad (1988) for details). When the mixture variable is continuous, as in CRSMs, numerical integration such as Gauss-Hermite quadrature (Abramowitz and Stegun (1987)) is needed. We only need to calculate the conditional means $E[\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)]$ and conditional variances $V[\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)]$ for a few selected fixed values of $z_2(t-\tau_1)$, which can be calculated using the Kalman filter. The marginal expectation is then a weighted sum of these values. The details are given in the appendix. Since this is one-dimensional integration and the number of knots needed for an accurate integration is very small (10 in our example), the resultant algorithm is computationally efficient.

The outline of the algorithm is as follows.

Approximate method:

(1.) Calculate $(\Gamma_2(N)|Y_1(N))$; apply the Kalman filter and smoothing algorithms to calculate the means and variances of $(X_2(N)|Y_2(N))$; calculate the means and variances of $(Z_2(N)|Y_2(N))$ from $(X_2(N)|Y_2(N))$.

(2.) Conditional on $(Z_2(N)|Y_2(N))$, apply the forward algorithm to Y_1, which is the Kalman filter with an approximation step using equations (4) and (5) at each filtering step; apply the smoothing method based on the approximated filtered results to obtain the means and variances of $(X_1(N)|Y_1(N))$; calculate the means and variances of $(Z_1(N)|Y_1(N))$.

(3.) Conditional on $(Z_1(N)|Y_1(N))$, apply the forward algorithm to Y_2, which involves an approximation step using equations similar to (4) and (5) at each filtering step; apply the smoothing method based on the approximated filtered results to obtain the means and variances of $(X_2(N)|Y_2(N))$; calculate the means and variances of $(Z_2(N)|Y_2(N))$.

(4.) Iterate between (2) and (3) until convergence.

The extension to m-step-ahead forecasting is immediate. That is, replace $(z_1(t-\tau_2)|Y_1(N))$ and $(z_2(t-\tau_1)|Y_2(N))$ by their forecasted distributions, when $t-\tau_1, t-\tau_2 > N$. Since no extra smoothing is required, the forecasting is carried out in parallel operations.

4. CRSMs with Conditionally Gaussian Univariate Models

In this section, we extend CRSMs to the case where the underlying univariate models are conditionally Gaussian themselves. One of the advantage of CRSMs is that, when the underlying univariate models are conditionally Gaussian, the resultant CRSMs are still conditionally Gaussian. That is, conditioned on the external signal and We focus on the case when the underlying univariate models are MDLMs, which are special conditionally Gaussian models where the mixture variable is discrete, although the methods we describe here can also apply to other types of conditionally Gaussian models (see Shepard (1994) for examples).

Consider the following model.

(A) The observation equations:

$$y_1(t) = F_1(t)\boldsymbol{x}_1(t) + v_1(t),\ v_1(t) \sim N(0, \sigma^2_{e1}(t))$$
$$y_2(t) = F_2(t)\boldsymbol{x}_2(t) + v_2(t),\ v_2(t) \sim N(0, \sigma^2_{e2}(t)).$$

(B) The system level equations:

$$\boldsymbol{x}_1(t) = H_1(t)\boldsymbol{x}_1(t-1) + \boldsymbol{w}_1(t),$$
$$\boldsymbol{x}_2(t) = H_2(t)\boldsymbol{x}_2(t-1) + \boldsymbol{w}_2(t),$$

with

$$\boldsymbol{w}_1(t) \sim N(\boldsymbol{\mu}_{1j}(t), \Sigma_{1j}(t)) \text{ if } i_1(t) = j, j = 0, ..., k_1 - 1,$$
$$\boldsymbol{w}_2(t) \sim N(\boldsymbol{\mu}_{2j}(t), \Sigma_{2j}(t)) \text{ if } i_2(t) = j, j = 0, ..., k_2 - 1,$$

where $i_1(t)$ and $i_2(t)$ are mixture variables that are indicators of the underlying processes.

