Gibbs Sampling in Linear Models #2

Econ 690

Purdue University

Outline

- In this lecture, we extend our previous lecture on Gibbs sampling in the linear regression model.
- In particular, we consider three additional variants of the linear model:
  1. A linear regression model with a single, unknown changepoint.
  2. The seemingly unrelated regressions [SUR] model.
  3. A linear model with inequality restrictions on the regression coefficients.

A Linear Model with a Single Changepoint

Suppose that the density for a time series y_t, $t = 1, 2, \cdots, T$, conditioned on its lags, the model parameters and other covariates, can be expressed as

-

In this model, λ is a changepoint - for periods until λ, one regression is assumed to generate y, and following λ, a new regression is assumed to generate y.

Suppose you employ priors of the form:

$$\begin{aligned} \theta = [\theta_1 \ \theta_2]' &\sim N(\mu_\theta, V_\theta) \\ \beta = [\beta_1 \ \beta_2]' &\sim N(\mu_\beta, V_\beta) \\ \sigma^2 &\sim IG(a_1, a_2) \\ \tau^2 &\sim IG(b_1, b_2) \\ \lambda &\sim \text{Uniform}\{1, 2, \cdots, T-1\}. \end{aligned}$$

Note that λ is treated as a parameter of the model, and by placing a uniform prior over the elements $1, 2, \cdots, T-1$, a changepoint is assumed to exist.

For this model, we will do the following:

1. (a) Derive the likelihood function.
2. (b) Describe how the Gibbs sampler can be employed to estimate the parameters of this model, given the priors above.

First, for the likelihood function, note that the data essentially divides into two parts, *conditioned on the changepoint* λ.

As such, standard results can be used for the linear regression model (as in our previous lectures) to simulate the regression and variance parameters within each of the two regimes. .

Specifically,

-

The joint posterior is proportional to the likelihood above times the given priors for the model parameters.

So, in this model, what will be the complete posterior conditionals for β and θ? Using our established results for the linear regression model, we know:

-

where

-

In the above, we have defined

-

Similarly, we obtain the posterior conditional for β:

$$\beta|\theta, \sigma^2, \tau^2, \lambda, y \sim N(D_\beta d_\beta, D_\beta),$$

where

$$D_\beta = \left(X_\beta' X_\beta/\tau^2 + V_\beta^{-1}\right)^{-1} \quad \text{and} \quad d_\beta = X_\beta' y_\beta/\tau^2 + V_\beta^{-1}\mu_\beta.$$

Where, again, we have defined

What about the complete conditional posterior distributions for the variance parameters σ^2 and τ^2? Can you derive these?
It is straightforward to show that

-

and

-

For the changepoint λ, simulation is not as standard.

What we do know is (under our uniform prior on λ):

-

where $\lambda \in \{1, 2, \cdots T-1\}$.

Since λ is discrete-valued, we can:

1. Calculate the above for each possible value of λ (given the current values of the other parameters).
2. Normalize these values into probabilities by dividing each value from the first step by the sum of all values obtained from the first step.
3. Sample λ by drawing from this discrete distribution.

A posterior simulator proceeds by iteratively sampling from all of these conditional posterior distributions.

Example With U.S. Average Annual Temperature Data

- We now provide an example of the previous method using temperature data.
- Specifically, we obtain data on the average annual temperature in the U.S. from 1895-2006 ($n = 112$).
- We seek to determine if there is a "change" or "break" in the historical temperature path, perhaps consistent with the idea of global warming.
- If there is such a break, and if global warming concerns are true, then we might expect that recent years have been characterized by more rapid increases in temperature than years in the distant past.
- In this example, we aren't thinking about a discrete jump in temperatures, but rather just a change in slopes.

To this end, we consider a restricted version of our previous model:

-

where

-

The changepoint in this model is λ, and periods before and after λ will have different slopes.

For this application, we focus attention on fitting a continuous function, and do not allow for discrete jumps as in our original presentation of the model.

