Bayesian Macroeconometrics

Methods and Applications

Joshua C. C. Chan

Sample: Preface and Chapters 1–4

September 2026

Contents

Preface

This book is written for readers who want to understand and implement the Bayesian models and methods used in modern empirical macroeconomics and forecasting. These methods are now standard tools in both academia and central banks, and the models applied in practice have become high-dimensional and computationally intensive. A solid understanding of the underlying algorithms is therefore essential. The book aims to cover these methods in enough depth that readers can both apply and extend them in their own research.

I have written this book in the way I would have wanted to learn. I rarely feel that I truly understand a method until I have derived it from first principles and implemented it myself in prototype code. I also want to understand which features of the data a model is designed to capture, how it is used in practice, and what its strengths and limitations are.

These preferences reflect a simple belief: econometric models and methods should be motivated by the demands of applications, by what the empirical researcher needs to accomplish. Methods serve applications, not the other way around. I have organized the material in this book according to this principle: models appear because they answer questions that macroeconomists ask, and estimation methods appear because those models require them.

The book grew out of a set of lecture notes for my PhD-level courses in Bayesian econometrics. The original goal of these notes was to prepare doctoral students to read the research literature and to conduct independent research in Bayesian macroeconometrics. Over time, the notes were expanded and reorganized, with an emphasis on practical implementation.

This book brings together Bayesian modeling, posterior derivations, algorithmic design, and computation. Derivations are presented in detail, and MATLAB code is integrated directly into the text to illustrate algorithms and empirical applications. The book pays close attention to how posterior samplers are constructed and implemented, especially in high-dimensional settings where these details matter for whether a method is usable at all. Each chapter is anchored by worked examples on real macroeconomic data and ends with exercises that reinforce the material. All the datasets and MATLAB code used in this book, along with R and Python implementations, are available on the companion website, `https://joshuachan.org/bayesmacro`, together with errata and other updates.

A central organizing principle of the book is the role of linear regression. Linear regression is both a foundational model and a building block that reappears, often in disguised form, throughout Bayesian macroeconometrics. Many models used in practice, including state space models and vector autoregressions, can be viewed as collections of linear regressions linked by latent states or dynamic dependence. Even nonlinear and nonparametric models often retain this underlying regression structure in embedded or conditional form. For this reason, this book begins with a careful and detailed treatment of Bayesian linear regression, including explicit derivations of posterior distributions and simulation algorithms. These derivations recur, in modified form, throughout the book, and tie together models that might otherwise look unrelated.

The material is organized into three parts. Part I, Core Bayesian Modeling and Assessment, develops the foundational Bayesian toolkit in regression settings, starting from linear regression, extending to mixture models, and concluding with methods for model comparison. In this part, Markov chain Monte Carlo methods such as the Gibbs sampler and Metropolis–Hastings algorithm are introduced concretely, through the models that motivate them, before their formal treatment in Part II. Part II, Bayesian Computation and High-Dimensional Methods, then develops the underlying theory of Markov chain Monte Carlo and presents more advanced sampling algorithms, along with flexible modeling approaches including Bayesian nonparametrics and modern shrinkage priors. Part III, Bayesian Time-Series Models, applies these tools to dynamic macroeconomic models, including time-varying parameter models, stochastic volatility models, factor models, and vector autoregressions, with particular emphasis on large, high-dimensional models.

The intended audience includes graduate students and researchers in econometrics, macroeconomics, and finance. Readers are expected to have completed a graduate-level or advanced undergraduate-level course in econometrics and to be familiar with matrix algebra and basic concepts from probability and statistics, such as conditional distributions, expectations, and likelihood functions. Appendix B collects the matrix algebra results used in the book, for readers who would like a refresher. The book is suitable both as a primary text for graduate courses in Bayesian macroeconometrics or Bayesian time-series analysis and as a reference for researchers seeking to apply Bayesian methods in empirical macroeconomic work.

If the book succeeds, readers will finish it ready to read the current literature critically, to use these methods in their own empirical work, and to extend them in directions that go beyond what is covered here.

Joshua Chan
West Lafayette
July 2026

Part I

Core Bayesian Modeling and Assessment

1

Foundations of Bayesian Econometrics

Bayesian methods are based on a few elementary rules of probability theory. At the core is Bayes' theorem, which describes how beliefs about unknown model parameters are updated in light of new information. This introductory chapter provides an overview of Bayesian theory and computation through a sequence of simple examples. The basic normal models considered here are special cases of the normal linear regression framework introduced in the next chapter. Many models used in empirical macroeconomics, including vector autoregressions and state space models, can be expressed as systems of regressions, often involving latent states. The derivations and computational ideas introduced here therefore provide the foundation for the macroeconometric models studied throughout the book.

1.1 Bayesian Inference and Computation

This section is organized into two parts. We first describe the basic components of Bayesian inference, including the posterior and posterior predictive distributions. We then discuss how these quantities can be computed in practice using simulation-based methods.

1.1.1 Posterior and Predictive Distributions

The fundamental organizing principle in Bayesian econometrics is Bayes' theorem. It provides the unifying framework for how Bayesian methods estimate model parameters, conduct inference, form predictions, and compare competing models.

Bayes' theorem states that for events A and B, the conditional probability of A given B is:

$$\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A)\mathbb{P}(B \,|\, A)}{\mathbb{P}(B)},$$

where $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are the marginal probabilities of events A and B, re-

spectively. This expression describes how beliefs about event A are updated when information contained in event B becomes available.

To apply Bayes' theorem to estimation and inference, we introduce some notation. Suppose the model is characterized by the density $p(\mathbf{y} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown model parameters. This function plays two roles: viewed as a function of $\mathbf{y}$ for a given parameter value, it specifies how the data are generated; viewed as a function of $\boldsymbol{\theta}$ at the observed data, it is the *likelihood function*, measuring how probable the observed sample would have been under each parameter value.

Now suppose an observed sample $\mathbf{y} = (y_1, \ldots, y_T)'$ is available, and the objective is to learn about $\boldsymbol{\theta}$. The goal of Bayesian methods is to obtain the *posterior distribution* $p(\boldsymbol{\theta} \mid \mathbf{y})$, which summarizes all the information about the parameter vector $\boldsymbol{\theta}$ given the data.

Applying Bayes' theorem, the posterior distribution can be computed as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\boldsymbol{\theta})\, p(\mathbf{y} \mid \boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\boldsymbol{\theta})\, p(\mathbf{y} \mid \boldsymbol{\theta}),$$

where $p(\boldsymbol{\theta})$ is the *prior distribution* and

$$p(\mathbf{y}) = \int p(\boldsymbol{\theta})\, p(\mathbf{y} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$$

is the *marginal likelihood*. The marginal likelihood is the density the model assigns to the observed data after averaging over the prior, and it plays a central role in Bayesian model comparison: the ratio of marginal likelihoods across two models gives the *Bayes factor*, which combines with prior model probabilities to yield posterior model probabilities. We return to these ideas in Chapter 5. Bayes' theorem shows that inference about $\boldsymbol{\theta}$ combines two sources of information: prior beliefs summarized by $p(\boldsymbol{\theta})$ and information in the observed sample $\mathbf{y}$ summarized by the likelihood.

The prior distribution $p(\boldsymbol{\theta})$ reflects beliefs about the parameters before observing the data. The need to specify a prior distribution is sometimes viewed as a weakness of the Bayesian approach. However, priors provide a formal mechanism for incorporating nondata information, such as economic theory or application-specific knowledge, and for imposing structure in high-dimensional settings. From a frequentist perspective, prior distributions can often be interpreted as a form of regularization.

A fundamental feature distinguishing the Bayesian method from the frequentist approach concerns what is treated as fixed and what as random. Frequentist inference treats the parameter $\boldsymbol{\theta}$ as fixed but unknown and the data $\mathbf{y}$ as the random object; estimators, standard errors, and confidence intervals are accordingly evaluated by their behavior across hypothetical repeated samples drawn from $p(\mathbf{y} \mid \boldsymbol{\theta})$. Bayesian inference reverses this conditioning: the

observed sample $\mathbf{y}$ is treated as given, and uncertainty about $\boldsymbol{\theta}$ is described by the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$. As a result, Bayesian probability statements such as $\mathbb{P}(\theta_i > 0 \mid \mathbf{y})$ have a direct interpretation as statements about the parameter conditional on the observed data, whereas frequentist confidence statements refer to long-run coverage across hypothetical repeated samples.

The posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$ characterizes all relevant information about $\boldsymbol{\theta}$ given the data. For example, point estimates may be obtained using the *posterior mean* $\mathbb{E}(\boldsymbol{\theta} \mid \mathbf{y})$ or the *posterior mode* $\operatorname{argmax}_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \mid \mathbf{y})$, while uncertainty about $\boldsymbol{\theta}$ may be summarized using *posterior standard deviations*. In particular, for the ith element of $\boldsymbol{\theta}$, uncertainty can be quantified by $\sqrt{\operatorname{Var}(\theta_i \mid \mathbf{y})}$.

In many applications, interest lies not only in inference about model parameters, but also in predicting future or unobserved outcomes. Bayesian prediction is based on the *posterior predictive distribution*, which integrates over uncertainty about the parameters. For a future observation y_{T+1}, the posterior predictive distribution is given by

$$p(y_{T+1} \mid \mathbf{y}) = \int p(y_{T+1} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{y})\, \mathrm{d}\boldsymbol{\theta}.$$

The posterior predictive distribution provides a *density forecast* for y_{T+1}, accounting for both intrinsic randomness in the data generating process and uncertainty about the parameters.

Point forecasts can be obtained by summarizing the posterior predictive distribution. A common choice is the *posterior predictive mean*,

$$\mathbb{E}(y_{T+1} \mid \mathbf{y}) = \int y_{T+1}\, p(y_{T+1} \mid \mathbf{y})\, \mathrm{d}y_{T+1},$$

which minimizes expected squared forecast error. Measures of predictive uncertainty can be obtained from the posterior predictive variance or from predictive quantiles. Posterior predictive distributions are therefore the basis for forecasting, uncertainty quantification, and model evaluation.

1.1.2 Monte Carlo and MCMC Methods

In principle, many quantities of interest—such as posterior moments, quantiles, and predictive distributions—can be computed once the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$ is known. In practice, however, they are often not available in closed form. In such cases, Monte Carlo simulation is required to approximate them.

To illustrate the main idea, suppose we obtain R *independent* draws from the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$, denoted by $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$. If the first mo-

ment exists, then by the weak law of large numbers the sample mean

$$\widehat{\boldsymbol{\theta}} = \frac{1}{R}\sum_{r=1}^{R} \boldsymbol{\theta}^{(r)}$$

converges in probability to $\mathbb{E}(\boldsymbol{\theta} \,|\, \mathbf{y})$ as $R \to \infty$. Since the simulation size R can be chosen by the researcher, posterior moments can in principle be approximated arbitrarily well, given sufficient computation. Other posterior quantities, such as variances or quantiles, can be estimated analogously using their sample counterparts.

More generally, inference on any function of the model parameters can be conducted in the same way. Given a function $g(\boldsymbol{\theta})$, posterior inference about $g(\boldsymbol{\theta})$ can be carried out by evaluating $g(\boldsymbol{\theta}^{(r)})$ for $r = 1, \ldots, R$ and summarizing the resulting sample.

Thus, Bayesian estimation and inference can often be viewed as a computational problem of obtaining draws from the posterior distribution. In general, sampling from an arbitrary distribution is difficult. Fortunately, a large class of algorithms—collectively known as *Markov chain Monte Carlo* (MCMC) methods—has been developed to sample from complex posterior distributions.

The basic idea behind these algorithms is to construct a Markov chain whose stationary distribution coincides with the target distribution, which in this context is the posterior distribution. By construction, draws produced by MCMC algorithms are autocorrelated. Nevertheless, under mild regularity conditions, ergodic theorems ensure that sample averages computed from these correlated draws converge to their population counterparts (see Section 6.1 for a formal statement). As a result, MCMC output can be used to consistently estimate posterior expectations and other quantities of interest, provided that the corresponding moments exist.

In the sections that follow, we use various simple examples to illustrate these ideas in a concrete setting. Basic MCMC methods, including the Gibbs sampler, are introduced to tackle the specific computational challenges that arise in these examples. A more systematic and theoretical treatment of MCMC methods is deferred to Chapter 6.

1.2 Normal Model with Known Variance

This section introduces Bayesian inference in its simplest parametric setting. The normal model with known variance admits analytical posterior results

and provides a transparent illustration of how prior beliefs are combined with information from the data. The example also serves as a building block for the regression-based and macroeconometric models developed in later chapters.

1.2.1 Model Setup

To illustrate the basic mechanics of Bayesian inference, consider the problem of estimating GDP growth next quarter using a survey of professional forecasters. Let $y_1, \ldots, y_T$ denote the T survey responses, where each y_t represents an annualized GDP forecast from one forecaster. We interpret these forecasts as unbiased but noisy measurements of the quarter-ahead expected GDP growth rate μ. As a simple starting point, suppose that conditional on μ, the survey responses are independent and normally distributed with known variance σ^2, reflecting forecast errors and cross-sectional disagreement. Normality is a convenient starting assumption that yields closed-form posterior results; heavier-tailed and more flexible error distributions are developed in Chapters 3 and 4.

A simple model for this measurement problem is the normal model

$$(y_t \,|\, \mu) \sim \mathcal{N}(\mu, \sigma^2), \quad t = 1, \ldots, T, \tag{1.1}$$

where the variance σ^2 is assumed to be known. The model (1.1) defines the likelihood function $p(\mathbf{y} \,|\, \mu)$.

We adopt the normal prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$, where μ_0 and σ_0^2 are *hyperparameters*, treated as known in this example. They can be calibrated from historical data: for instance, μ_0 may be set equal to the average realized GDP growth rate over the past two decades, while σ_0^2 reflects uncertainty about how informative this long-run average is for the current quarter-ahead forecast. A transparent way to set it is $\sigma_0^2 = \sigma^2/T_0$ for some $T_0 > 0$, so that the prior carries the same information as T_0 hypothetical survey responses; the numerical example below adopts $T_0 = 1$. The normal distribution and its parameterization are summarized in Appendix A.

1.2.2 Posterior Distribution and Interpretation

All the relevant information about μ is summarized by the posterior distribution, which is obtained by Bayes' theorem:

$$p(\mu \,|\, \mathbf{y}) = \frac{p(\mu)\, p(\mathbf{y} \,|\, \mu)}{p(\mathbf{y})}.$$

In this case, the posterior distribution $p(\mu \,|\, \mathbf{y})$ is itself normal. We establish this result explicitly below. Although the algebra may appear tedious at first,

working through this derivation is instructive, as the same calculations and insights recur throughout the book in more general regression and state space models.