(C) The cross-regressed equations:

$$\gamma_1(t) = g_1\{\alpha_1 + \beta_1 z_2(t - \tau_1)\}$$
$$\gamma_2(t) = g_2\{\alpha_2 + \beta_2 z_1(t - \tau_2)\}.$$

Denote $I_1(N) = \{i_1(1), i_1(2), \ldots, i_1(N)\}$ and $I_2(N) = \{i_2(1), \ldots, i_2(N)\}$. The block Gibbs sampler is now extended as follows.

Block Gibbs Sampler:

(1.) Generate a candidate point $(Z_1^*(N), Z_2^*(N), I_1^*(N), I_2^*(N))$

(2.) Calculate $(\Gamma_1(N)|Z_2^*(N))$ and generate $I_1^*(N)$ from $p(I_1(N)|Y_1(N), Z_1^*(N), \Gamma_1(N))$; draw $X_1^*(N)$ from $p(X_1(N)|Y_1(N), \Gamma_1, I_1^*(N))$; calculate $Z_1^*(N)$ from $X_1^*(N)$.

(3.) Calculate $(\Gamma_2(N) \,|\, Z_1^*(N))$ and generate $I_2^*(N)$ from $(I_2(N) \,|\, Y_2(N), Z_2^*(N), \Gamma_2(N))$. Then draw $X_2^*(N)$ from $p(X_2(N)|Y_2(N), \Gamma_2(N), I_2^*(N))$; calculate $Z_2^*(N)$ from $X_2^*(N)$.

(4.) Go to (2.)

Readers are referred to Carter and Kohn (1994, 1996) and Shephard (1994) for details on how to use the Kalman filter and smoothing algorithm to generate the indicator variables $I_1(N)$ and $I_2(N)$ and the state variables $X_1(N)$ and $X_2(N)$.

Because of the conditionally Gaussian structure of CRSMs, the approximate method can also be extended here. The method follows the same iterative scheme given in Section 3, except a filtering step now involves two steps of approximations: the first with respect to the integration over the external latent signals, the second with respect to the normal approximation to the mixture of normals. The second step is a regular Harrison-Stevens approximation. Since the calculation is straight-forward, we omit the details.

5. Estimation of the Unknown Parameters

Because CRSMs are conditionally Gaussian, the parametric estimation methods for conditionally Gaussian models can be applied. Some of the references are Carter and Kohn (1994, 1996, 1997) for a Bayesian approach, Shepard (1994)

for simulated EM, and Billio and Monfort (1997) for simulated likelihood. The Bayesian methods connect naturally with the block Gibbs sampler. Though, the priors have to be specified case-by-case. Since the number of simulations needed to obtain $3-4$ decimal place accuracy for the criterion function in simulated EM and likelihood steps is very large, these methods are only practical when the number of unknown parameters is small. In the problems that we consider, there are usually a large number of unknown parameters to be estimated. We therefore focus on the approximate likelihood method using the normal approximation to the posterior distributions method. Denote the collection of unknown parameters as $\Theta = \{\Theta_1, \Theta_2\}$, where Θ_1 contains all the parameters associated with series one and Θ_2 includes all the parameters associated with series two. The likelihood can be calculated using the one-step-ahead-prediction density $p(y_i(t)|Y_i(t-1), \Theta_i)$ which is available in the Kalman filter:

$$L(\Theta_i|Y_i(N)) = \prod_{t=1}^{N} p(y_i(t)|Y_i(t-1), \Theta_i),\ i = 1, 2, \tag{6}$$

where $Y_i(t-1)$ denotes the collection of whole history up to $t-1$. Because of the normal approximation to the filtered distributions, this is only an approximate likelihood.

The iterative scheme becomes this.

(1a) Conditional on $Y_1(N)$, obtain $\hat{\Theta}_2$ that maximizes $L(\Theta_2|Y_2(N))$, which only involves the forward filtering algorithm.