We restrict $\lambda \in \{3, 4, \ldots, T-3\}$.

Thus, we assume that a changepoint exists, and force the changepoint to occur toward the "interior" of the sample period.

For our priors, we specify a uniform prior for λ and also specify

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \sim N\left[\begin{pmatrix} 52 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{pmatrix} \right].$$

In this model, the complete posterior conditionals for α and σ^2 are straightforward.

That is, given λ, the covariate matrix X_λ can be constructed, and the conditionals for these parameters are of standard forms.

For the changepoint λ, under our flat prior, we obtain:

-

Thus, we must calculate the likelihood (for a given α and σ^2) for each possible value of $\lambda \in \{3, 4, \ldots, T-3\}$, and then normalize these quantities.

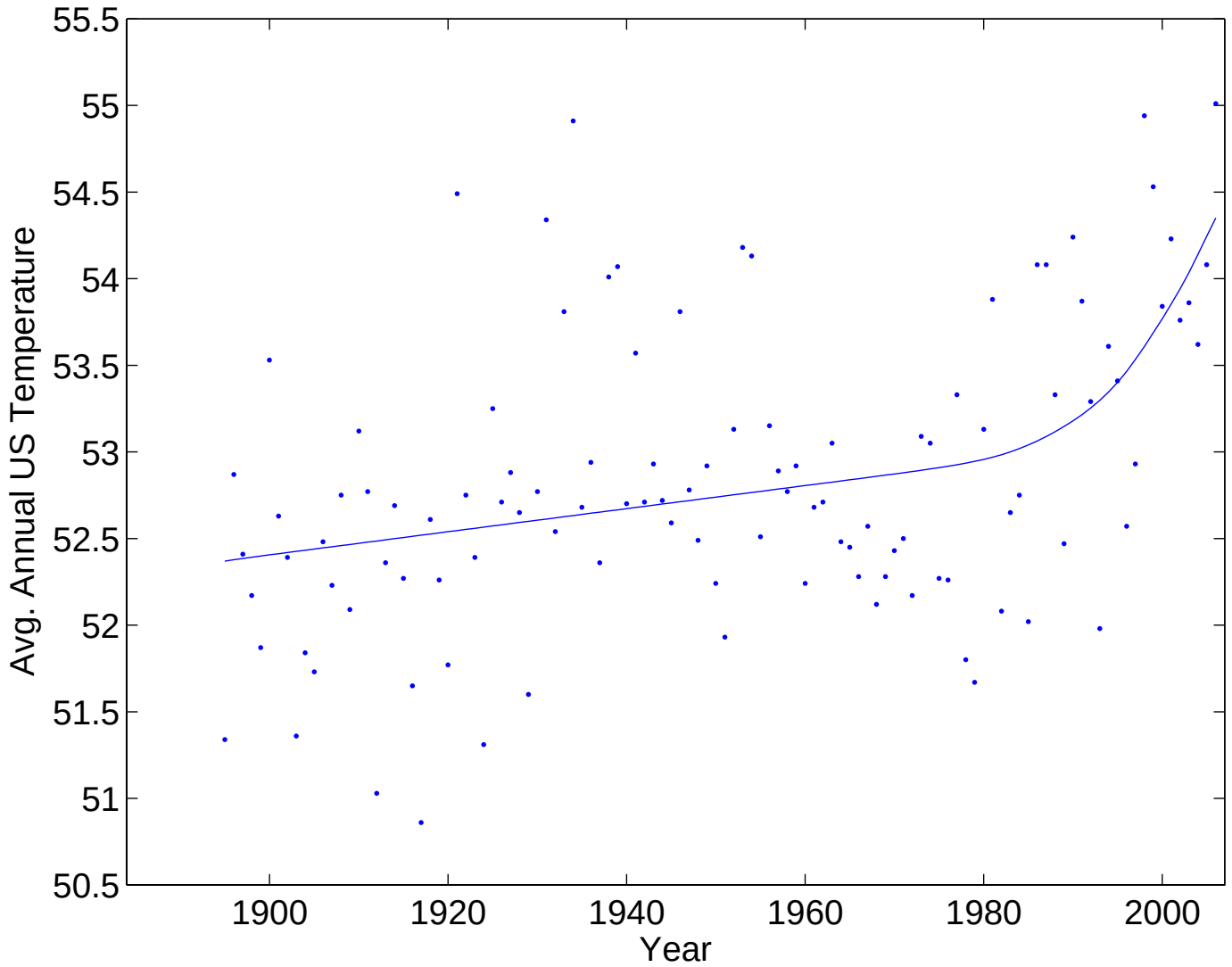
The changepoint λ is then drawn from this discrete distribution.