In the first step, we write out the likelihood function $p(\mathbf{y} \,|\, \mu)$. Recall that conditional on μ, the observations $y_1, \ldots, y_T$ are independently distributed according to a normal distribution with mean μ and variance σ^2; see Appendix A for the definition and parameterization of the normal distribution. It then follows from (1.1) that the likelihood function is given by

$$\begin{aligned} p(\mathbf{y} \,|\, \mu) &= \prod_{t=1}^{T} (2\pi\sigma^2)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(y_t-\mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2}. \end{aligned} \tag{1.2}$$

Similarly, the prior density $p(\mu)$ is given by

$$p(\mu) = (2\pi\sigma_0^2)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}. \tag{1.3}$$

We now combine the likelihood (1.2) and the prior (1.3) to obtain the posterior distribution. Since the variable of interest in $p(\mu \,|\, \mathbf{y})$ is μ, multiplicative constants that do not depend on μ can be ignored; see Exercise 1.3. By Bayes' theorem,

$$\begin{aligned} p(\mu \,|\, \mathbf{y}) &\propto p(\mu)\, p(\mathbf{y} \,|\, \mu) \\ &\propto \mathrm{e}^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2} \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{\mu^2 - 2\mu\mu_0}{\sigma_0^2} + \frac{T\mu^2 - 2\mu\sum_{t=1}^{T} y_t}{\sigma^2}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}\right)\mu^2 - 2\mu\left(\frac{\mu_0}{\sigma_0^2} + \frac{T\bar{y}}{\sigma^2}\right)\right)\right], \end{aligned} \tag{1.4}$$

where $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$ denotes the sample mean. Because the exponent in (1.4) is quadratic in μ, the posterior distribution $p(\mu \,|\, \mathbf{y})$ is normal. We next determine its mean and variance.

In this example, the posterior distribution belongs to the same parametric family as the prior—both are normal. More generally, when the prior and posterior distributions belong to the same family, the prior is referred to as a *conjugate prior* for the likelihood.

To derive the mean and variance of the posterior distribution, suppose that

$$(\mu \,|\, \mathbf{y}) \sim \mathcal{N}(\widehat{\mu}, D_\mu)$$

for some posterior mean $\widehat{\mu}$ and variance D_μ. Using the normal density, the

posterior density can be written as

$$\begin{aligned} p(\mu \,|\, \mathbf{y}) &= (2\pi D_\mu)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2D_\mu}(\mu-\widehat{\mu})^2} \\ &\propto \mathrm{e}^{-\frac{1}{2}\left(\frac{1}{D_\mu}\mu^2 - 2\mu\frac{\widehat{\mu}}{D_\mu}\right)}. \end{aligned}$$

Comparing this expression with (1.4), and noting that the two expressions must coincide for all $\mu \in \mathbb{R}$, we obtain

$$D_\mu = \left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}\right)^{-1}.$$

Matching the linear terms in μ yields

$$\begin{aligned} \frac{\widehat{\mu}}{D_\mu} &= \frac{\mu_0}{\sigma_0^2} + \frac{T\bar{y}}{\sigma^2} \\ \widehat{\mu} &= D_\mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{T\bar{y}}{\sigma^2}\right). \end{aligned}$$

Substituting for D_μ, the posterior mean can be written as

$$\widehat{\mu} = \left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}\right)^{-1} \left(\frac{\mu_0}{\sigma_0^2} + \frac{T\bar{y}}{\sigma^2}\right) = \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}}\mu_0 + \frac{\frac{T}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}}\bar{y}.$$

Thus, the posterior mean is a weighted average of the prior mean μ_0 and the sample mean $\bar{y}$, with weights proportional to their respective precisions. As the sample size T increases, a greater weight is placed on $\bar{y}$. Conversely, for a fixed sample size and measurement variance, a more informative prior—that is, a smaller σ_0^2—receives greater weight. In this example, the posterior mean combines a historical benchmark for GDP growth with information from the survey of professional forecasters, with the relative weights determined by the precision of the prior calibration and the cross-sectional dispersion of the survey forecasts.

This weighted-average form has a useful interpretation: the prior acts as a shrinkage device, pulling estimates toward economically meaningful benchmark values when the data are noisy or the sample size is small. In the present example, the posterior mean shrinks the survey-based estimate toward the historical average growth rate, with the strength of this shrinkage governed by the relative precision of the prior and the data. The same principle underlies many Bayesian macroeconometric methods, where priors stabilize estimation in high-dimensional or short-sample settings.

Beyond inference on model parameters, Bayesian analysis also yields predictions for future or unobserved outcomes through the posterior predictive distribution. In the normal model with known variance, the posterior predictive distribution for a future observation can be derived analytically by integrating out posterior uncertainty in μ. The details of this derivation are given in Exercise 1.4.

Example 1.1 (Posterior Updating in a Survey-Forecast Model)**.** As a numerical illustration, suppose the prior mean is set to $\mu_0 = 2$, reflecting a long-run average annualized GDP growth rate of 2 percent, while the sample mean of the survey forecasts is $\bar{y} = 0.5$, indicating more pessimistic expectations for growth next quarter. Suppose further that $\sigma_0^2 = \sigma^2 = 1$, so that the prior variance is equal to the variance of a single survey forecast. Figure 1.1 overlays the prior with a sequence of posterior distributions of μ for $T = 1, 2, 5$, and 25, where successively darker curves correspond to larger sample sizes.

The prior is centered at $\mu_0 = 2$ and is relatively diffuse. As T grows, the posterior shifts toward the sample mean $\bar{y} = 0.5$ and becomes increasingly concentrated, reflecting the growing weight placed on the survey information. For $T = 1$, the posterior mean equals 1.25, exactly halfway between the prior and sample means. For $T = 25$, the posterior is sharply peaked near $\bar{y}$ and the influence of the prior is small.

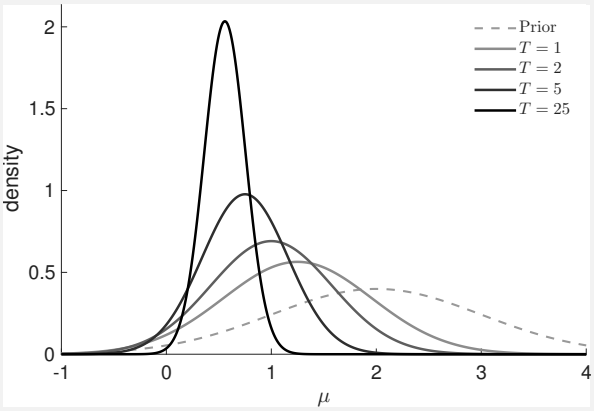

FIGURE 1.1: Prior $p(\mu)$ (dashed) and posterior distributions $p(\mu \,|\, \mathbf{y})$ for $T = 1, 2, 5, 25$, with darker shades corresponding to larger T.

1.2.3 Credible Sets and Monte Carlo Integration

Since $(\mu \,|\, \mathbf{y}) \sim \mathcal{N}(\widehat{\mu}, D_\mu)$, a variety of posterior quantities can be computed directly, including the posterior mean $\mathbb{E}(\mu \,|\, \mathbf{y}) = \widehat{\mu}$, posterior standard deviation $\sqrt{\mathrm{Var}(\mu \,|\, \mathbf{y})} = \sqrt{D_\mu}$, posterior probabilities such as $\mathbb{P}(\mu > 0 \,|\, \mathbf{y})$, and credible sets for μ. A $(1 - \alpha)$ credible set is any subset $C \subseteq \mathbb{R}$ satisfying $\mathbb{P}(\mu \in C \,|\, \mathbf{y}) = 1 - \alpha$. Two common choices are the *equal-tailed* interval, formed by the $\alpha/2$ and $1 - \alpha/2$ posterior quantiles, and the *highest posterior density* region, defined as the smallest such set and equivalently the set of values with posterior density above a threshold. The two coincide when the

posterior is symmetric and unimodal, as in the present normal example, but can differ noticeably when the posterior is skewed.

More generally, interest may lie in the posterior expectation of a function $g(\mu)$, which need not be available in closed form. In particular, consider

$$\mathbb{E}(g(\mu) \,|\, \mathbf{y}) = \int g(\mu)\, p(\mu \,|\, \mathbf{y}) \mathrm{d}\mu.$$

In many cases, this integral cannot be evaluated analytically. However, it can be approximated using *Monte Carlo integration*. Specifically, suppose we generate R draws $\mu^{(1)}, \ldots, \mu^{(R)}$ from the posterior distribution $p(\mu \,|\, \mathbf{y})$. The posterior expectation can then be estimated by

$$\widehat{g} = \frac{1}{R} \sum_{r=1}^{R} g(\mu^{(r)}).$$

By the weak law of large numbers, $\widehat{g}$ converges in probability to $\mathbb{E}(g(\mu) \,|\, \mathbf{y})$ as $R \to \infty$. Because the simulation size R is under the researcher's control, posterior expectations can in principle be approximated to arbitrary accuracy.

Example 1.2 (Monte Carlo Approximation of a Posterior Probability)**.** As an illustration, suppose $g(x) = \mathbf{1}\{x < 0\}$, so that the posterior expectation $\mathbb{E}(g(\mu) \,|\, \mathbf{y})$ corresponds to the posterior probability that GDP growth next quarter is negative: $\mathbb{E}(g(\mu) \,|\, \mathbf{y}) = \mathbb{P}(\mu < 0 \,|\, \mathbf{y})$. This probability provides a simple measure of downside risk that is often of direct interest in macroeconomic analysis.

Suppose further that $(\mu \,|\, \mathbf{y}) \sim \mathcal{N}(\widehat{\mu}, D_\mu)$, with $\widehat{\mu} = 0.64$ and $D_\mu = 0.09$. The following MATLAB script `normal_known_variance.m` implements Monte Carlo integration using $R = 10,000$ draws.

```
R = 10000;
mu_hat = 0.64; Dmu = 0.09;
mu = mu_hat + sqrt(Dmu)*randn(R,1);
g_hat = mean(mu<0);
```

Listing 1.1: Monte Carlo approximation of the posterior probability $\mathbb{P}(\mu < 0 \,|\, \mathbf{y})$ (`normal_known_variance.m`).

The command `randn(R,1)` generates an $R \times 1$ vector of standard normal random variables. Recall that if $Z \sim \mathcal{N}(0,1)$, then $Y = a + bZ$ follows a $\mathcal{N}(a, b^2)$ distribution. Accordingly, the third line generates R draws from the posterior distribution $\mathcal{N}(\widehat{\mu}, D_\mu)$. The fourth line evaluates the indicator function $\mathbf{1}\{\mu < 0\}$ for each draw and computes its sample mean, yielding a Monte Carlo approximation to $\mathbb{P}(\mu < 0 \,|\, \mathbf{y})$, which is

approximately 0.0169. This value is close to the corresponding probability obtained by numerically evaluating the normal distribution function.

1.3 Normal Model with Unknown Variance

We now extend the normal model studied in the previous section by allowing the error variance to be unknown. We first specify the model and prior distributions and derive the joint posterior distribution. We then discuss posterior sampling using Gibbs sampling and illustrate the resulting algorithm with a numerical example.

1.3.1 Model and Joint Posterior

Consider the normal model

$$(y_t \,|\, \mu, \sigma^2) \sim \mathcal{N}(\mu, \sigma^2), \quad t = 1, \ldots, T,$$

where both μ and σ^2 are now unknown. We assume μ and σ^2 are *a priori* independent, so that $p(\mu, \sigma^2) = p(\mu)\, p(\sigma^2)$. For μ, we retain the same prior as in the previous section, $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Since σ^2 takes values on the positive real line, a conventional choice is the inverse-gamma prior. We assume $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$, with density

$$p(\sigma^2) = \frac{S_0^{\nu_0}}{\Gamma(\nu_0)} (\sigma^2)^{-(\nu_0+1)} \mathrm{e}^{-S_0/\sigma^2}, \quad \sigma^2 > 0.$$

The inverse-gamma distribution is widely used as a prior for variance parameters in Gaussian models, in part because it leads to analytically convenient posterior updates.

The inverse-gamma distribution and its properties are discussed in more detail in Appendix A. Multiple parameterizations of the inverse-gamma distribution are common in the literature; throughout this book we adopt the parameterization shown above. When comparing results across sources, it is important to verify the parameterization being used. Under this parameterization, the prior mean $\mathbb{E}(\sigma^2) = S_0/(\nu_0 - 1)$ exists for $\nu_0 > 1$, and smaller values of ν_0 correspond to heavier-tailed, more diffuse prior beliefs about the scale of the measurement error.

Given the likelihood and prior distributions, we can derive the joint posterior distribution of (μ, σ^2). By Bayes' theorem, the joint posterior density is

proportional to the joint density of the parameters and the data:

$$\begin{aligned} p(\mu, \sigma^2 \,|\, \mathbf{y}) &\propto p(\mu, \sigma^2, \mathbf{y}) \\ &\propto p(\mu)p(\sigma^2)p(\mathbf{y} \,|\, \mu, \sigma^2) \\ &\propto \mathrm{e}^{-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2}(\sigma^2)^{-(\nu_0+1)}\mathrm{e}^{-\frac{S_0}{\sigma^2}} \prod_{t=1}^{T}(\sigma^2)^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2\sigma^2}(y_t-\mu)^2}, \end{aligned} \tag{1.5}$$

where proportionality holds up to a normalizing constant that does not depend on (μ, σ^2).

Although an explicit expression for the joint posterior density is available, it is generally not straightforward to compute analytically posterior quantities of interest, such as the posterior mean $\mathbb{E}(\mu \,|\, \mathbf{y})$ or the posterior variance $\mathrm{Var}(\sigma^2 \,|\, \mathbf{y})$. A natural way forward is therefore to rely on Monte Carlo simulation to approximate these quantities.

1.3.2 Introducing the Gibbs Sampler

To compute posterior quantities of interest in this model, we require draws from the joint posterior distribution $p(\mu, \sigma^2 \,|\, \mathbf{y})$. For example, to approximate the posterior variance $\mathrm{Var}(\sigma^2 \,|\, \mathbf{y})$, suppose we obtain R draws from the joint posterior distribution $p(\mu, \sigma^2 \,|\, \mathbf{y})$, denoted by $(\mu^{(1)}, \sigma^{2(1)}), \ldots, (\mu^{(R)}, \sigma^{2(R)})$. An estimator of the posterior variance is then given by

$$\frac{1}{R}\sum_{r=1}^{R}(\sigma^{2(r)} - \bar{\sigma}^2)^2,$$

where $\bar{\sigma}^2$ denotes the sample mean of the simulated draws.

The remaining challenge is therefore computational: how can we generate draws from the posterior distribution? This question motivates Markov chain Monte Carlo (MCMC) methods, a broad class of algorithms designed to sample from complex probability distributions by constructing a Markov chain whose stationary distribution coincides with the target posterior distribution. Foundational contributions include the Metropolis algorithm (Metropolis et al., 1953) and its generalization by Hastings (1970). In Bayesian econometrics, influential early expositions of simulation-based posterior inference include Gelfand and Smith (1990) and Chib and Greenberg (1995). We begin by introducing one such method, the *Gibbs sampler*.

Suppose we wish to sample from a target distribution $p(\boldsymbol{\theta})$, where the parameter vector $\boldsymbol{\theta}$ is partitioned into B blocks $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_B$. A Gibbs sampler constructs a Markov chain $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$ by iteratively sampling each block from its full conditional distribution

$$p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_B), \quad i = 1, \ldots, B.$$

Under mild regularity conditions, the stationary distribution of the resulting Markov chain coincides with the target distribution $p(\boldsymbol{\theta})$.

Operationally, the algorithm proceeds as follows. Starting from an initial value $\boldsymbol{\theta}^{(0)}$, the Gibbs sampler iterates through the conditional distributions to generate a sequence of draws. Algorithm 1.1 summarizes the procedure.

Algorithm 1.1 Gibbs Sampler

Input: Initial value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \ldots, \boldsymbol{\theta}_B^{(0)})$ and number of iterations R.