(1b) Calculate $(X_2(N)|Y_2(N), \hat{\Theta}_2)$ using the smoothing algorithm; calculate $(Z_2(N)|Y_2(N), \hat{\Theta}_2)$.

(2a) Conditional on $(Z_2(N)|Y_2(N),\ \hat{\Theta}_2)$, obtain $\hat{\Theta}_1$ that maximizes $L(\Theta_1|\ Y_1(N))$.

(2b) Calculate $(X_1(N)|Y_1(N),\ \hat{\Theta}_1)$ using the smoothing algorithm; calculate $(Z_1(N)|\ Y_1(N),\ \hat{\Theta}_1)$.

(3a) Conditional on $(Z_1(N)|Y_1(N),\ \hat{\Theta}_1)$, obtain $\hat{\Theta}_2$ that maximizes $L(\Theta_2|\ Y_2(N))$.

(3b) Calculate $(X_2(N)|Y_2(N), \hat{\Theta}_2)$ using the smoothing algorithm; calculate $(Z_2(N)|Y_2(N), \hat{\Theta}_2)$.

(4) Iterate between (2) and (3) until convergence.

In our experience, this algorithm usually converges within two or three iterations. Then $(Z_1(N)|Y_1(N), \hat{\Theta}_1)$ and $(Z_2(N)|Y_2(N),\ \hat{\Theta}_2)$ in the final stage serve as estimates of the latent signals.

So far we have treated the lags as known. In practical applications, researchers usually have a rough idea about the range of the lags. Estimation of the lags can be done by fitting models with different values for the lags and choosing the ones that maximize the profile likelihoods.

6. Application to Hormone Data

We illustrate the methods proposed above by modeling the bivariate hormone time series shown in Figure 1. One of the objectives is to study the relationship of the pulsatile secretions between the two hormones. Using a CRSM, we explore whether a rise or a drop in the concentration of one hormone has an effect on the release of the other hormone. As an illustration, we show the result of the analysis of portal GnRH and jugular LH of ewe 1. The results of the other five ewes are consistent with those of the first.

Under the exponential-decay-toward-baseline assumption, which is a result from the one compartmental model, we have the following model.

(A) The observation equations:

$$\begin{aligned} y_1(t) &= x_1(t) + b_1 + v_1(t),\ v_1(t) \sim N(0, \sigma_{e1}^2) \\ y_2(t) &= x_2(t) + b_2 + v_2(t), v_2(t) \sim N(0, \sigma_{e2}^2), \end{aligned} \tag{7}$$

where b_1 and b_2 are the baselines; $x_1(t)$ and $x_2(t)$ are the pulsatile components; $v_1(t)$ and $v_2(t)$ are the measurement errors.

(B) The system level equations:

$$\begin{aligned} x_1(t) &= a_1 x_1(t-1) + w_1(t) \\ x_2(t) &= a_2 x_2(t-1) + w_2(t), \end{aligned} \tag{8}$$

with

$$\begin{aligned} w_1(t) &\sim N(\mu_{1j}, \sigma_{1j}^2) \text{ if } i_1(t) = j, j = 0, 1, \\ w_2(t) &\sim N(\mu_{2j}, \sigma_{2j}^2) \text{ if } i_2(t) = j, j = 0, 1, \end{aligned}$$

where a_1 and a_2 are the decay factors which are functions of the half-lifes, $i_1(t) = 1$ or $i_2(t) = 1$ indicates a pulse at time t while $i_1(t) = 0$ or $i_2(t) = 0$ indicates no pulse. With μ_{10}, μ_{20}, σ_{10}^2 and σ_{20}^2 set to zero, $w_1(t)$ or $w_2(t)$ is zero with probability one when $i_1(t) = 0$ or $i_2(t) = 0$; μ_{11}, μ_{21}, σ_{11}^2 and σ_{21}^2 are the mean pulse amplitudes and their associated variances. Denote the prior probabilities of being pulses as $\pi_1(t) = p(i_1(t) = 1)$ and $\pi_2(t) = p(i_2(t) = 1)$.