- The following slides present results of this estimation exercise.
- The simulator was run for 50,000 iterations, and the first 1,000 were discarded as the burn-in.
- We present two graphs: The first plots the posterior mean of

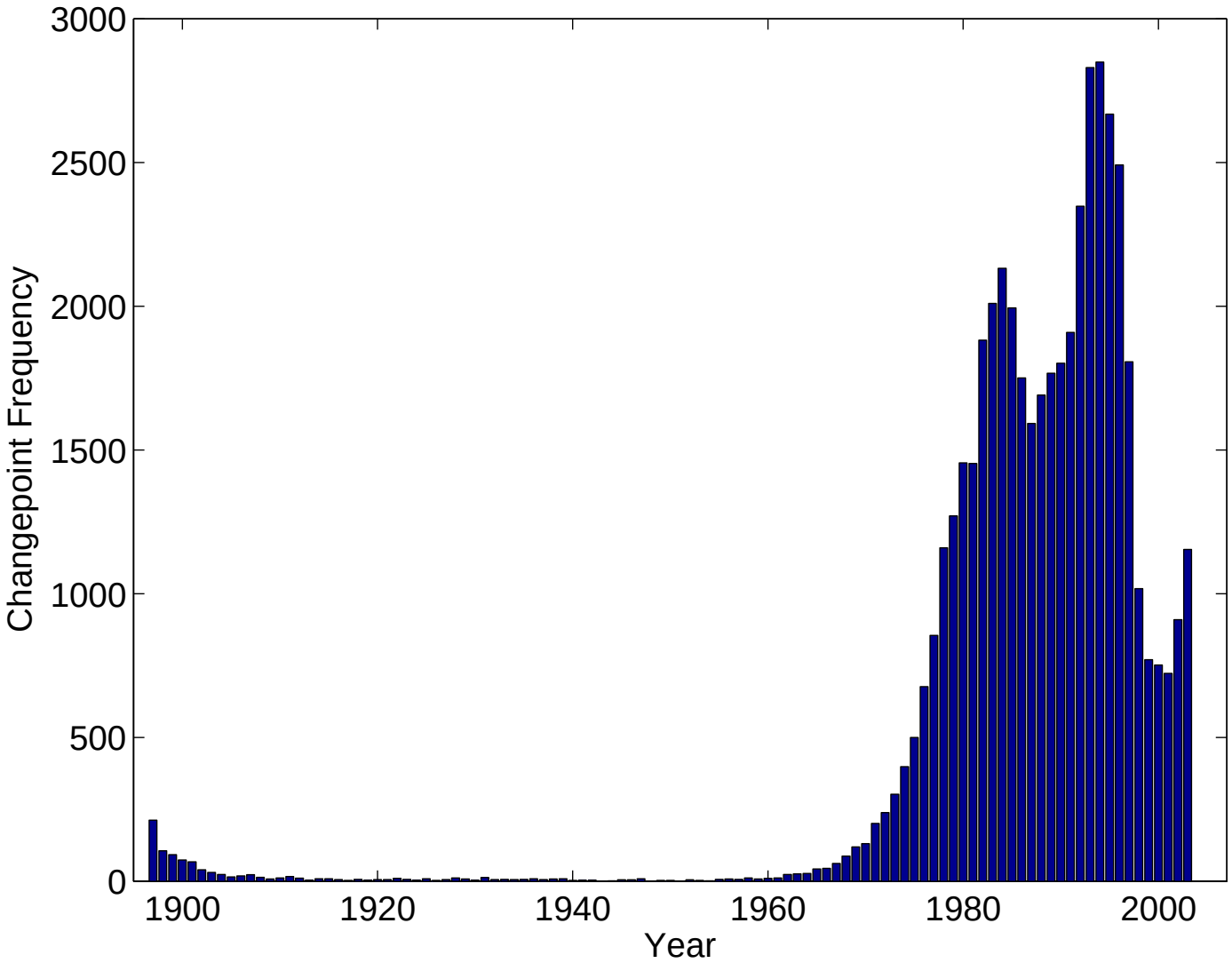
$$\hat{y} \equiv X_\lambda \alpha.$$

- Note that this function also averages over the uncertainty surrounding the location of the changepoint λ.
- The second graph plots the posterior frequencies associated with the location of λ.

Posterior Mean of Expected Temperature



Posterior Frequencies of Changepoint



- The first graph clearly indicates the location of a changepoint at the latter-end of the sample period (around 1980), and a strong upswing in average temperatures after that period.
- The second graph reaches a similar conclusion, as the posterior distribution of the changepoint λ places most mass toward the end of the sample period.
- Note the revision of our prior, here, which was specified to be uniform over all possible discrete values. If no learning had taken place, we would expect a similar uniform shape for the posterior distribution.

The SUR Model

Consider a two-equation version of the Seemingly Unrelated Regression (SUR) model [Zellner (1962)]:

-

where

$$\epsilon_i = [\epsilon_{i1}\ \epsilon_{i2}]' \overset{iid}{\sim} N(0, \Sigma), \quad i = 1, 2, \cdots, n,$$

x_{i1} and x_{i2} are $1 \times k_1$ and $1 \times k_2$, respectively, and

-

Suppose you employ priors of the form