1. For $r = 0, 1, \ldots, R-1$, repeat the following updates:

   (a) Draw
   $$\boldsymbol{\theta}_1^{(r+1)} \sim p\left(\boldsymbol{\theta}_1 \,|\, \boldsymbol{\theta}_2^{(r)}, \ldots, \boldsymbol{\theta}_B^{(r)}\right).$$

   (b) For $i = 2, \ldots, B-1$, draw
   $$\boldsymbol{\theta}_i^{(r+1)} \sim p\left(\boldsymbol{\theta}_i \,|\, \boldsymbol{\theta}_1^{(r+1)}, \ldots, \boldsymbol{\theta}_{i-1}^{(r+1)}, \boldsymbol{\theta}_{i+1}^{(r)}, \ldots, \boldsymbol{\theta}_B^{(r)}\right).$$

   (c) Draw
   $$\boldsymbol{\theta}_B^{(r+1)} \sim p\left(\boldsymbol{\theta}_B \,|\, \boldsymbol{\theta}_1^{(r+1)}, \ldots, \boldsymbol{\theta}_{B-1}^{(r+1)}\right).$$

2. Return the draws $\{\boldsymbol{\theta}^{(r)}\}_{r=1}^{R}$.

It is important to emphasize that the Markov chain $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$ does not converge to a fixed vector of constants. Rather, as the number of iterations increases, the distribution of $\boldsymbol{\theta}^{(r)}$ converges to the target distribution $p(\boldsymbol{\theta})$.

In practice, the initial draws may be influenced by the starting value $\boldsymbol{\theta}^{(0)}$, typically set to a rough estimate such as the prior mean or a maximum likelihood estimate; ergodicity ensures the chain eventually forgets its initialization. It is therefore common to discard the first R_0 iterations, referred to as the *burn-in* period. A variety of convergence diagnostics have been proposed to assess whether the Markov chain has reached its stationary distribution. One widely used diagnostic is the Geweke convergence test described in Geweke (1992). These diagnostics, together with a more systematic treatment of convergence and Markov chain theory, are discussed in greater detail in Chapter 6.

1.3.3 Gibbs Sampler for the Two-Parameter Normal Model

We now return to the estimation of the two-parameter normal model. To construct a Gibbs sampler for drawing from the joint posterior distribution

$p(\mu, \sigma^2 \,|\, \mathbf{y})$ in (1.5), we require the two full conditional distributions $p(\mu \,|\, \mathbf{y}, \sigma^2)$ and $p(\sigma^2 \,|\, \mathbf{y}, \mu)$.

We begin with the conditional distribution of μ. Conditional on σ^2, the model reduces to the normal model with known variance studied in the previous section. Applying the same derivation as before yields

$$\begin{aligned} p(\mu \,|\, \mathbf{y}, \sigma^2) &\propto p(\mu) p(\mathbf{y} \,|\, \mu, \sigma^2) \\ &\propto \mathrm{e}^{-\frac{1}{2}\left(\left(\frac{1}{\sigma_0^2}+\frac{T}{\sigma^2}\right)\mu^2 - 2\mu\left(\frac{\mu_0}{\sigma_0^2}+\frac{T\bar{y}}{\sigma^2}\right)\right)}. \end{aligned}$$

It follows that

$$(\mu \,|\, \mathbf{y}, \sigma^2) \sim \mathcal{N}(\widehat{\mu}, D_\mu),$$

where

$$D_\mu = \left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2}\right)^{-1}, \quad \widehat{\mu} = D_\mu \left(\frac{\mu_0}{\sigma_0^2} + \frac{T\bar{y}}{\sigma^2}\right).$$

Next, we derive the conditional distribution of σ^2. Combining the prior and likelihood terms that depend on σ^2 gives

$$\begin{aligned} p(\sigma^2 \,|\, \mathbf{y}, \mu) &\propto p(\sigma^2)\, p(\mathbf{y} \,|\, \mu, \sigma^2) \\ &\propto (\sigma^2)^{-(\nu_0+1)} \mathrm{e}^{-\frac{S_0}{\sigma^2}} (\sigma^2)^{-T/2} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-\mu)^2} \\ &\propto (\sigma^2)^{-(\nu_0+T/2+1)} \mathrm{e}^{-\frac{S_0+\sum_{t=1}^{T}(y_t-\mu)^2/2}{\sigma^2}}. \end{aligned}$$

This expression corresponds to an inverse-gamma distribution. Comparing with the inverse-gamma density, we obtain

$$(\sigma^2 \,|\, \mathbf{y}, \mu) \sim \mathcal{IG}\left(\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \mu)^2\right).$$

The resulting Gibbs sampler for the two-parameter normal model proceeds as follows. Choose initial values $\mu^{(0)} = a_0$ and $\sigma^{2(0)} = b_0 > 0$. Then, for $r = 1, \ldots, R$, iterate: (1) Draw $\mu^{(r)} \sim p(\mu \,|\, \mathbf{y}, \sigma^{2(r-1)})$; and (2) Draw $\sigma^{2(r)} \sim p(\sigma^2 \,|\, \mathbf{y}, \mu^{(r)})$.

After discarding the first R_0 draws as burn-in, the remaining draws $\{\mu^{(r)}, \sigma^{2(r)}\}_{r=R_0+1}^{R}$ can be used to approximate posterior quantities of interest. For example, the posterior mean of μ may be estimated by

$$\frac{1}{R - R_0} \sum_{r=R_0+1}^{R} \mu^{(r)}$$

as an estimate of $\mathbb{E}(\mu \,|\, \mathbf{y})$.

This two-step Gibbs sampler is a direct application of Algorithm 1.1 to

the present model. The parameter vector is $\boldsymbol{\theta} = (\mu, \sigma^2)$, and each iteration alternates between sampling from the full conditional distribution of μ given σ^2 and the full conditional distribution of σ^2 given μ. Because both conditional distributions belong to standard families—the normal and inverse-gamma distributions—each Gibbs update can be implemented by direct simulation. This conjugate structure makes the sampler particularly simple and efficient, and it highlights how Gibbs sampling exploits conditional tractability even when the joint posterior distribution is not easily summarized analytically.

In more elaborate macroeconometric models, the same principle applies, although the conditional distributions may no longer have closed-form expressions and may require additional numerical methods. The basic logic, however, remains unchanged and will recur throughout the book.

Example 1.3 (Gibbs Sampling for the Two-Parameter Normal Model)**.** As an illustration, the following MATLAB script `normal_gibbs.m` generates a synthetic dataset of $T = 100$ observations from the two-parameter normal model with $\mu = 3$ and $\sigma^2 = 0.1$. It then implements the Gibbs sampler derived above by alternately sampling from the two conditional distributions, $p(\mu \mid \mathbf{y}, \sigma^2)$ and $p(\sigma^2 \mid \mathbf{y}, \mu)$.

```
% MCMC settings
nsim = 10000;
burnin = 1000;

% simulate data from the normal model
T = 100;
mu = 3;
sig2 = .1;
y = mu + sqrt(sig2)*randn(T,1);

% prior hyperparameters
mu0   = 0;   % prior mean of mu
sig20 = 100; % prior variance of mu
nu0   = 3;   % IG shape parameter for sig2
S0    = 0.5; % IG scale parameter for sig2

% initialize the Markov chain
mu = 0;
sig2 = 1;

store_theta = zeros(nsim,2); % columns: [mu, sig2]

for isim = 1:nsim + burnin
    % sample mu
    Dmu = 1/(1/sig20 + T/sig2);
    mu_hat = Dmu*(mu0/sig20 + sum(y)/sig2);
    mu = mu_hat + sqrt(Dmu)*randn;
```

```
    % sample sig2
    sig2 = 1/gamrnd(nu0+T/2,1/(S0+sum((y-mu).^2)/2));

    % store draws after burn-in
    if isim > burnin
        isave = isim - burnin;
        store_theta(isave,:) = [mu sig2];
    end
end
% posterior means
theta_mean = mean(store_theta)';
```

Listing 1.2: Gibbs sampler for the normal model (`normal_gibbs.m`).

Sampling from the conditional distribution $(\mu \,|\, \mathbf{y}, \sigma^2)$ proceeds exactly as in the known-variance case discussed earlier. Sampling from the conditional distribution $(\sigma^2 \,|\, \mathbf{y}, \mu)$ is implemented by exploiting the relationship between the inverse-gamma and gamma distributions: an inverse-gamma draw can be obtained by taking the reciprocal of a gamma draw; see Exercise 1.7. In the code, this is accomplished using the built-in `gamrnd` function, whose parameterization of the gamma distribution differs from the inverse-gamma parameterization used in this book and therefore requires a simple transformation.

Using 10,000 posterior draws after discarding the first 1,000 iterations as burn-in, the posterior means of μ and σ^2 are estimated to be 2.99 and 0.11, respectively, which are close to the true data generating values. Figure 1.2 displays trace plots of the post-burn-in draws: both chains fluctuate around stable levels with no apparent trend or sticky behavior, consistent with the sampler having reached its stationary distribution.

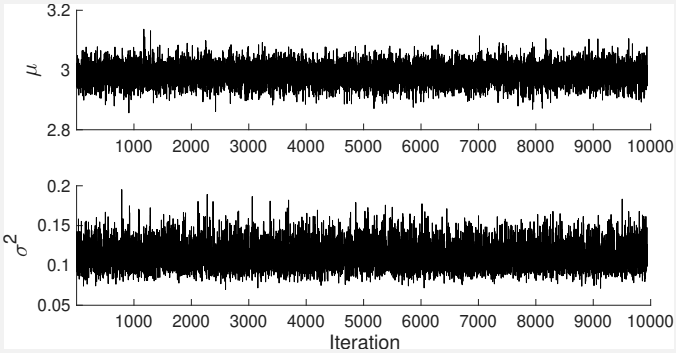

FIGURE 1.2: Trace plots of the post-burn-in draws of μ (top) and σ^2 (bottom).

1.4 Further Reading

Introductory treatments of Bayesian inference can be found in many standard texts. Gelman et al. (2013) provide a broad and influential overview of Bayesian modeling, prior specification, and posterior computation across a wide range of applications. For readers interested in Bayesian econometrics, Koop (2003) and Greenberg (2013) offer accessible introductions that emphasize regression-based models commonly used in applied work. Geweke (2005) provides a more comprehensive, graduate-level treatment that integrates Bayesian inference, simulation methods, and econometric modeling. Additional widely used references include Berger (1985) and Bernardo and Smith (1994) on the foundations of Bayesian inference, and Poirier (1995) and Lancaster (2004) on Bayesian econometrics.

The Gibbs sampler originates with Geman and Geman (1984) in the context of image restoration and was brought into mainstream Bayesian statistics by Gelfand and Smith (1990). Markov chain Monte Carlo methods are surveyed in book-length form by Robert and Casella (2004) and in the handbooks of Brooks et al. (2011) and Kroese, Taimre and Botev (2011). These references develop theoretical foundations, convergence diagnostics, and practical implementation issues. A systematic treatment of Markov chains and MCMC theory is given in Chapter 6.

Exercises

Exercise 1.1 (Geometric Likelihood with Uniform Prior)**.** Suppose $X \sim \mathcal{GEO}(p)$ denotes a geometric distribution with success probability p; see Appendix A for the definition and basic properties. Assume the prior distribution for p is uniform on $(0, 1)$, that is, $p \sim \mathcal{U}(0, 1)$.

(a) Derive the posterior distribution of p.

(b) Find the posterior mode.

(c) Compute the posterior expectation.

Exercise 1.2 (Gamma Likelihood with Gamma Prior)**.** Suppose $x_1 = 1.11$, $x_2 = 0.53$, $x_3 = 11.14$, $x_4 = 0.49$, and $x_5 = 2.47$ constitute an observed sample from a $\mathcal{G}(1, \lambda)$ distribution, where $\mathcal{G}(\alpha, \beta)$ denotes the gamma distribution as defined in Appendix A. Consider Bayesian inference for the rate parameter λ, using the gamma prior $\lambda \sim \mathcal{G}(1, 2)$.

(a) Derive the posterior distribution of λ.

(b) Compute the posterior mean.

Exercise 1.3 (Normal Mean Posterior with Known Variance)**.** In deriving the posterior distribution of μ in the normal model with known variance in Section 1.2, we exploit the computational shortcut of ignoring multiplicative constants that do not depend on μ. In this exercise, you will verify that retaining all constants leads to the same posterior distribution. To that end, first compute the marginal likelihood $p(\mathbf{y}) = \int p(\mu)\, p(\mathbf{y} \,|\, \mu)\, \mathrm{d}\mu$. Then, use Bayes' theorem to show that the posterior distribution $p(\mu \,|\, \mathbf{y})$ is normal with mean $\widehat{\mu}$ and variance D_μ as defined in Section 1.2.

Exercise 1.4 (Normal Posterior Predictive Distribution)**.** Consider the normal model with known variance

$$(y_t \,|\, \mu) \sim \mathcal{N}(\mu, \sigma^2), \quad t = 1, \ldots, T,$$

and prior $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Let y_{T+1} denote a future observation generated from the same model. Using the posterior distribution of μ derived in Section 1.2, compute the posterior predictive distribution $p(y_{T+1} \,|\, \mathbf{y}) = \int p(y_{T+1} \,|\, \mu)\, p(\mu \,|\, \mathbf{y})\, \mathrm{d}\mu$, and show that $(y_{T+1} \,|\, \mathbf{y})$ is normally distributed. State its mean and variance.

Exercise 1.5 (Monte Carlo Estimation of a Moment)**.** Suppose $X \sim \mathcal{N}(2, 3)$ and we wish to estimate $\mathbb{E}(X-1)^4$ using Monte Carlo simulation. Specifically, we generate $R = 10,000$ independent draws $x^{(1)}, \ldots, x^{(R)}$ from the $\mathcal{N}(2, 3)$ distribution and compute $\widehat{m} = R^{-1} \sum_{r=1}^{R} (x^{(r)} - 1)^4$. Give a Monte Carlo estimate of $\mathbb{E}(X-1)^4$. Compare the simulation-based estimate with the analytical value.

Exercise 1.6 (Monte Carlo Standard Error and Simulation Size)**.** Let $\mu^{(1)}, \ldots, \mu^{(R)}$ be independent draws from the posterior distribution, and consider the Monte Carlo estimator $\widehat{g} = R^{-1} \sum_{r=1}^{R} g(\mu^{(r)})$ of the posterior expectation $\mathbb{E}(g(\mu) \,|\, \mathbf{y})$.

(a) Show that $\widehat{g}$ is unbiased and that $\mathrm{Var}(\widehat{g}) = \mathrm{Var}(g(\mu) \,|\, \mathbf{y})/R$, so that the Monte Carlo standard error decreases at the rate $R^{-1/2}$.

(b) Consider the downside-risk probability $\mathbb{P}(\mu < 0 \,|\, \mathbf{y})$ estimated in the Monte Carlo illustration of Section 1.2, for which $g(\mu) = \mathbf{1}\{\mu < 0\}$. Show that $g(\mu)$ is a Bernoulli random variable, so that $\mathrm{Var}(\widehat{g}) = p(1-p)/R$ with $p = \mathbb{P}(\mu < 0 \,|\, \mathbf{y})$. Using $p \approx 0.0165$ and $R = 10,000$, report the Monte Carlo standard error.