(C) The cross-regressed equations:

$$\begin{aligned} \text{logit}\{\pi_1(t)\} &= \alpha_1 + \beta_1 z_2(t - \tau_1) \\ \text{logit}\{\pi_2(t)\} &= \alpha_2 + \beta_2 z_1(t - \tau_2), \end{aligned} \tag{9}$$

where $\text{logit}\{\pi_1(t)\} = \log\{\pi_1(t)/(1 - \pi_1(t))\}$, and $z_1(t)$ and $z_2(t)$ are the latent concentrations, modeled by $b_1 + x_1(t)$ and $b_2 + x_2(t)$ respectively.

The unknown parameters besides $\{\tau_1, \tau_2\}$ are $\Theta_1 = \{b_1, a_1, \sigma_{e1}, \mu_{11}, \sigma_{11}, \alpha_1, \beta_1\}$ and $\Theta_2 = \{b_2, a_2, \sigma_{e2}, \mu_{21}, \sigma_{21}, \alpha_2, \beta_2\}$. The initial states $x_1(0)$ and $x_2(0)$ are assigned diffuse prior distributions $N(0, 10000)$. When $z_2(t - \tau_1)$ and $z_1(t - \tau_2)$

are not yet available in the first few values, we assign $\pi_1(t)$ and $\pi_2(t)$ to their average values, obtained by running the two univariate models separately. The two averages are 0.323 for GnRH (series 1) and 0.296 for LH (series 2). Because of the large number of unknown parameters, we use the approximate likelihood method to estimate the parameters. A grid of 1 to 12 lags is used to search for the optimal estimates $\hat{\tau}_1$ and $\hat{\tau}_2$. Using the approximate maximum profile likelihoods, $\hat{\tau}_1 = 2$ and $\hat{\tau}_2 = 1$. This means that pulses in GnRH lead those in LH by 1 time point (5 minutes), while LH has a 2 time point (10 minutes) lagged feedback to GnRH.

Table 1. Point estimates of the parameters with their 95% bootstrap confidence intervals.

	GnRH		LH	
Parameters	Estimate	95% CI	Estimate	95% CI
a	0.11	(0.10, 0.13)	0.81	(0.75, 0.86)
σ_e	0.10	(0.07, 0.12)	0.75	(0.58, 0.92)
β	-2.78	$(-6.51, -1.42)$	6.09	(2.93, 22.88)
α	24.09	(12.07, 57.62)	-5.94	$(-21.87, -3.87)$
μ	4.33	(2.98, 6.06)	3.54	(3.03, 4.08)
σ_w	3.44	(2.25, 4.34)	0.95	(0.05, 1.39)
b	0.27	(0.23, 0.30)	4.64	(2.46, 6.15)

The estimates of parameters, along with their 95% bootstrap confidence intervals, are shown in Table 1. Because of the normal approximations to the filtered distributions, the likelihoods are approximate and the naive observed information matrices do not provide correct variance estimates. We used a parametric bootstrap to obtain the 95% confidence intervals. We generated 1000 replicates from the estimated parameters and re-estimated the parameters. We excluded 6 extreme outliers. The boxplots of the approximate MLEs are shown in Figure 2. The approximate method appears to yield unbiased estimates. From Table 1, the rises in GnRH appear to trigger the pulsatile secretions of LH, since $\hat{\beta}_2$ is positive and its 95% confidence interval does not include zero. The increase in LH may have a negative feedback to the GnRH release, since the $\hat{\beta}_1$ is negative and its 95% confidence interval does not include zero. Other parameters of interest are the mean pulse amplitudes $\mu_1 = 4.33$ and $\mu_2 = 3.54$, and the baselines $b_1 = 0.27$ and $b_2 = 4.64$. They are all significantly different from zero since their 95% confidence intervals do not include zero. For simplicity, we did not re-estimate the lags in the parametric bootstrap. This can be done easily to quantify the variability of the estimates of the lags.