$$\beta = [\beta_1' \ \beta_2']' \sim N(\mu_\beta, V_\beta)$$

and

-

where W denotes a Wishart distribution.

Note that the Wishart distribution is a multivariate generalization of the gamma distribution, and is the conjugate prior for Σ^{-1} in a multivariate normal sampling model.

The kernel of the Wishart is

-

First, note that the likelihood function can be expressed as

$$L(\beta, \Sigma) = (2\pi)^{-n}|\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\tilde{y}_i - \tilde{X}_i\beta)'\Sigma^{-1}(\tilde{y}_i - \tilde{X}_i\beta)\right),$$

where

$$\tilde{y}_i = [y_{i1}\ y_{i2}],' \quad \tilde{X}_i = \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix}$$

To implement the Gibbs sampler for this model, we must derive $p(\beta|\Sigma, y)$ and $p(\Sigma^{-1}|\beta, y)$. We will now consider the first of these. Observe that we can stack the SUR model as

-

where

-

and y_j, X_j and ϵ_j, for $j = 1, 2$ have been staked over $i = 1, 2, \ldots, n$. In addition, stacked in this way, we obtain:

-

In this form, we can apply our established result for the linear regression model to obtain:

$$\beta|\Sigma, y \sim N(D_\beta d_\beta, D_\beta),$$

where

$$D_\beta = \left(X'(\Sigma^{-1} \otimes I_N)X + V_\beta^{-1}\right)^{-1}$$

and

$$d_\beta = X'(\Sigma^{-1} \otimes I_N)y + V_\beta^{-1}\mu_\beta.$$

The derivation of the conditional posterior distribution for Σ^{-1} is new, and so we will go through the details associated with this derivation.