(c) Determine the number of draws R required for the Monte Carlo standard error to fall below 10^{-3}.

Exercise 1.7 (Inverse-Gamma as Reciprocal of Gamma)**.** Let $X \sim \mathcal{G}(\alpha, \beta)$ denote a gamma-distributed random variable with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$; see Appendix A for the definition. Define $Z = X^{-1}$. Show that Z follows an inverse-gamma distribution, that is, $Z \sim \mathcal{IG}(\alpha, \beta)$.

Exercise 1.8 (Equal-Tailed and Highest Posterior Density Intervals)**.** The equal-tailed and highest posterior density intervals introduced in Section 1.2 coincide for a symmetric, unimodal posterior but differ when it is skewed. Suppose the posterior of a variance parameter is inverse-gamma, $\sigma^2 \sim \mathcal{IG}(5, 2)$.

(a) Generate $R = 10,000$ draws from the posterior (an inverse-gamma draw is the reciprocal of a gamma draw; see Exercise 1.7). Report the 90% equal-tailed interval from the empirical quantiles, and the 90% highest posterior density interval, computed as the shortest window of $9,000$ consecutive sorted draws.

(b) Verify that the highest posterior density interval is the shorter of the two, and explain how its location reflects the right skewness of the inverse-gamma distribution.

Exercise 1.9 (Gibbs Sampling from a Bivariate Normal)**.** Let $(x, y)'$ follow a bivariate normal distribution with mean vector $\mathbf{0} = (0, 0)'$ and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

(a) Show that the full conditional distributions are

$$(y \,|\, x) \sim \mathcal{N}(\rho x, 1 - \rho^2), \qquad (x \,|\, y) \sim \mathcal{N}(\rho y, 1 - \rho^2).$$

(b) Construct a Gibbs sampler to generate $10,000$ draws from the joint distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Plot the simulated draws for $\rho = 0$, 0.7, and 0.9, and comment on how the dependence between X and Y changes with ρ.

Exercise 1.10 (Gibbs Sampling from a Nonstandard Density)**.** Suppose we wish to generate a random sample from the joint density

$$f(x, y) \propto x \mathrm{e}^{-x - 2y - 10xy}, \quad x > 0,\ y > 0.$$

(a) Derive the full conditional densities $f(x \,|\, y)$ and $f(y \,|\, x)$.

(b) Construct a Gibbs sampler to generate $10,000$ draws from the joint distribution $f(x, y)$. Plot histograms of the marginal draws for x and y.

2

Normal Linear Regression

The normal linear regression model is a fundamental building block in statistics and econometrics. Many widely used nonlinear and multivariate models can be viewed as extensions of this basic framework, so a clear understanding of estimation and inference in the linear model is essential. This chapter derives the likelihood under the normal linear regression, introduces two common prior specifications, and shows how to carry out Bayesian inference using the resulting posterior distributions—either in closed form or via Markov chain Monte Carlo methods. These results provide the foundation for the more sophisticated Bayesian models developed in later chapters.

2.1 Normal Linear Regression and Its Likelihood

We begin by formally defining the normal linear regression model that will be used throughout this chapter.

Definition 2.1 (Normal Linear Regression Model)**.** *A* normal linear regression model *is defined by*

$$y_t = \beta_1 + x_{2t}\beta_2 + \cdots + x_{kt}\beta_k + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $\beta_1, \ldots, \beta_k$ *are regression coefficients,* $x_{2t}, \ldots, x_{kt}$ *are the regressors, and the disturbances* $\varepsilon_1, \ldots, \varepsilon_T$ *are independent and identically distributed as* $\mathcal{N}(0, \sigma^2)$*. Equivalently, in matrix notation,*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_T)'$*,* $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$*,* $\mathbf{X}$ *is the* $T \times k$ *regressor matrix with* t*th row* $(1, x_{2t}, \ldots, x_{kt})$*, and* $(\boldsymbol{\varepsilon} \mid \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T)$*.*

Throughout, the regressor matrix $\mathbf{X}$ is treated as known and is therefore suppressed from the conditioning set to simplify notation. For example, we write the likelihood as $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2)$ rather than $p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2)$, and adopt the same convention for the posterior and predictive densities.

Under the normal linear regression model in Definition 2.1, $\mathbf{y}$ is an affine transformation of $\boldsymbol{\varepsilon}$ and therefore follows a multivariate normal distribution (see, e.g., Theorem 3.6 in Chan and Kroese, 2025). In particular,

$$(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T).$$

Applying the multivariate normal density formula (see Appendix A) yields the likelihood function for the normal linear regression model:

$$p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-T/2} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}. \tag{2.1}$$

This likelihood, together with a prior specification for $(\boldsymbol{\beta}, \sigma^2)$, fully characterizes the Bayesian linear regression model.

To complete the model specification, prior distributions are assigned to the parameters $\boldsymbol{\beta}$ and σ^2. In the normal linear regression model, two prior formulations are particularly common: the natural conjugate prior and the independent normal and inverse-gamma prior.

2.2 Natural Conjugate Prior

This section introduces the natural conjugate prior for the normal linear regression model. This prior is the benchmark in Bayesian regression analysis: it yields closed-form posterior and predictive distributions and is the point of departure for the more flexible prior specifications discussed later.

2.2.1 Prior Specification

We begin by describing the natural conjugate prior for the normal linear regression model and its connection to the normal-inverse-gamma distribution, along with guidance on hyperparameter interpretation. The natural conjugate prior is a joint distribution for $(\boldsymbol{\beta}, \sigma^2)$ and is most conveniently expressed in terms of the normal-inverse-gamma distribution.

The normal-inverse-gamma distribution can be defined hierarchically as follows. Let $\mathbf{w}$ be an $n \times 1$ random vector and let s be a positive random variable. Conditional on s, $\mathbf{w}$ is normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $s\boldsymbol{\Sigma}$, and marginally s follows an inverse-gamma distribution $\mathcal{IG}(\alpha, \beta)$. We write $(\mathbf{w}, s) \sim \mathcal{NIG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta)$. The corresponding joint density is given in Appendix A.

In the context of the normal linear regression model, the natural conjugate

prior for $(\boldsymbol{\beta}, \sigma^2)$ is obtained by specifying $(\boldsymbol{\beta} \,|\, \sigma^2) \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$, or equivalently, $(\boldsymbol{\beta}, \sigma^2) \sim \mathcal{NIG}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}, \nu_0, S_0)$.

The normal-inverse-gamma prior is parameterized by four hyperparameters, namely $\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}, \nu_0$, and S_0. Below we describe a common approach to eliciting these hyperparameters.

We begin by choosing a small value for ν_0 (for example, $\nu_0 = 3$), which yields a large prior variance for σ^2 and thus a relatively diffuse prior on the error variance. Conditional on ν_0, the scale parameter S_0 is then selected to match a desired prior mean of σ^2 via $\mathbb{E}(\sigma^2) = S_0/(\nu_0 - 1)$ for $\nu_0 > 1$, or equivalently $S_0 = (\nu_0 - 1)\mathbb{E}(\sigma^2)$.

Next, the prior mean $\boldsymbol{\beta}_0$ is specified to reflect prior beliefs about the regression coefficients. In many applications, setting $\boldsymbol{\beta}_0 = \mathbf{0}_k$ provides a reasonable default choice. Finally, the matrix $\mathbf{V}_{\boldsymbol{\beta}}$ governs the prior dispersion of $\boldsymbol{\beta}$ relative to the error variance σ^2. Since the ith diagonal element of $\mathbf{V}_{\boldsymbol{\beta}}$ can be interpreted as the prior variance of β_i scaled by σ^2, $\mathbf{V}_{\boldsymbol{\beta}}$ is often taken to be diagonal with large diagonal elements, reflecting weak prior information about individual coefficients.

The hyperparameters need not be fixed by judgment alone. They can be assigned their own priors and estimated jointly with the model parameters, leading to the hierarchical priors of Chapter 7; they can be chosen to maximize the marginal likelihood, as done for the shrinkage hyperparameters of the vector autoregressions in Chapters 12 and 14; or they can be selected on predictive grounds, using cross-validation-based criteria of the kind discussed in Chapter 5.

2.2.2 Posterior Analysis under Conjugacy

The normal-inverse-gamma prior leads to closed-form posterior updating in the normal linear regression model. The following theorem characterizes the resulting joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$ and provides the basis for both analytical inference and efficient posterior sampling.

Theorem 2.1 (Normal-Inverse-Gamma Conjugacy)**.** Consider the normal linear regression model with likelihood given in (2.1) and suppose the prior distribution for $(\boldsymbol{\beta}, \sigma^2)$ is normal-inverse-gamma,

$$(\boldsymbol{\beta}, \sigma^2) \sim \mathcal{NIG}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}, \nu_0, S_0).$$

Then the posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$ is also normal-inverse-gamma,

$$(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S}),$$

where $\widehat{\nu} = \nu_0 + T/2$,

$$\mathbf{D}_{\boldsymbol{\beta}} = (\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{y}),$$

and

$$\widehat{S} = S_0 + \tfrac{1}{2}\left(\mathbf{y}'\mathbf{y} + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\widehat{\boldsymbol{\beta}}\right).$$

The derivation below is algebraically involved but worth working through carefully: the conjugate results it establishes are used repeatedly in later chapters—in model comparison, in efficient computation for high-dimensional settings, and in trans-dimensional algorithms for exploring model space.

Proof. To verify conjugacy, we derive the posterior distribution by combining the likelihood and prior kernels and completing the square in $\boldsymbol{\beta}$ (see Appendix B.4). Although the algebra is involved, the derivation highlights the key mechanism underlying conjugacy in the linear regression model—namely, the quadratic structure induced by Gaussian assumptions.

Step 1: Combine the likelihood and prior kernels. Combining the normal likelihood in (2.1) with the normal-inverse-gamma prior gives the unnormalized posterior density

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) &\propto p(\boldsymbol{\beta}, \sigma^2)\, p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) \\ &\propto (\sigma^2)^{-(\nu_0+1+k/2)} \mathrm{e}^{-\frac{1}{\sigma^2}\left[S_0 + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right]} \\ &\quad \times (\sigma^2)^{-T/2} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \\ &\propto (\sigma^2)^{-(\widehat{\nu}+1+k/2)} \mathrm{e}^{-\frac{1}{\sigma^2}\left[S_0 + \frac{1}{2}Q(\boldsymbol{\beta})\right]}, \end{aligned} \tag{2.2}$$

where $\widehat{\nu} = \nu_0 + T/2$ and

$$Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

To establish conjugacy, it suffices to show that $Q(\boldsymbol{\beta})$ can be written as a single quadratic form centered at a posterior mean.

Step 2: Expand and collect quadratic terms in $\boldsymbol{\beta}$. First expand the likelihood quadratic form:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \mathbf{y}'\mathbf{y}, \tag{2.3}$$

where we use the fact that $\mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ is a scalar and hence equals its transpose. Next expand the prior quadratic form:

$$(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = \boldsymbol{\beta}'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0. \tag{2.4}$$

Collecting terms yields

$$Q(\boldsymbol{\beta}) = \boldsymbol{\beta}'(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})\boldsymbol{\beta} - 2\boldsymbol{\beta}'(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{y}) + \mathbf{y}'\mathbf{y} + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0. \quad (2.5)$$

Step 3: Complete the square. Define

$$\mathbf{D}_{\boldsymbol{\beta}} = (\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{y}).$$

Adding and subtracting $\widehat{\boldsymbol{\beta}}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\widehat{\boldsymbol{\beta}}$ in (2.5) gives

$$Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'\mathbf{D}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \mathbf{y}'\mathbf{y} + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\widehat{\boldsymbol{\beta}}. \quad (2.6)$$

Substituting (2.6) into (2.2) yields

$$p(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \propto (\sigma^2)^{-(\widehat{\nu}+1+k/2)} \mathrm{e}^{-\frac{1}{\sigma^2}\left[\widehat{S} + \frac{1}{2}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\mathbf{D}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})\right]},$$

where

$$\widehat{S} = S_0 + \tfrac{1}{2}\left(\mathbf{y}'\mathbf{y} + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\widehat{\boldsymbol{\beta}}\right).$$

Comparing this kernel with the normal-inverse-gamma density in Appendix A establishes that

$$(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S}),$$

which completes the proof. □

The conjugacy result provides explicit expressions for the posterior hyperparameters $(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S})$. Each has a natural interpretation in terms of how prior information and sample evidence are combined.

The vector $\widehat{\boldsymbol{\beta}}$ is the posterior mean of the regression coefficients. When $\mathbf{X}'\mathbf{X}$ is invertible, it can be interpreted as a precision-weighted average of the prior mean $\boldsymbol{\beta}_0$ and the ordinary least squares (OLS) estimator $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, with weights determined by the prior precision $\mathbf{V}_{\boldsymbol{\beta}}^{-1}$ and the data precision $\mathbf{X}'\mathbf{X}$. As the sample size increases, the likelihood becomes increasingly informative and $\widehat{\boldsymbol{\beta}}$ converges to the least squares estimator. In contrast, in small samples or when the regressors are weakly informative, the prior exerts a more substantial influence on $\widehat{\boldsymbol{\beta}}$. If $\mathbf{X}'\mathbf{X}$ is singular, the least squares estimator is not uniquely defined, but the posterior mean $\widehat{\boldsymbol{\beta}}$ remains well defined provided $\mathbf{V}_{\boldsymbol{\beta}}$ is positive definite. In this case, the prior regularizes the problem and ensures a proper posterior distribution.

The matrix $\mathbf{D}_{\boldsymbol{\beta}} = (\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$ governs the posterior dispersion of the regression coefficients conditional on σ^2: the conditional posterior covariance of $\boldsymbol{\beta}$ is $\sigma^2\mathbf{D}_{\boldsymbol{\beta}}$, while the marginal posterior covariance, obtained after integrating out σ^2, differs from $\mathbf{D}_{\boldsymbol{\beta}}$. It can be interpreted as the posterior analogue of a covariance matrix scaled by the error variance, reflecting both prior uncertainty and the information content of the regressors. Multicollinearity in $\mathbf{X}$

or weak prior precision translates directly into increased posterior uncertainty through this matrix.

The scalar hyperparameter $\widehat{\nu} = \nu_0 + T/2$ is an effective degrees-of-freedom parameter for the error variance. Each additional observation contributes one-half degree of freedom, so $\widehat{\nu}$ grows linearly with the sample size, leading to increasing concentration of the posterior distribution of σ^2.

Finally, the scale parameter $\widehat{S}$ combines the prior scale S_0 with information from the data. It can be equivalently expressed as (see Exercise 2.2):

$$\widehat{S} = S_0 + \tfrac{1}{2}\left[(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right],$$

which shows that $\widehat{S}$ depends both on the overall fit of the model, as measured by the residual sum of squares, and on the discrepancy between the prior mean $\boldsymbol{\beta}_0$ and the posterior mean $\widehat{\boldsymbol{\beta}}$. The second term makes explicit how prior misspecification is penalized in the posterior distribution of σ^2.

The conjugacy result in Theorem 2.1 immediately delivers several useful consequences. Because the posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$ is available in closed form, one can integrate out components of the parameter vector analytically. In particular, marginalizing over σ^2 yields Student-t distributions for both the regression coefficients and future observations, while integrating over all parameters produces a closed-form expression for the marginal likelihood. These results are summarized in the following corollaries.