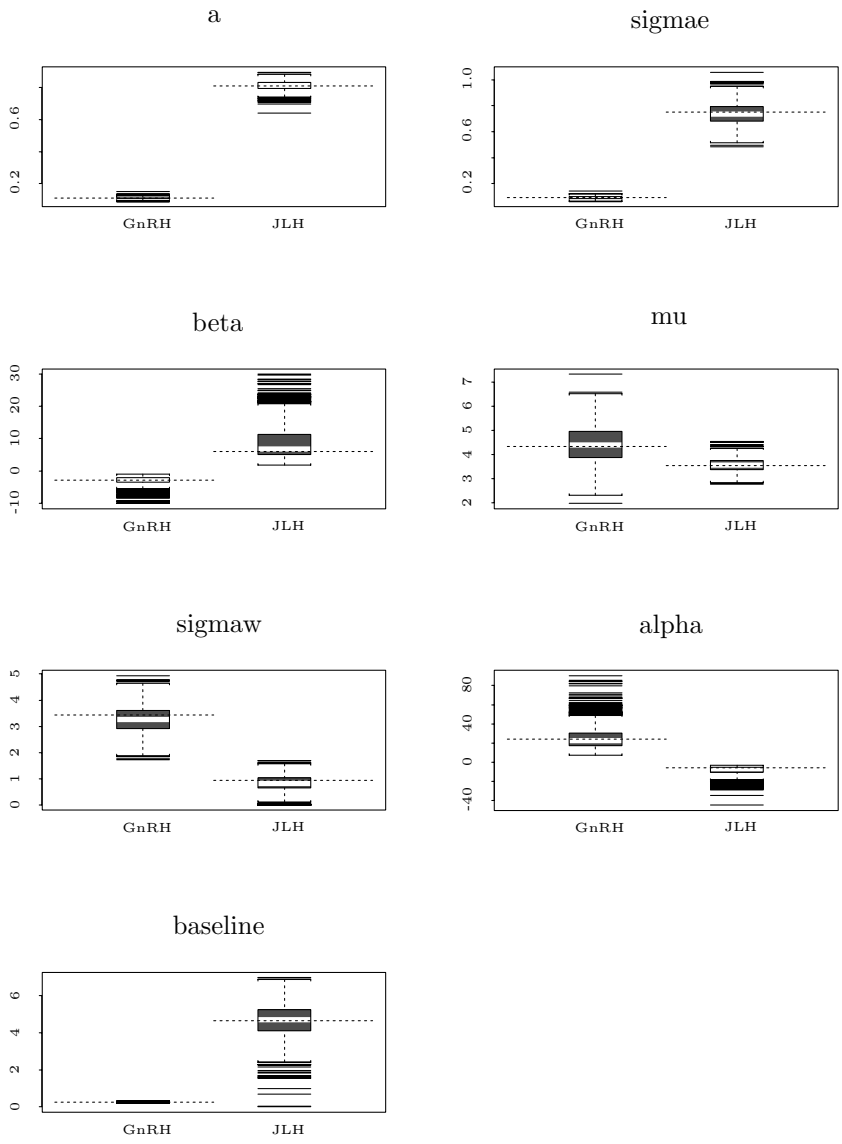

Figure 2. Boxplots of the approximate MLEs in the 1000 parametric bootstraps. The dotted lines are the true values.

The upper panels of Figure 1 show GnRH and LH with their fitted values. The lower panels plot the posterior probabilities of the pulse indicators. The smoothed values of GnRH almost completely overlap the observed ones, since the portal GnRH data are very clean. In LH, the model only picks up the large jumps and ignores the small blips. From the concentration data, every GnRH pulse is followed by a LH pulse, the secretion of GnRH seems to be stopped by

the rise of LH. When LH drops to a certain level, it appears to trigger another GnRH pulse. This phenomenon is confirmed by the posterior probabilities of the pulse indicators, and also matches the parameter estimates of the model.