Combining prior with likelihood, we obtain:

$$p(\Sigma^{-1}|\beta, y) \propto p(\Sigma^{-1})|\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\tilde{y}_i - \tilde{X}_i\beta)'\Sigma^{-1}(\tilde{y}_i - \tilde{X}_i\beta)\right),$$

Using the specific form of our Wishart prior, we get (letting $\epsilon_i = \tilde{y}_i - \tilde{X}_i\beta$) :

-

In this form, it follows that

-

Thus, implementation of the Gibbs sampler for the SUR model only requires sampling from a Normal and a Wishart distribution.

The latter of these (i.e., the Wishart), though perhaps unfamiliar, is easy to draw from.

For example, when ν is an integer (at least as large as the dimension of Σ), as is often the case, we can obtain a draw from a $W(\Omega, \nu)$ density as follows:

First, sample

$$x_i \sim \mathcal{N}(0, \Omega), \quad i = 1, 2, \ldots, \nu.$$

Then, let

$$P = \sum_{i=1}^{\nu} x_i x_i'.$$

It follows that

$$P \sim W(\Omega, \nu).$$

Inequality Constraints

Consider the regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, 2, \cdots, n,$$

where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.
Let $\beta = [\beta_0 \ \beta_1 \ \cdots \ \beta_k]'$. Suppose that the regression coefficients are known to satisfy constraints of the form

$$a < H\beta < b,$$

where a and b are known $(k+1) \times 1$ vectors of lower and upper limits, respectively, and H is a non-singular matrix which selects elements of β to incorporate the known constraints.

The vectors a and b may also contain elements equal to $-\infty$ or ∞, respectively, if a particular linear combination of β is not to be bounded from above or below (or both).

Thus, in this formulation of the model, we are capable of imposing up to $k+1$ inequality restrictions, but no more.

We seek to describe how the Gibbs sampler can be employed to fit this model.

Following Geweke (1996), we will first reparameterize the model in terms of $\gamma = H\beta$.

To this end, let us write the regression model as

$$y = X\beta + \epsilon,$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad \text{and} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Since $\gamma = H\beta$ and H is non-singular, write $\beta = H^{-1}\gamma$.

The reparameterized regression model thus becomes

-

with

-

In the γ-parameterization, note that independent priors can be employed which satisfy the stated inequality restrictions.

To this end, we specify independent truncated Normal priors for elements of γ of the form:

-

with $I(\cdot)$ denoting the standard indicator function.

The posterior distribution is proportional to the product of the priors (with $p(\sigma^2) \propto \sigma^{-2}$) and likelihood:

-

We can make use of our often-used formula to derive the posterior conditional for γ:

-

where

-

and V is the $(k+1) \times (k+1)$ diagonal matrix with j^{th} diagonal element equal to V_j.

The density on the last slide is a multivariate Normal density truncated to the regions defined by a and b.

Drawing directly from this multivariate truncated Normal is non-trivial in general, but as Geweke [1991] points out, the posterior conditional distributions: $\gamma_j|\gamma_{-j}, \sigma^2, y$ (with γ_{-j} denoting all elements of γ other than γ_j) are univariate truncated Normal.

Thus, one can sample each γ_j individually (rather than sampling γ in a single step).

Let

$$\Omega = [\overline{H}]^{-1} = \tilde{X}'\tilde{X}/\sigma^2 + V^{-1},$$

ω_{ij} denote the (i,j) element of Ω, and $\overline{\gamma}_j$ the j^{th} element of $\overline{\gamma}$.

With a bit of work, one can show:

-

where $TN_{(a,b)}(\mu, \sigma^2)$ denotes a Normal density with mean μ and variance σ^2 truncated to the interval (a, b).

The derivation of the posterior conditional for the variance parameter is reasonably standard by this point:

-

A Gibbs sampling algorithm for the regression model with linear inequality constraints thus follows by independently sampling each regression parameter γ_j and then sampling the variance parameter σ^2.