Corollary 2.1 (Marginal Posterior Distribution of $\boldsymbol{\beta}$)**.** Suppose that the joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$ is

$$(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S}).$$

Then the marginal posterior distribution of $\boldsymbol{\beta}$ is multivariate Student-t,

$$(\boldsymbol{\beta} \,|\, \mathbf{y}) \sim \mathcal{T}_{2\widehat{\nu}}\left(\widehat{\boldsymbol{\beta}}, \frac{\widehat{S}}{\widehat{\nu}}\mathbf{D}_{\boldsymbol{\beta}}\right).$$

The proof of Corollary 2.1 is left as an exercise. It follows by integrating out σ^2 from the joint normal-inverse-gamma posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$.

Corollary 2.2 (Marginal Likelihood under the Natural Conjugate Prior)**.** Consider the normal linear regression model with likelihood given in (2.1) and suppose the prior distribution for $(\boldsymbol{\beta}, \sigma^2)$ is normal-inverse-gamma,

$$(\boldsymbol{\beta}, \sigma^2) \sim \mathcal{NIG}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}, \nu_0, S_0).$$

Then the marginal likelihood admits the closed-form expression

$$p(\mathbf{y}) = (2\pi)^{-T/2}|\mathbf{V}_{\boldsymbol{\beta}}|^{-1/2}|\mathbf{D}_{\boldsymbol{\beta}}|^{1/2}\frac{\Gamma(\widehat{\nu})}{\Gamma(\nu_0)}\frac{S_0^{\nu_0}}{\widehat{S}^{\widehat{\nu}}}, \tag{2.7}$$

where $\widehat{\nu}$, $\mathbf{D}_{\boldsymbol{\beta}}$, and $\widehat{S}$ are given in Theorem 2.1.

A proof of Corollary 2.2 is left as an exercise. The key idea is to combine Theorem 2.1 with Bayes' theorem to express the marginal likelihood as a ratio of normalizing constants.

Corollary 2.3 (Posterior Predictive Distribution)**.** Consider the normal linear regression model with likelihood given in (2.1) and prior

$$(\boldsymbol{\beta}, \sigma^2) \sim \mathcal{NIG}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}, \nu_0, S_0).$$

Let $(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S})$, where $(\widehat{\nu}, \widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{S})$ are given in Theorem 2.1. For a new regressor vector $\mathbf{x}_{T+1}$, the posterior predictive distribution of y_{T+1} is Student-t,

$$(y_{T+1} \,|\, \mathbf{y}, \mathbf{x}_{T+1}) \sim \mathcal{T}_{2\widehat{\nu}}\left(\mathbf{x}_{T+1}'\widehat{\boldsymbol{\beta}}, \frac{\widehat{S}}{\widehat{\nu}}\Big(1 + \mathbf{x}_{T+1}'\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+1}\Big)\right).$$

Proof. By Theorem 2.1, the joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$ is normal-inverse-gamma,

$$(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S}).$$

It follows immediately from the hierarchical representation of the normal-inverse-gamma distribution that the conditional posterior distributions are

$$(\boldsymbol{\beta} \,|\, \sigma^2, \mathbf{y}) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \sigma^2\mathbf{D}_{\boldsymbol{\beta}}), \quad (\sigma^2 \,|\, \mathbf{y}) \sim \mathcal{IG}(\widehat{\nu}, \widehat{S}).$$

From the regression model,

$$y_{T+1} = \mathbf{x}_{T+1}'\boldsymbol{\beta} + \varepsilon_{T+1}, \quad (\varepsilon_{T+1} \,|\, \sigma^2) \sim \mathcal{N}(0, \sigma^2),$$

and hence, conditional on $(\sigma^2, \mathbf{y})$, y_{T+1} is a linear combination of normal random variables. Marginalizing over $(\boldsymbol{\beta} \,|\, \sigma^2, \mathbf{y})$ yields

$$(y_{T+1} \,|\, \sigma^2, \mathbf{y}) \sim \mathcal{N}\left(\mathbf{x}_{T+1}'\widehat{\boldsymbol{\beta}}, \sigma^2\big(1 + \mathbf{x}_{T+1}'\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+1}\big)\right).$$

Combining this conditional normal distribution with $(\sigma^2 \,|\, \mathbf{y}) \sim \mathcal{IG}(\widehat{\nu}, \widehat{S})$ shows that

$$(y_{T+1}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}\left(\mathbf{x}_{T+1}'\widehat{\boldsymbol{\beta}}, 1 + \mathbf{x}_{T+1}'\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+1}, \widehat{\nu}, \widehat{S}\right).$$

Applying Corollary 2.1 then gives the stated Student-t posterior predictive distribution,

$$(y_{T+1} \,|\, \mathbf{y}, \mathbf{x}_{T+1}) \sim \mathcal{T}_{2\widehat{\nu}} \left(\mathbf{x}'_{T+1}\widehat{\boldsymbol{\beta}}, \frac{\widehat{S}}{\widehat{\nu}}\big(1 + \mathbf{x}'_{T+1}\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+1}\big) \right).$$

□

2.2.3 Application: Forecasting PCE Inflation

To illustrate how conjugate Bayesian regression extends to time-series settings, we consider a simple autoregressive model written in normal linear regression form with lagged regressors. Using quarterly US PCE inflation over the sample period 1960Q1–2019Q4, we estimate an AR(2) model under a normal-inverse-gamma prior and construct a one-step-ahead posterior predictive distribution for inflation in 2020Q1.

Definition 2.2 (Autoregressive Model of Order p)**.** *A univariate time series* $\{y_t\}_{t=1}^{T}$ *follows an* autoregressive model *of order p, denoted AR(p), if*

$$y_t = \beta_1 + \beta_2 y_{t-1} + \cdots + \beta_{p+1} y_{t-p} + \varepsilon_t, \quad (\varepsilon_t \,|\, \sigma^2) \sim \mathcal{N}(0, \sigma^2),$$

for $t = 1, \ldots, T$, where the presample observations $y_0, \ldots, y_{1-p}$ are treated as fixed initial conditions.

The AR(p) model can be written in the normal linear regression form of Definition 2.1 by constructing the regressor matrix from lagged values of y_t. For simplicity, we set $p = 2$ (and hence $k = 3$) and adopt the normal-inverse-gamma prior for $(\boldsymbol{\beta}, \sigma^2)$.

The hyperparameters are chosen so that the normal-inverse-gamma prior is weakly informative. The prior mean of the regression coefficients is set to $\boldsymbol{\beta}_0 = \mathbf{0}_k$, while the prior covariance matrix is $\mathbf{V}_{\boldsymbol{\beta}} = 100 \cdot \mathbf{I}_k$, implying that the prior variances of the coefficients are large relative to the error variance σ^2. For the error variance, we set $\nu_0 = 4$ and $S_0 = 1$, which results in an inverse-gamma prior with finite mean $\mathbb{E}(\sigma^2) = S_0/(\nu_0 - 1) = 1/3$ and relatively large dispersion.

Using the posterior distribution implied by conjugacy, we compute the one-step-ahead posterior predictive density for US PCE inflation in 2020Q1. The script `linreg_NIG_predictive.m` loads the dataset `USPCE.csv` and implements the closed-form posterior and predictive calculations implied by the normal-inverse-gamma prior for the linear regression model.

We make a few comments on the implementation. First, the autoregressive model is cast explicitly as a normal linear regression by constructing the

regressor matrix from lagged inflation, with the first four observations treated as presample values. Second, all posterior quantities, including the predictive density, are computed analytically using closed-form expressions implied by conjugacy, with no reliance on Monte Carlo methods. Third, the predictive distribution is Student-t rather than Gaussian, reflecting uncertainty about both the regression coefficients and the error variance. This feature is evident in the scale term $1+\mathbf{x}'_{T+1}\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+1}$, which inflates the predictive variance relative to a plug-in forecast that conditions on point estimates of the parameters.

```
% load data
data = readmatrix('USPCE.csv', 'Range', 'B2:B241');
y0 = data(1:4);  % initial conditions: [y_{-3},...,y_0]
y = data(5:end); % sample used for estimation
T = size(y,1);

% regressors for AR(2): [1,y_{t-1},y_{t-2}]
xlag1 = [y0(4); y(1:end-1)];
xlag2 = [y0(3:4); y(1:end-2)];
X = [ones(T,1), xlag1, xlag2];
k = size(X,2);   % number of regressors

% prior hyperparameters (normal-inverse-gamma)
beta0  = zeros(k,1);   % prior mean
iVbeta = speye(k)/100; % prior precision
nu0    = 4;            % prior degrees of freedom
S0     = 1;            % prior scale

% Posterior hyperparameters
Dbeta = (iVbeta + X'*X)\speye(k); % inverse computed via \
beta_hat = Dbeta*(iVbeta*beta0 + X'*y);
S_hat = S0 + (y'*y + beta0'*iVbeta*beta0 ...
    - beta_hat'*(Dbeta\beta_hat))/2;
nu_hat = nu0 + T/2;

% Posterior predictive density for y_{T+1}
xTp1  = [1; y(end); y(end-1)]; % regressor vector at T+1
ygrid = linspace(-3,8,500)';   % evaluation grid
fy = tpdfLS(ygrid, xTp1'*beta_hat, ...
    S_hat/nu_hat*(1 + xTp1'*Dbeta*xTp1),2*nu_hat);
```

Listing 2.1: Posterior predictive density under the normal-inverse-gamma prior (`linreg_NIG_predictive.m`).

In the implementation, the posterior covariance matrix $\mathbf{D}_{\boldsymbol{\beta}} = (\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$ is computed using the backslash operator in MATLAB, which avoids explicit matrix inversion and improves numerical stability. Specifically, the expression `(iVbeta + X'*X)\speye(k)` computes $\mathbf{D}_{\boldsymbol{\beta}}$ by solving

$$(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})\mathbf{D}_{\boldsymbol{\beta}} = \mathbf{I}_k,$$

where `speye(k)` denotes the $k \times k$ sparse identity matrix.

To evaluate the posterior predictive density, we use a small utility function `tpdfLS.m` that computes the univariate Student-t density in location-scale form, evaluating the predictive density implied by Corollary 2.3 on a grid of candidate values. The implementation avoids MATLAB's built-in `tpdf`, which is defined for standardized Student-t variables. The function is available on the companion website.

The posterior means of the first- and second-lag autoregressive coefficients are 0.673 and 0.186, respectively, indicating a fairly persistent inflation process. Figure 2.1 displays the posterior predictive density for 2020Q1. The posterior predictive mean is $\mathbf{x}_{T+1}'\widehat{\boldsymbol{\beta}} = 1.69$. The realized inflation rate of 1.39 lies at the 41.5th percentile of the posterior predictive distribution. This outcome falls well within the central region of the predictive distribution and therefore does not represent an unusually large forecast error. In particular, the predictive density assigns substantial probability mass to values close to the realized outcome.

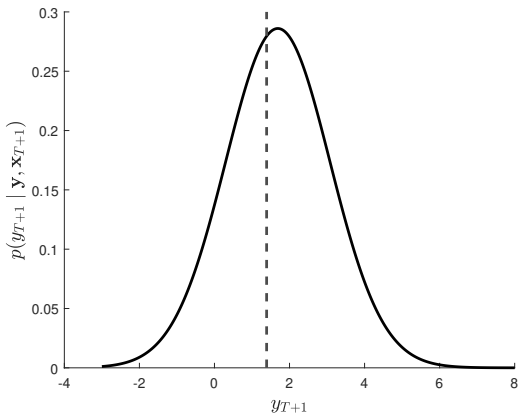

FIGURE 2.1: Posterior predictive density for US PCE inflation in 2020Q1 under an AR(2) model with a normal-inverse-gamma prior. The solid line shows the Student-t predictive density, while the dashed vertical line marks the realized inflation rate $y_{T+1} = 1.39$.

2.3 Independent Normal and Inverse-Gamma Prior

The second common joint prior for $(\boldsymbol{\beta}, \sigma^2)$ is often called the independent normal and inverse-gamma prior, reflecting the assumption of prior independence

between $\boldsymbol{\beta}$ and σ^2:

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta})p(\sigma^2).$$

Specifically, we assume $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$ with prior densities

$$p(\boldsymbol{\beta}) = (2\pi)^{-\frac{k}{2}} |\mathbf{V}_{\boldsymbol{\beta}}|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)' \mathbf{V}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)}, \tag{2.8}$$

$$p(\sigma^2) = \frac{S_0^{\nu_0}}{\Gamma(\nu_0)} (\sigma^2)^{-(\nu_0+1)} \mathrm{e}^{-\frac{S_0}{\sigma^2}}. \tag{2.9}$$

Unlike the natural conjugate prior, this specification does not give a posterior distribution of standard form. As a result, posterior inference is carried out using simulation methods, most commonly a Gibbs sampler. Despite this loss of analytical tractability, the independent prior is often preferred in more elaborate models, such as those discussed in Chapter 3, because it decouples prior beliefs about the regression coefficients and the error variance and integrates more naturally into hierarchical and shrinkage frameworks.

2.3.1 Gibbs Sampler for the Independent Prior

We now derive a Gibbs sampler for the normal linear regression model with likelihood given in (2.1) and independent prior distributions in (2.8)-(2.9). The sampler is based on alternately drawing from the two full conditional distributions $p(\boldsymbol{\beta} \,|\, \mathbf{y}, \sigma^2)$ and $p(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta})$, both of which take standard forms.

First, we derive the conditional density of $\boldsymbol{\beta}$ given $\mathbf{y}$ and σ^2. Ignoring terms that do not depend on $\boldsymbol{\beta}$, we have

$$\begin{aligned} p(\boldsymbol{\beta} \,|\, \mathbf{y}, \sigma^2) &\propto p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) p(\boldsymbol{\beta}) p(\sigma^2) \\ &\propto \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \times \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)' \mathbf{V}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta}-\boldsymbol{\beta}_0)} \end{aligned}$$

Expanding the quadratic forms as in (2.3)-(2.4) yields

$$\begin{aligned} p(\boldsymbol{\beta} \,|\, \mathbf{y}, \sigma^2) &\propto \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\beta}' \mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}_0)} \times \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}' \mathbf{X}'\mathbf{X} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \mathbf{X}'\mathbf{y})} \\ &= \mathrm{e}^{-\frac{1}{2}\left[\boldsymbol{\beta}' \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}' \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right)\right]}. \end{aligned} \tag{2.10}$$

Because the exponent is quadratic in $\boldsymbol{\beta}$, the conditional posterior distribution is multivariate normal:

$$(\boldsymbol{\beta} \,|\, \mathbf{y}, \sigma^2) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}})$$

for some mean vector $\widehat{\boldsymbol{\beta}}$ and covariance matrix $\mathbf{D}_{\boldsymbol{\beta}}$.

One can obtain explicit expressions for $\widehat{\boldsymbol{\beta}}$ and $\mathbf{D}_{\boldsymbol{\beta}}$ by completing the square in (2.10). Instead of repeating that algebra, we adopt a more direct and computationally convenient approach below.