For comparison, we re-fitted the two time series separately using two multi-process dynamic linear models without a feedback relationship. The results are shown in Figure 3. It can be seen that the estimation of pulses in GnRH is still relatively clean because of the small measurement errors in GnRH. The estimation of the pulse locations in the LH is very noisy and the one-to-one relationship between the hormones becomes difficult to see. We conclude that the CRSM not only identifies the potential feedback relationship between GnRH, but also produces better estimates of the pulses by borrowing the external information from the other hormone.

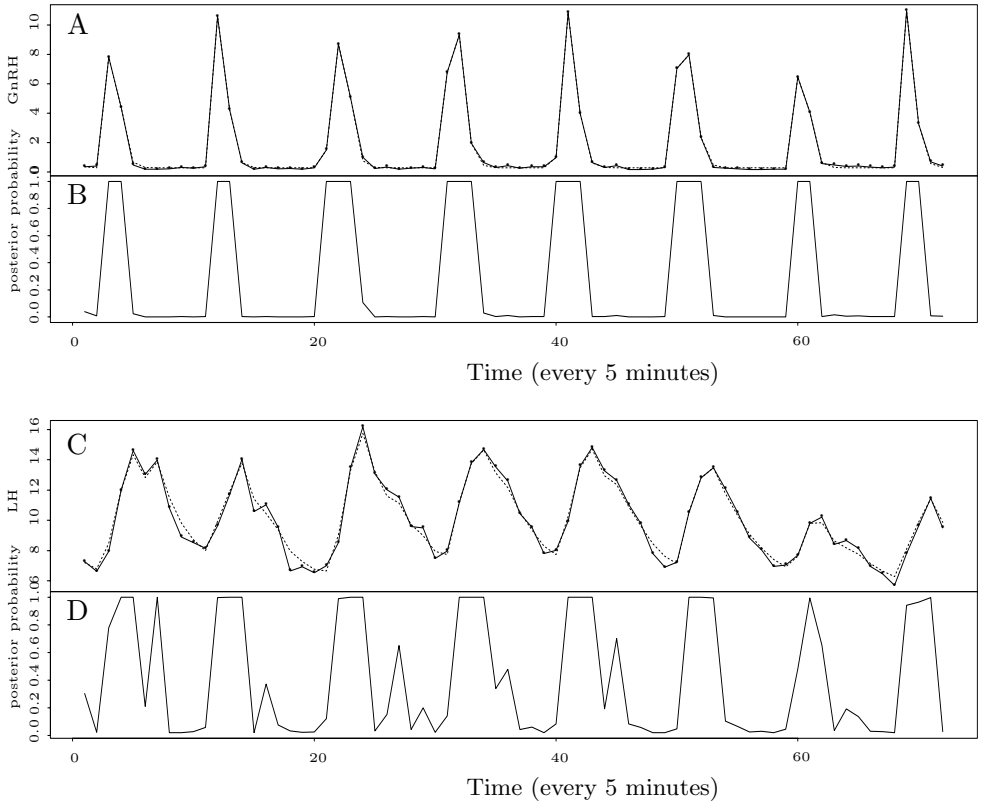

Figure 3. Separate fits without feedback relationship. A: observed and fitted values of GnRH. The dots with solid lines are observed concentrations and the dotted lines are the smoothed values. B: posterior probabilities of the pulse indicators. C: observed and fitted values of LH. D: posterior probabilities of the pulse indicators.

Despite the statistical significance of β_1 and β_2, we have to point out that the probabilistic relationships in our model do not ensure causality. For example, the apparent LH feedback to GnRH can be an artifact of a constant latency between GnRH peaks. Guo and Brown (2000) showed that regularity can be explained by a negative feedback relationship. The stronger the negative feedback, the more regular the system may appear. Such a negative feedback relationship needs to be confirmed by controlled experiments.