To this end, recall that the kernel of a multivariate normal distribution $\mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}})$ is

$$\mathrm{e}^{-\frac{1}{2}\left(\boldsymbol{\beta}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{D}_{\boldsymbol{\beta}}^{-1}\widehat{\boldsymbol{\beta}}\right)}.$$

This expression must coincide with the kernel in (2.10) for all $\boldsymbol{\beta} \in \mathbb{R}^k$. Equating the quadratic and linear terms in $\boldsymbol{\beta}$ therefore yields

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}\left(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right).$$

Next, using (2.1) and (2.9), we derive the conditional density $p(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta})$. Since $p(\boldsymbol{\beta})$ does not depend on σ^2, it can be ignored in the conditional kernel:

$$\begin{aligned} p(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}) &\propto p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2)\, p(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \times (\sigma^2)^{-(\nu_0+1)} \mathrm{e}^{-\frac{S_0}{\sigma^2}} \\ &= (\sigma^2)^{-\left(\frac{T}{2}+\nu_0+1\right)} \mathrm{e}^{-\frac{1}{\sigma^2}\left(S_0 + \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)}. \end{aligned}$$

This is the kernel of an inverse-gamma density. Hence,

$$(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}) \sim \mathcal{IG}\left(\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

Having derived the two full conditional distributions required for the Gibbs sampler, we now briefly comment on their practical implementation.

Although one may rely on built-in routines such as `mvnrnd` in MATLAB to sample from a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, it is useful to recall how such draws can be generated directly from independent standard normal variables, as summarized in Algorithm 2.1. This construction clarifies the mechanics of multivariate simulation and will be employed in later chapters when custom sampling steps are required. The Cholesky factorization on which it relies is reviewed in Appendix B.3.

Algorithm 2.1 Sampling from a Multivariate Normal Distribution

Input: Mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$, covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, number of draws R.

1. Compute the Cholesky factorization $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$, where $\mathbf{C}$ is lower triangular.
2. For $r = 1, \ldots, R$, repeat:
   (a) Draw $\mathbf{z}^{(r)} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$.
   (b) Set $\mathbf{u}^{(r)} = \boldsymbol{\mu} + \mathbf{C}\mathbf{z}^{(r)}$.
3. Return the draws $\{\mathbf{u}^{(r)}\}_{r=1}^{R}$.

The dominant computational cost in Algorithm 2.1 is the Cholesky factorization of the covariance matrix $\mathbf{\Sigma}$, which requires $\mathcal{O}(n^3)$ operations for a dense $n \times n$ matrix. This factorization is performed once, after which each draw requires $\mathcal{O}(n^2)$ operations to compute $\mathbf{C}\mathbf{z}^{(r)}$. When many draws are needed, the one-time cost of the Cholesky decomposition is amortized across iterations.

Algorithm 2.1 also presumes direct access to the covariance matrix $\mathbf{\Sigma}$. In many regression and time-series models, however, the natural quantity is the *precision matrix* $\mathbf{\Sigma}^{-1}$, which arises directly from Gaussian likelihoods and prior specifications. Obtaining $\mathbf{\Sigma}$ therefore implicitly requires solving a linear system or forming an inverse of the precision matrix, an operation that is costly and can be numerically unstable.

In the Gibbs sampler for the linear regression model, $\mathbf{\Sigma}$ corresponds to the conditional covariance matrix $\mathbf{D}_{\boldsymbol{\beta}}$, whose dimension equals the number of regression coefficients. For moderate k, this step is inexpensive. In high-dimensional regressions, however, forming $\mathbf{D}_{\boldsymbol{\beta}}$ and computing its Cholesky factorization can become a computational bottleneck. In such settings, it is more efficient to work directly with the precision matrix $\mathbf{D}_{\boldsymbol{\beta}}^{-1}$, exploiting structural features such as sparsity and bandedness. Such precision-based samplers reappear in later chapters, in the context of high-dimensional regression and state space models.

Applying the general Gibbs sampler described in Algorithm 1.1 in Chapter 1, the Gibbs sampler for the linear regression model with the independent normal and inverse-gamma prior proceeds as follows. Choose initial values $\boldsymbol{\beta}^{(0)} = \mathbf{a}_0$ and $\sigma^{2(0)} = b_0 > 0$. Then, for $r = 1, \ldots, R$, iterate:

1. Draw $\boldsymbol{\beta}^{(r)} \sim p(\boldsymbol{\beta} \,|\, \mathbf{y}, \sigma^{2(r-1)})$ (multivariate normal), using Algorithm 2.1.
2. Draw $\sigma^{2(r)} \sim p(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}^{(r)})$ (inverse-gamma).

2.3.2 Application: Forecasting PCE Inflation with Independent Prior

To illustrate the linear regression under the independent normal and inverse-gamma prior, we revisit the PCE inflation forecasting exercise of Section 2.2.3. The data, autoregressive specification, and forecasting objective are unchanged, allowing a direct comparison between analytical inference under the natural conjugate prior and simulation-based inference under the independent prior. With comparable diffuse hyperparameters, simulation-based inference under the independent prior should closely match the closed-form conjugate results. The exercise therefore serves mainly to validate the Gibbs sampler before applying it in settings where closed-form posteriors and predictive distributions are unavailable.

Specifically, quarterly US PCE inflation over the sample period 1960Q1–2019Q4 is modeled using an AR(2) specification,

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t, \quad (\varepsilon_t \,|\, \sigma^2) \sim \mathcal{N}(0, \sigma^2),$$

which can be written in the normal linear regression form of Definition 2.1. We adopt independent prior distributions, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$, with hyperparameters chosen to align as closely as possible with those used in the natural conjugate prior case.

In particular, the prior mean is set to $\boldsymbol{\beta}_0 = \mathbf{0}_k$ and the prior covariance matrix is taken to be $\mathbf{V}_{\boldsymbol{\beta}} = 100 \cdot \mathbf{I}_k$, reflecting weak prior information about the regression coefficients. Unlike the natural conjugate prior, however, $\mathbf{V}_{\boldsymbol{\beta}}$ here represents the *unconditional* prior covariance of $\boldsymbol{\beta}$ and does not scale with the error variance σ^2. As a result, the same numerical choice of $\mathbf{V}_{\boldsymbol{\beta}}$ has a different interpretation than in the normal-inverse-gamma prior, where $\mathbf{V}_{\boldsymbol{\beta}}$ governs dispersion relative to σ^2. Finally, we set $\nu_0 = 4$ and $S_0 = 1$, matching the inverse-gamma prior used in the conjugate analysis.

In contrast to the normal-inverse-gamma prior, this specification breaks joint conjugacy. Nevertheless, the conditional posterior distributions remain standard: conditional on σ^2, the regression coefficients follow a multivariate normal distribution, while conditional on $\boldsymbol{\beta}$, the error variance follows an inverse-gamma distribution. Posterior inference can therefore be conducted using a Gibbs sampler, as derived in the last section. After discarding an initial burn-in period, the resulting draws $\left\{(\boldsymbol{\beta}^{(r)}, \sigma^{2(r)})\right\}_{r=1}^{R}$ provide a Monte Carlo approximation to the joint posterior distribution.

The one-step-ahead posterior predictive density can be approximated by evaluating the conditional predictive density on a grid and averaging over posterior draws. Specifically, given each draw $(\boldsymbol{\beta}^{(r)}, \sigma^{2(r)})$, the conditional predictive density is Gaussian

$$p(y_{T+1} \,|\, \boldsymbol{\beta}^{(r)}, \sigma^{2(r)}, \mathbf{y}) = f_{\mathcal{N}}\left(y_{T+1};\, \mathbf{x}_{T+1}'\boldsymbol{\beta}^{(r)}, \sigma^{2(r)}\right),$$

so that

$$p(y_{T+1} \,|\, \mathbf{y}) \approx \frac{1}{R}\sum_{r=1}^{R} f_{\mathcal{N}}\left(y_{T+1};\, \mathbf{x}_{T+1}'\boldsymbol{\beta}^{(r)}, \sigma^{2(r)}\right).$$

Evaluating this expression on a grid yields a smooth approximation to the posterior predictive density without requiring kernel smoothing of simulated outcomes.

The script `linreg_indep_predictive.m` implements the Gibbs sampler described in Section 2.3.1, drawing iteratively from the conditional posterior distributions of $\boldsymbol{\beta}$ and σ^2 and constructing the predictive density by averaging conditional Gaussian forecasts across posterior draws.

```
nsim = 20000; burnin = 1000;

% load data
data = readmatrix('USPCE.csv', 'Range', 'B2:B241');
y0 = data(1:4);  % initial conditions: [y_{-3},y_{-2},y_{-1},y_0]
y = data(5:end); % sample used for estimation
T = size(y,1);

% regressors for AR(2): [1, y_{t-1}, y_{t-2}]
xlag1 = [y0(4); y(1:end-1)];
xlag2 = [y0(3:4); y(1:end-2)];
X = [ones(T,1), xlag1, xlag2];
k = size(X,2);   % number of regressors

% prior hyperparameters (indep normal and inverse-gamma)
beta0  = zeros(k,1);
iVbeta  = 1/100*speye(k);  % prior precision of beta
nu0    = 4;
S0     = 1;

% one-step-ahead regressor vector (T+1)
xTp1 = [1; y(end); y(end-1)];
y_obs = 1.39;
ygrid = linspace(-3,8,300)';

% initialize chain
beta = (X'*X)\(X'*y);
e    = y - X*beta;
sig2 = (e'*e)/T;

store_theta = zeros(nsim, k+1);     % [beta' sig2]
store_fy = zeros(size(ygrid)); % store only the sum

% Gibbs sampler
for isim = 1:nsim + burnin
    % sample beta
    Dbeta = (iVbeta + X'*X/sig2)\speye(k);
    beta_hat = Dbeta*(iVbeta*beta0 + X'*y/sig2);
    C = chol(Dbeta,'lower');
    beta = beta_hat + C*randn(k,1);

    % sample sig2
    e = y - X*beta;
    sig2 = 1/gamrnd(nu0 + T/2, 1/(S0 + 0.5*(e'*e)));

    if isim > burnin
        isave = isim - burnin;
        store_theta(isave,:) = [beta' sig2];

        % store one-step-ahead predictive density
        mu1 = xTp1'*beta;
        s2_1 = sig2;
        fy = normpdf(ygrid, mu1, sqrt(s2_1));
        store_fy = store_fy + fy;
    end
end
```

```
% posterior summaries
theta_mean = mean(store_theta,1);
theta_CI = quantile(store_theta,[.025 .975],1);
fy_hat = store_fy/nsim; % averaged posterior predictive density
```

Listing 2.2: Posterior predictive density under the independent normal and inverse-gamma prior (`linreg_indep_predictive.m`).

When the hyperparameters are chosen to align closely with those used under the natural conjugate prior, the posterior predictive density obtained via Gibbs sampling under the independent prior is nearly indistinguishable from the closed-form Student-t predictive density derived in Section 2.2.3. In effect, the Student-t predictive density is approximated here by averaging conditional Gaussian predictive densities over posterior draws of $(\boldsymbol{\beta}, \sigma^2)$. This averaging over uncertainty in the error variance produces heavier tails than a plug-in Gaussian forecast. As a result, the predictive distribution for 2020Q1 looks essentially the same as Figure 2.1, and we omit a duplicate plot. This representation—viewing a non-Gaussian predictive density as an average of Gaussian densities with respect to a mixing distribution—will reappear in a more general form in later chapters.

The posterior means of the first- and second-lag autoregressive coefficients are 0.673 and 0.186, respectively, which are virtually identical to those under the natural conjugate prior. The posterior predictive mean is 1.69, and the realized inflation rate of 1.39 lies at the 41.6th percentile of the predictive distribution, matching the conjugate analysis up to Monte Carlo error.

Under the independent prior, all inference, including prediction, is carried out via simulation. The same approach applies when conjugacy is unavailable—for example, in hierarchical, shrinkage, or nonlinear specifications—and it is the template for the Bayesian computation developed in later chapters.

2.4 Further Reading

Bayesian linear regression and conjugate analysis are classic topics with a well-developed literature. The foundational monograph of Zellner (1971) provides the original and most systematic treatment of the natural conjugate prior and its implications for inference, prediction, and model comparison.

Modern textbook treatments include Gelman et al. (2013), a standard reference on Bayesian data analysis that covers posterior simulation and hierarchical modeling, and Koop (2003), a comprehensive Bayesian econometrics

text with applications to regression, panel data, discrete choice, and time-series models.

Chan et al. (2019) adopt a complementary, problem-driven approach. Their textbook is organized around worked examples and exercises that guide readers through the derivation and implementation of standard Bayesian models, including the normal linear regression model discussed in this chapter.

Exercises

Exercise 2.1 (Normal-Inverse-Gamma Density)**.** Suppose $s \sim \mathcal{IG}(\alpha, \beta)$ and $(\mathbf{w} \,|\, s) \sim \mathcal{N}(\boldsymbol{\mu}, s\boldsymbol{\Sigma})$, i.e., $(\mathbf{w}, s) \sim \mathcal{NIG}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta)$. Verify that the joint density of $(\mathbf{w}, s)$ is given by

$$f(\mathbf{w}, s; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \beta) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} s^{-(\alpha+1+n/2)} \mathrm{e}^{-\frac{1}{s}\left[\beta + \frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\right]}.$$

Exercise 2.2 (Equivalent Forms of the Posterior Scale)**.** Using the definitions in Theorem 2.1, show that the posterior scale parameter $\widehat{S}$ can be written equivalently as

$$\widehat{S} = S_0 + \tfrac{1}{2}\Big[(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\Big].$$

Hint: Use the definitions of $\widehat{\boldsymbol{\beta}}$ and $\mathbf{D}_{\boldsymbol{\beta}}$ and expand the quadratic forms.

Exercise 2.3 (Student-t Marginal for $\boldsymbol{\beta}$)**.** Verify Corollary 2.1 by integrating out σ^2 from the joint normal-inverse-gamma posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$.

Exercise 2.4 (Marginal Likelihood under Conjugacy)**.** Using Theorem 2.1 and Corollary 2.2, derive the closed-form expression for the marginal likelihood $p(\mathbf{y})$ under the natural conjugate prior. *Hint:* Start from Bayes' theorem,

$$p(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) = \frac{p(\boldsymbol{\beta}, \sigma^2)\, p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2)}{p(\mathbf{y})},$$

and evaluate the ratio of normalizing constants by substituting the prior, likelihood, and posterior normal-inverse-gamma densities. The identity holds for any $(\boldsymbol{\beta}, \sigma^2)$, so it is convenient to evaluate it at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$, at which point both the quadratic form in $\boldsymbol{\beta}$ and the powers of σ^2 cancel.

Exercise 2.5 (Noninformative Prior and the Classical OLS Connection)**.** Consider the normal linear regression model with likelihood (2.1) under the improper prior $p(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$, with $\mathbf{X}$ of full column rank and $T > k$. Let $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\mathrm{RSS} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})$.

(a) Show that $(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}, (\mathbf{X}'\mathbf{X})^{-1}, \frac{T-k}{2}, \frac{\mathrm{RSS}}{2})$, and hence $(\boldsymbol{\beta} \,|\, \mathbf{y}) \sim \mathcal{T}_{T-k}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}, s^2(\mathbf{X}'\mathbf{X})^{-1})$ with $s^2 = \mathrm{RSS}/(T-k)$.

(b) Conclude that the credible intervals for β_i coincide with the usual classical confidence intervals, and explain why the degrees of freedom are $T-k$ rather than the $2\widehat{\nu}$ of the proper conjugate prior.