7. Concluding Remarks

We have introduced a class of bivariate time series models, in which each time series is modeled by a structural time series model and the structural parameters are modeled as regressions over the history of the other series. Traditional vector time series models require all time series to share the same model structure and characterize the relationship between series through marginal correlation coefficients. In contrast, CRSMs characterize the relationship through cross-regressions. As a result, these models allow very flexible relationships that can be potentially across time series models with very different structures. Another unique structure of CRSMs is that the cross-regressions model the relationship through the structural parameters instead of directly through the outcomes. This not only increases the flexibility in modeling different relationships, but also introduces a robust probabilistic relationship which is particularly useful when the relationship is not one-to-one. Finally, CRSMs are conditionally Gaussian in general when the underlying univariate time series models are Gaussian or conditionally Gaussian. This unified structure enables us to adapt existing efficient estimation procedures for state space models.

For simplicity, we introduced the framework for bivariate time series with a single pair of crossed lagged relationships. This can be extended to multivariate time series and cases with multiple cross-relationships. The cross-regression should be viewed as a general way to parameterize a complex relationship, which can include internal history as well as external history as regressors. These relationships can be either linear or nonlinear. The feedback relationship can also include multiple time points. Covariates can also be introduced into the cross-regressions. If we incorporate time dependent covariates into the cross-regressions, the relationship can then also be changing over time. The cross-regression structure introduces a new concept of modeling relationships between time series.

A special case of interest is when the relationship is one directional. In this case, we only need one equation in the cross-regression and the lag can be any nonnegative integer. The iterative estimation procedure reduces to a two-step procedure: first estimate the influential time series, then incorporate

the posterior estimates of this time series in estimating the second time series. When the observed signal of the influential time series is used, the structural parameters of interest in the second time series are the mean parameters and the lag is set to zero, this model reduces to a Multiregression Dynamic Model (MDM) (Queen and Smith (1993)).

Similar to other regression models, causality cannot be proven by the use of these models. For example, in the physiological data set presented in this paper, the role of GnRH as a precursor to LH has been established, while the feedback mechanism of LH on GnRH is yet to be confirmed by controlled experiments. Therefore, we emphasize that the use of these models is only an aid in exploring many potential relationships. Although some model selection criteria, such as AIC and Bayes factor, can be used as supplements in choosing among alternative models, physical (or physiological) relevance is the key in model selection.

Theoretically, CRSMs can be extended to general non-Gaussian cases. However, it is well-known that estimation of general non-Gaussian state space models is very computationally intensive. This extension of CRSMs can only intensify the needed computation. It will only be practical when more efficient estimation procedures for the general non-Gaussian state space models become available.

Appendix. Numerical integration for equations (4) and (5)

To calculate the marginal expectation and variance of the filtered distribution, we need to calculate the conditional expectations and conditional variances for a few selected values of $z_2(t-\tau_1)$, then calculate the marginal values as weighted sums. The selected knots b_i and corresponding weights w_i, for $i = 1, \ldots, T$ are listed in Abramowitz and Stegun (1987). The number of knots T only needs to be 10 or 20 to produce accurate estimates in our example. The elements needed in equations (4) and (5) are:

$$E[\boldsymbol{x}_1(t)|Y_1(t)] = \sum_{i=1}^{T} w_i exp(b_i^2) E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1) = b_i) f(z_2(t-\tau_1) = b_i);$$

$$V[E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1)] = E[E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1))]^2 - [E(\boldsymbol{x}_1(t)|Y_1(t))]^2;$$

$$E[E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1))]^2 = \sum_{i=1}^{T} w_i exp(b_i^2) [E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1) = b_i)]^2 f(z_2(t-\tau_1) = b_i);$$

$$E[V(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1))] = \sum_{i=1}^{T} w_i exp(b_i^2) V(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1) = b_i) f(z_2(t-\tau_1) = b_i).$$

The elements $E(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1) = b_i)$ and $V(\boldsymbol{x}_1(t)|Y_1(t), z_2(t-\tau_1) = b_i)$ are available in the Kalman filter. The distribution $f(z_2(t-\tau_1) = b_i)$ is Gaussian.