Exercise 2.6 (Gibbs versus Conjugate Inference)**.** Replicate the application in Section 2.2.3 using a Gibbs sampler rather than the closed-form normal-inverse-gamma formulas.

(a) Using the same AR(2) specification and the same natural conjugate normal-inverse-gamma prior, implement a Gibbs sampler for $(\boldsymbol{\beta}, \sigma^2)$.

(b) Use the simulated draws to estimate the posterior means of $\boldsymbol{\beta}$ and σ^2, and compare them with the analytical posterior means implied by Theorem 2.1.

(c) Using the Gibbs output, simulate the one-step-ahead posterior predictive distribution for 2020Q1 and compare its mean and predictive percentile at $y_{T+1} = 1.39$ with the closed-form Student-t predictive results.

Exercise 2.7 (Two-Step-Ahead Predictive Density)**.** Extend the PCE forecasting application in Section 2.2.3 to compute the two-step-ahead posterior predictive density for PCE inflation in 2020Q2 under the AR(2) model and the normal-inverse-gamma prior.

(a) Show that, conditional on y_{T+1} and the data $\mathbf{y}$, the predictive distribution of y_{T+2} is Student-t,

$$(y_{T+2} \,|\, y_{T+1}, \mathbf{y}) \sim \mathcal{T}_{2\widehat{\nu}}\left(\mathbf{x}'_{T+2}\widehat{\boldsymbol{\beta}}, \frac{\widehat{S}}{\widehat{\nu}}\Big(1 + \mathbf{x}'_{T+2}\mathbf{D}_{\boldsymbol{\beta}}\mathbf{x}_{T+2}\Big)\right),$$

where $\mathbf{x}_{T+2} = (1, y_{T+1}, y_T)'$ and $(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}, \widehat{\nu}, \widehat{S})$ are defined in Theorem 2.1. Conditional on a realized or simulated value of y_{T+1}, the regressor vector $\mathbf{x}_{T+2}$ is treated as fixed.

(b) Generate draws $y^{(r)}_{T+1} \sim p(y_{T+1} \,|\, \mathbf{y})$, $r = 1, \ldots, R$. For each draw, evaluate the conditional density $p(y_{T+2} \,|\, y^{(r)}_{T+1}, \mathbf{y})$ on a grid of values for y_{T+2}, and average across r to obtain a Monte Carlo approximation of the two-step-ahead predictive density $p(y_{T+2} \,|\, \mathbf{y})$.

Exercise 2.8 (Ridge and Lasso as Bayesian MAP Estimators)**.** Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is $T \times 1$, $\mathbf{X}$ is a $T \times k$ matrix with full column rank and $(\boldsymbol{\varepsilon} \,|\, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2\mathbf{I}_T)$, with σ^2 known.

(a) *Ridge regression.* The ridge estimator is defined as

$$\boldsymbol{\beta}_{\text{ridge}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{k} \beta_j^2 \right\}.$$

Show that ridge regression can be interpreted as a Bayesian maximum a posteriori (MAP) estimator by assuming the Gaussian prior

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}_k, \gamma \mathbf{I}_k).$$

Derive the posterior mode of $p(\boldsymbol{\beta} \,|\, \mathbf{y})$ and determine how γ must relate to λ for the two estimators to coincide.

(b) *Lasso.* The Lasso estimator is defined as

$$\boldsymbol{\beta}_{\text{Lasso}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \sum_{j=1}^{k} |\beta_j| \right\}.$$

Show that the Lasso can be interpreted as a Bayesian MAP estimator under independent Laplace (double-exponential) priors on the regression coefficients:

$$p(\boldsymbol{\beta}) = \prod_{j=1}^{k} \frac{1}{2\tau} \mathrm{e}^{-|\beta_j|/\tau}.$$

Exercise 2.9 (Linear Regression with a g-Prior)**.** Consider the linear regression model

$$(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_T),$$

where $\mathbf{X}$ is a $T \times k$ matrix with full column rank. Suppose the prior distributions are $p(\sigma^2) \propto 1/\sigma^2$ and $(\boldsymbol{\beta} \,|\, \sigma^2) \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma^2 g(\mathbf{X}'\mathbf{X})^{-1})$, where $\boldsymbol{\beta}_0$ and $g > 0$ are known. Derive the joint posterior distribution $p(\boldsymbol{\beta}, \sigma^2 \,|\, \mathbf{y})$.

Exercise 2.10 (Gibbs Sampling under a g-Prior)**.** Consider the linear regression model and prior specification in Exercise 2.9.

(a) Derive the conditional posterior distributions $p(\boldsymbol{\beta} \,|\, \sigma^2, \mathbf{y})$ and $p(\sigma^2 \,|\, \boldsymbol{\beta}, \mathbf{y})$.

(b) Using quarterly US PCE inflation data as in Section 2.2.3, estimate an AR(2) model under the g-prior using a Gibbs sampler. Report the posterior mean and posterior variance of σ^2, and compare them with the corresponding quantities obtained under the normal-inverse-gamma prior.

3

Linear Regression with General Errors

The linear regression model with independent Gaussian errors provides a convenient starting point for Bayesian inference, as it yields closed-form posterior distributions and straightforward computation. Empirical macroeconomic data, however, often display features that fall outside this setting, such as serial correlation and excess kurtosis. This chapter extends the Bayesian regression model of Chapter 2 to accommodate such departures while preserving a conditional structure that basic Markov chain Monte Carlo methods can exploit.

3.1 Gaussian Likelihood with Structured Dependence

We considered in Chapter 2 the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (\boldsymbol{\varepsilon} \,|\, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T).$$

This specification assumes independent, homoskedastic, and Gaussian errors. Empirical macroeconomic time series, however, often exhibit heavy-tailed disturbances and serial correlation that violate this assumption. To accommodate such departures within a tractable Gaussian framework, we assume

$$(\boldsymbol{\varepsilon} \,|\, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}_T, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is a positive-definite $T \times T$ covariance matrix indexed by a parameter vector $\boldsymbol{\theta}$. The standard linear regression model is recovered as the special case $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2 \mathbf{I}_T$. Serial correlation is captured directly through off-diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, whereas heavy-tailed errors are accommodated by a conditionally Gaussian representation in which $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is diagonal but observation-specific, as developed in Section 3.2. For now, we leave the structure of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ unspecified and introduce concrete examples in the sections that follow.

Under this assumption, the joint distribution of the data satisfies

$$(\mathbf{y} \,|\, \boldsymbol{\beta}, \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}),$$

with likelihood function

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{-T/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}. \tag{3.1}$$

We do not require prior independence between $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Instead, we specify a conditional prior for $\boldsymbol{\beta}$ given $\boldsymbol{\theta}$ and assume $(\boldsymbol{\beta} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$, where the prior mean $\boldsymbol{\beta}_0$ and covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}$ may depend on $\boldsymbol{\theta}$. The prior distribution for $\boldsymbol{\theta}$ will be specified once a particular covariance structure is imposed. Estimation proceeds via MCMC methods by iteratively sampling from the conditional distributions $p(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\theta})$ and $p(\boldsymbol{\theta} \mid \mathbf{y}, \boldsymbol{\beta})$. The following result characterizes the conditional distribution of $\boldsymbol{\beta}$.

Theorem 3.1 (Conditional Posterior of Regression Coefficients)**.** Under the normal linear regression model with $(\boldsymbol{\varepsilon} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}_T, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ and conditional prior $(\boldsymbol{\beta} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$, the conditional posterior distribution of $\boldsymbol{\beta}$ given $(\mathbf{y}, \boldsymbol{\theta})$ is

$$(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\theta}) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}),$$

where

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X} \right)^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}} \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}_0 + \mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{y} \right).$$

Proof. The result follows by combining the Gaussian likelihood in (3.1) with the normal prior for $\boldsymbol{\beta}$ and completing the square in $\boldsymbol{\beta}$. The derivation is identical to that in Section 2.3, except that $\sigma^{-2}\mathbf{I}_T$ is replaced by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$. □

The expressions in Theorem 3.1 involve the inverse of the $T \times T$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$. Direct inversion is computationally expensive when T is large. In many empirically relevant models, however, either $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ or $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$ has a *band matrix* structure, meaning that its nonzero elements are confined to a small number of diagonals (see Appendix B.2). Exploiting this sparsity can dramatically reduce computational costs and is central to efficient implementation in time-series applications.

3.2 Linear Regression with Student-t Errors

The first example we consider is a linear regression model with Student-t errors. Relative to the normal distribution, the Student-t distribution exhibits heavier tails and is therefore more robust to outliers and other forms

of model misspecification. A large empirical literature has shown that models with heavier-tailed disturbances often provide a better fit to financial and macroeconomic time series.

We assume that the regression errors $\{\varepsilon_t\}$ are independently and identically distributed as $\mathcal{T}_\nu(0, \sigma^2)$, where the Student-$t$ distribution is defined in Appendix A. Under this parameterization, σ^2 is a scale parameter rather than the variance; when $\nu > 2$, the variance of ε_t is $\nu\sigma^2/(\nu - 2)$. Under this assumption, the likelihood function can be derived in a straightforward manner. Unlike the Gaussian case, however, the posterior distribution of the regression coefficients $\boldsymbol{\beta}$ is no longer normal, which complicates posterior sampling.

3.2.1 Latent-Variable Representation and Data Augmentation

To facilitate estimation, we exploit a latent-variable representation of the Student-t distribution. The key idea is to introduce a set of *latent variables* such that, conditional on these variables, the model reduces to a Gaussian regression with heteroskedastic errors. This representation allows us to retain simple conditional sampling steps within an MCMC algorithm. The approach was first implemented in a Bayesian regression context by Geweke (1993).

Theorem 3.2 (Scale-Mixture Representation of the Student-t Distribution)**.** Suppose

$$(\varepsilon \mid \lambda) \sim \mathcal{N}(0, \lambda\sigma^2), \quad \lambda \sim \mathcal{IG}(\nu/2, \nu/2).$$

Then the marginal distribution of ε is $\mathcal{T}_\nu(0, \sigma^2)$.

The result in Theorem 3.2 follows by integrating out the latent scale parameter λ from the joint density of (ε, λ). A detailed derivation is provided in Exercise 3.1.

This representation implies that a linear regression model with Student-t errors can be estimated by applying Theorem 3.2 to each error separately: introducing independent latent scale variables $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$, $t = 1, \ldots, T$, yields a conditionally Gaussian regression with observation-specific error variances. This is an instance of a general computational strategy known as *data augmentation*, in which latent variables are introduced to simplify posterior computation. We will encounter this strategy repeatedly in later chapters.

The latent variable λ_t acts as an observation-specific scale factor: conditional on λ_t, the error is Gaussian with variance $\lambda_t\sigma^2$, so a large λ_t inflates the variance of observation t. This lets the model accommodate occasional large disturbances without letting them unduly influence the estimated regression coefficients.

3.2.2 Posterior Sampling with Known ν

Using the latent-variable representation introduced above, a linear regression with Student-t errors can be written as:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t,$$

where $\mathbf{x}_t = (1, x_{2t}, \ldots, x_{kt})'$ is the vector of covariates, $(\varepsilon_t \,|\, \lambda_t, \sigma^2) \sim \mathcal{N}(0, \sigma^2\lambda_t)$ and $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$. Throughout this subsection, we assume that the degrees-of-freedom parameter ν is known.

We consider the following independent normal and inverse-gamma prior on $(\boldsymbol{\beta}, \sigma^2)$:

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}}), \quad \sigma^2 \sim \mathcal{IG}(\nu_0, S_0).$$

Relative to the Gaussian linear regression model, the introduction of Student-t errors adds an additional block of latent variables $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ to the parameter space.

Posterior simulation is carried out using a Gibbs sampler that iteratively draws from the following conditional distributions: $p(\boldsymbol{\beta} \,|\, \mathbf{y}, \boldsymbol{\lambda}, \sigma^2)$, $p(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\lambda})$, and $p(\boldsymbol{\lambda} \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2)$.

Conditional on $(\boldsymbol{\lambda}, \sigma^2)$, the model can be written in vector form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $(\boldsymbol{\varepsilon} \,|\, \boldsymbol{\lambda}, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2\boldsymbol{\Lambda})$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_T)$. This falls within the general framework developed in Section 3.1. Applying Theorem 3.1 with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2\boldsymbol{\Lambda}$ yields

$$(\boldsymbol{\beta} \,|\, \mathbf{y}, \boldsymbol{\lambda}, \sigma^2) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}), \tag{3.2}$$

where

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\sigma^2}\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{X}\right)^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}\left(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma^2}\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{y}\right).$$

Here, $\boldsymbol{\Lambda}^{-1} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_T^{-1})$ is a diagonal—and hence sparse—matrix, which is straightforward to construct computationally.

The update for $\boldsymbol{\beta}$ shows this downweighting concretely: its posterior precision depends on the weighted cross-product $\mathbf{X}'\boldsymbol{\Lambda}^{-1}\mathbf{X}$, in which observations with large λ_t receive smaller weights and therefore exert less influence on the regression coefficients.

Next, conditional on $(\boldsymbol{\beta}, \boldsymbol{\lambda})$, the variance parameter σ^2 follows an inverse-gamma distribution:

$$(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\lambda}) \sim \mathcal{IG}\left(\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Lambda}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right),$$

the derivation of which closely parallels that of Section 2.3.1 and is omitted.

Finally, we consider the conditional distribution of the latent scale variables. Although $p(\boldsymbol{\lambda} \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2)$ is T-dimensional, the latent variables $\lambda_1, \ldots, \lambda_T$ are conditionally independent given the data and $(\boldsymbol{\beta}, \sigma^2)$. As a result, this step reduces to sampling T independent univariate distributions.

Combining the likelihood contribution of observation t with the prior $\lambda_t \sim \mathcal{IG}(\nu/2, \nu/2)$, we have

$$\begin{aligned} p(\boldsymbol{\lambda} \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2) &\propto p(\mathbf{y} \mid \boldsymbol{\beta}, \boldsymbol{\lambda}, \sigma^2) p(\boldsymbol{\lambda} \mid \nu) \\ &\propto \prod_{t=1}^{T} \left[(2\pi\lambda_t\sigma^2)^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\lambda_t\sigma^2}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2} \times (\lambda_t)^{-\left(\frac{\nu}{2}+1\right)} \mathrm{e}^{-\frac{\nu}{2\lambda_t}} \right] \\ &\propto \prod_{t=1}^{T} \left[(\lambda_t)^{-\left(\frac{\nu+1}{2}+1\right)} \mathrm{e}^{-\frac{1}{2\lambda_t}\left(\nu + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma^2}\right)} \right], \end{aligned}$$

which corresponds to a product of inverse-gamma kernels. Therefore,

$$(\lambda_t \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2) \sim \mathcal{IG}\left(\frac{\nu+1}{2}, \frac{1}{2}\left(\nu + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma^2} \right) \right), \quad t = 1, \ldots, T.$$

Because λ_t is drawn from its conditional posterior, a large squared residual $(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2$ tends to produce a larger λ_t, which downweights that observation in the next $\boldsymbol{\beta}$ update. Because $\boldsymbol{\Lambda}$ stays diagonal, this robustness comes at little computational cost. Exercise 3.2 makes the mechanism precise by deriving the expected precision weight $\mathbb{E}(\lambda_t^{-1} \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2)$, which shrinks toward zero as the standardized residual grows.