Acknowledgement

This research was supported by NIH grants R01 CA84438, R01 NIMH62298, P60 DK20572 and P30 HD18258. We thank the two referees for their helpful comments that substantially improved this article. The authors also thank Dr. A. Rees Midgley for allowing us to use the hormone data.

References

Abramowitz, M. and Stegun, I. A. (1987). *Handbook of Mathematical Functions*. Dover Publications, New York.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Englewood Cliffs.

Billio, M. and Monfort, A. (1998). Switching state space models likelihood function, filtering and smoothing. *J. Statist. Plann. Inference* **68**, 65-103.

Bolstad, W. (1988). The multiprocess dynamic linear model with biased perturbations: a real time model for growth hormone level. *Biometrika* **75**, 685-692.

Carlin, B. P., Polson, N. G. and Stoffer, D. (1992). A Monte Carlo approach to non-normal and nonlinear state space modeling. *J. Amer. Statist. Assoc.* **87**, 493-500.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541-553.

Carter, C. K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* **83**, 589-601.

Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *J. Roy. Statist. Soc. Ser. B* **59**, 255-268.

De Jong, P. (1989). Smoothing and interpolation with the state-space model. *J. Amer. Statist. Assoc.* **84**, 1085-1088.

De Jong, P. and Shephard, N. (1995). Simulation smoother for time series models. *Biometrika* **82**, 339-350.

Fruhwirth-Schnatter, S. (1994). Data augmentatation and dynamic linear models. *J. Time Ser. Anal.* **15**, 183-202.

Gordon, A. E. and Smith, A. F. M. (1990). Monitoring and modeling biomedical time series. *J. Amer. Statist. Assoc.* **85**, 328-337.

Guo and Brown (2000). Structural time series models with feedback mechanisms. *Biometrics* **56**, 686-691.

Guo, W., Wang, Y. and Brown, M. B. (1999). A signal extraction approach to modeling hormone time series with pulses and a changing baseline. *J. Amer. Statist. Assoc.* **94**, 746-756.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser. B* **38**, 205-247.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, A. C. and Shephard, N. (1992). Structural time series models. *Handbook of Statistics* **11**. North-Holland, Amsterdam.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series (with discussion). *J. Amer. Statist. Assoc.* **82**, 1032-1063.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika* **80**, 117-126.

Midgley A. Rees Jr., McFadden K., Ghazzi M., Karsch F. J., Brown M. B., Mauger D. T. and Padmanabhan V. (1997). Nonclassical secretory dynamics of LH revealed by hypothalamo-hypophyseal portal sampling of sheep. *Endocrine* **6**, 133-143.

Queen, C. M. and Smith, J. Q. (1993). Multiregression dynamic models. *J. Roy. Statist. Soc. Ser. B* **55**, 849-870.

Shephard, N. (1994). Partially non-Gaussian state space models. *Biometrika* **81**, 115-131.

Shumway, R. H. and Stoffer, D. S. (1991). Dynamic linear models with switching. *J. Amer. Statist. Assoc.* **86**, 763-769.

Smith, A. F. M. and West, M. (1983). Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics* **39**, 867-878.

Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Ann. Statist.* **21**, 1701-1762.

Weiss, J., Jameson, J. L., Burrin, J. M. and Crowley, W. F. (1990). Divergent responses of gonadotropin subunit messenger RNA's to continuous versus pulsatile gonadotropin-releasing hormone in vitro. *Molecular Endocrinology* **4**, 557-564.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* 2nd edition. Springer-Verlag, New York.

Department of Biostatistics and Epidemiology, University of Pennsylvania, 613 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021, U.S.A.

E-mail: wguo@cceb.upenn.edu

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.

(Received October 1999; accepted April 2001)