The Gibbs sampler for the linear regression model with Student-t errors and known ν proceeds as follows. Choose initial values

$$\boldsymbol{\beta}^{(0)}, \quad \sigma^{2(0)} > 0, \quad \lambda_t^{(0)} > 0 \ \text{ for } t = 1, \ldots, T.$$

A convenient choice is to initialize $\boldsymbol{\beta}^{(0)}$ at the ordinary least squares estimate, set $\sigma^{2(0)}$ equal to the corresponding residual variance, and take $\lambda_t^{(0)} = 1$ for all t. For $r = 1, \ldots, R$, repeat the following steps:

1. Draw $\boldsymbol{\beta}^{(r)} \sim p(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\lambda}^{(r-1)}, \sigma^{2(r-1)})$.
2. Draw $\sigma^{2(r)} \sim p(\sigma^2 \mid \mathbf{y}, \boldsymbol{\beta}^{(r)}, \boldsymbol{\lambda}^{(r-1)})$.
3. Draw $\lambda_t^{(r)} \sim p(\lambda_t \mid \mathbf{y}, \boldsymbol{\beta}^{(r)}, \sigma^{2(r)}), t = 1, \ldots, T$.

3.2.3 Posterior Sampling with Unknown ν

In practice, the degrees-of-freedom parameter ν is typically unknown and must be estimated from the data. In this section, we treat ν as an unknown param-

eter and extend the Gibbs sampler accordingly. Specifically, estimation with unknown ν requires adding to the Gibbs sampler of the previous section a single additional block that samples ν conditional on the remaining parameters and latent variables. Because all other conditional distributions are evaluated given ν, the remaining steps of the Gibbs sampler remain unchanged. This modular structure is a key strength of MCMC methods: model extensions can often be implemented by making only minor modifications to existing algorithms.

To complete the model specification, we need to assign a prior distribution to ν. We adopt a uniform prior on the interval $(2, \bar{\nu})$. The restriction $\nu > 2$ ensures that the variance of the Student-t error distribution exists. The upper bound $\bar{\nu}$ is chosen to be sufficiently large—e.g., $\bar{\nu} = 50$—so that the Student-t distribution can closely approximate the normal distribution when supported by the data.

Under this prior, the conditional density of ν given the remaining quantities is

$$\begin{aligned} p(\nu \,|\, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \sigma^2) &\propto p(\boldsymbol{\lambda} \,|\, \nu) p(\nu) \\ &\propto \prod_{t=1}^{T} \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} \lambda_t^{-\left(\frac{\nu}{2}+1\right)} \mathrm{e}^{-\frac{\nu}{2\lambda_t}} \\ &= \frac{(\nu/2)^{\frac{T\nu}{2}}}{\Gamma(\nu/2)^T} \left(\prod_{t=1}^{T} \lambda_t \right)^{-\left(\frac{\nu}{2}+1\right)} \mathrm{e}^{-\frac{\nu}{2} \sum_{t=1}^{T} \lambda_t^{-1}} \end{aligned} \tag{3.3}$$

for $2 < \nu < \bar{\nu}$, and 0 otherwise. The resulting conditional density is nonstandard and does not correspond to a distribution from which direct simulation is straightforward. Consequently, an additional numerical method is required to sample ν within the Gibbs sampler.

To this end, we introduce the *inverse-transform method*, which provides a general approach for simulating a random variable from a specified cumulative distribution function (cdf).

Algorithm 3.1 Inverse-Transform Method

Input: A cumulative distribution function F of the target random variable.

1. Generate $U \sim \mathcal{U}(0, 1)$.
2. Return a draw X defined by

$$X = \begin{cases} \min\{x : F(x) \geq U\}, & \text{if } X \text{ is discrete,} \\ F^{-1}(U), & \text{if } X \text{ is continuous.} \end{cases}$$

It is straightforward to verify that the random variable $X = F^{-1}(U)$ has cdf F. In particular, for any x,

$$\mathbb{P}(X \leqslant x) = \mathbb{P}(F^{-1}(U) \leqslant x) = \mathbb{P}(U \leqslant F(x)) = \int_0^{F(x)} 1 \, \mathrm{d}u = F(x),$$

which establishes the result.

Figure 3.1 illustrates the inverse-transform method for both discrete and continuous distributions. In the left panel, the random variable takes values on the finite support $\{1, 2, \ldots, 6\}$, and the cdf is a step function that increases at each integer value according to the associated probability mass. Given a uniform draw $U \sim \mathcal{U}(0, 1)$, the inverse-transform method selects the smallest integer x such that the cumulative probability $F(x)$ exceeds U. In the right panel, the random variable has a continuous distribution with a smooth cdf. In this case, the same uniform draw is mapped to an outcome by evaluating the inverse cdf at U, yielding a continuous-valued realization.

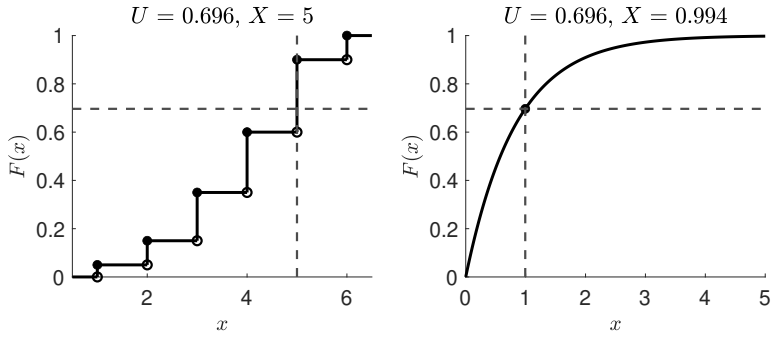

FIGURE 3.1: Inverse-transform method for discrete and continuous distributions. The left panel shows a discrete cdf on $\{1, \ldots, 6\}$, where a draw is the smallest value with cumulative probability exceeding a uniform draw. The right panel shows a continuous cdf, where a draw is obtained by evaluating the inverse cdf at the same uniform realization.

The inverse-transform method is particularly convenient when the target distribution is discrete or when the cdf F is available in closed form. In such cases, simulation reduces to evaluating the inverse cdf at a single uniform draw. When F is not analytically invertible, however, numerical inversion is required, which can substantially limit the practical usefulness of the method.

In many applications, the inverse-transform method is therefore difficult to implement. As an alternative, the *Griddy-Gibbs sampler* (Ritter and Tanner, 1992) provides a practical approach for approximating draws from univariate distributions with bounded support. The Griddy-Gibbs sampler can be viewed

as a discretized version of the inverse-transform method and requires only the evaluation of the target density, up to a normalizing constant.

Algorithm 3.2 Griddy-Gibbs Sampler

Input: An unnormalized density $f(x)$ on (a, b) and a grid $\{x_1, \ldots, x_n\}$.

1. Evaluate $f(x_j)$ on the grid and form the cumulative sums

$$F_i = \sum_{j=1}^{i} f(x_j), \quad i = 1, \ldots, n.$$

2. Normalize so that $F_n = 1$.

3. Draw $U \sim \mathcal{U}(0, 1)$ and return $X = x_k$, where $k = \min\{i : F_i \geq U\}$.

The basic idea of the Griddy-Gibbs sampler is to approximate the cumulative distribution function of a bounded univariate target density on a fine grid and then apply the inverse-transform method to the resulting discrete approximation.

The following MATLAB function `griddy_gibbs.m` implements a generic Griddy-Gibbs step. It takes a function handle `logf` that evaluates the log of an unnormalized target density, lower and upper bounds (a, b), and the number of grid points `n_grid`, so it applies to any bounded univariate conditional distribution known up to a normalizing constant.

To improve numerical stability, the log-density is shifted by its maximum value before exponentiation, which avoids underflow or overflow when forming the normalized weights.

```
function [x_draw, x_grid, logf_grid] = ...
    griddy_gibbs(logf, a, b, n_grid)
% jittered grid to avoid repeated draws
step = (b - a) / (n_grid + 1);
x_grid = a + step*(1:n_grid)';
  % add small jitter
x_grid = x_grid + (rand(n_grid,1) - 0.5)*step;
    % enforce strict bounds
x_grid = min(max(x_grid, a + eps), b - eps);

% evaluate logf on the grid and normalize
logf_grid = logf(x_grid);
w = exp(logf_grid - max(logf_grid));
w = w / sum(w); % normalize to sum to 1

cdf_grid = cumsum(w);
u = rand;
idx = find(cdf_grid >= u, 1, 'first');
```

```
x_draw = x_grid(idx);
end
```

Listing 3.1: Griddy-Gibbs sampler for a bounded univariate target (`griddy_gibbs.m`).

Introducing a small amount of random jitter into the grid helps avoid repeatedly drawing from the same fixed discrete support across MCMC iterations. When the grid is sufficiently fine, the Griddy-Gibbs sampler typically provides an accurate approximation to the desired conditional distribution. Its main limitation is scalability: extending the approach beyond the univariate case becomes infeasible because the number of grid points required for a given level of accuracy grows exponentially with dimension. In such settings, the Metropolis–Hastings algorithm, introduced in Section 3.3 and developed further in Chapter 6, provides a general alternative.

To apply this generic routine to sample the degrees-of-freedom parameter ν, we define a wrapper function `sample_nu_griddy.m`. This function constructs the log kernel of the conditional density $p(\nu \mid \boldsymbol{\lambda})$ and calls `griddy_gibbs.m` with support $(2, \bar{\nu})$.

```
function nu = sample_nu_griddy(lam, nu_ub, n_grid)
T = length(lam);
sum_loglam = sum(log(lam));
sum_ilam = sum(1./lam);

% log kernel of p(nu | lambda)
log_kernel = @(x) T*((x/2).*log(x/2) - gammaln(x/2)) ...
    - (x/2 + 1) * sum_loglam - (x/2) * sum_ilam;

nu = griddy_gibbs(log_kernel, 2, nu_ub, n_grid);
end
```

Listing 3.2: Griddy-Gibbs update for ν using the generic sampler (`sample_nu_griddy.m`).

We now turn to an empirical illustration that brings together the modeling and computational tools developed in this section.

3.2.4 Application: Modeling PCE Inflation via an AR Model with Student-t Errors

To illustrate the practical value of allowing for Student-t disturbances, we revisit the PCE inflation example previously analyzed under Gaussian errors in Section 2.3.2. Using quarterly US PCE inflation over the sample period 1960Q1–2019Q4, we now estimate the same AR(2) specification but replace the normal error assumption with a Student-t distribution. This comparison

highlights how heavier-tailed disturbances alter inference in the presence of large, infrequent shocks—an empirically relevant feature of inflation data.

Specifically, we consider the following AR(2) model with Student-t errors:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t,$$

where the disturbances admit the scale-mixture representation $(\varepsilon_t \mid \lambda_t, \sigma^2) \sim \mathcal{N}(0, \lambda_t \sigma^2)$ and $(\lambda_t \mid \nu) \sim \mathcal{IG}(\nu/2, \nu/2)$. Estimation is carried out using the Gibbs sampler described earlier, with an additional Griddy-Gibbs step to sample the degrees-of-freedom parameter ν. The MATLAB script `linreg_t.m`, reported below, implements this algorithm. A key computational feature of the implementation is that the inverse covariance or precision matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} = \sigma^{-2} \boldsymbol{\Lambda}^{-1}, \quad \boldsymbol{\Lambda}^{-1} = \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_T^{-1}),$$

is diagonal and therefore sparse. In the code, this structure is exploited by constructing $\boldsymbol{\Lambda}^{-1}$ using

```
iLam = sparse(1:T,1:T,1./lam);
```

which substantially reduces both memory usage and computational cost.

The code listing below implements the full Gibbs sampler and stores posterior draws of the regression coefficients, variance, degrees-of-freedom parameter, and latent scale variables. For brevity, the plotting commands are omitted from the listing. The shaded 95% credible band for λ_t in Figure 3.2 is produced using the auxiliary MATLAB function `shaded_band`, which we reuse in later chapters.

```
nsim = 20000; burnin = 1000;

% load data
data = readmatrix('USPCE.csv', 'Range', 'B2:B241');
y0 = data(1:4);  % initial conditions: [y_{-3},y_{-2},y_{-1},y_0]
y = data(5:end); % sample used for estimation
T = size(y,1);

% regressors for AR(2): [1, y_{t-1}, y_{t-2}]
xlag1 = [y0(4); y(1:end-1)];
xlag2 = [y0(3:4); y(1:end-2)];
X = [ones(T,1), xlag1, xlag2];
k = size(X,2);   % number of regressors

% prior hyperparameters (indep normal and inverse-gamma)
beta0  = zeros(k,1);
iVbeta  = 1/100*speye(k);  % prior precision of beta
nu0    = 4;
S0     = 1;
nu_ub = 50;                % prior upperbound of nu
```

```
% initialize the Markov chain
nu = 5;
beta = (X'*X)\(X'*y);
sig2 = sum((y-X*beta).^2)/T;
lam = ones(T,1);
iLam = sparse(1:T,1:T,1./lam);

store_theta = zeros(nsim,k+2);  % [beta', sig2, nu]
store_lam = zeros(nsim,T);
n_grid = 500;

for isim = 1:nsim + burnin
    % sample beta
    Dbeta = (iVbeta + X'*iLam*X/sig2)\speye(k);
    beta_hat = Dbeta*(iVbeta*beta0 + X'*iLam*y/sig2);
    C = chol(Dbeta,'lower');
    beta = beta_hat + C*randn(k,1);

    % sample sig2
    e = y - X*beta;
    sig2 = 1/gamrnd(nu0+T/2,1/(S0 + e'*iLam*e/2));

    % sample lam
    lam = 1./gamrnd((nu+1)/2,2./(nu+e.^2/sig2));
    iLam = sparse(1:T,1:T,1./lam);

    % sample nu
    nu = sample_nu_griddy(lam,nu_ub,n_grid);

        % store the parameters
    if isim > burnin
        isave = isim - burnin;
        store_theta(isave,:) = [beta' sig2 nu];
        store_lam(isave,:) = lam';
    end
end
% Posterior summaries
theta_mean = mean(store_theta,1);
theta_lo  = quantile(store_theta,.025,1)';
theta_hi  = quantile(store_theta,.975,1)';
```

Listing 3.3: Gibbs sampler for an AR(2) model with Student-t errors (`linreg_t.m`).

Using a sample of $20,000$ posterior draws, the posterior means of the first- and second-lag autoregressive coefficients and the degrees-of-freedom parameter are estimated to be 0.662, 0.223, and 3.9, respectively. The left panel of Figure 3.2 displays posterior draws of the degrees-of-freedom parameter ν, while the right panel plots the posterior mean of the latent scale λ_t together with a pointwise 95% credible band on a logarithmic scale. As is evident from the left panel, most of the posterior mass of ν lies well below 10, indicating that the error distribution exhibits substantially heavier tails than the Gaussian distribution.

The posterior means of the latent scale variables $\{\lambda_t\}$ indicate that, for most quarters in the sample, inflation dynamics are well described by moderately scaled Gaussian disturbances. The majority of the λ_t values lie between approximately 1 and 2.5, implying that typical observations receive weights similar to those in a homoskedastic normal regression. This suggests that, outside of exceptional episodes, the AR(2) model captures the conditional mean dynamics of inflation reasonably well, with only modest adjustments for tail thickness.

At the same time, the estimated latent scales exhibit a small number of pronounced spikes, with several quarters associated with very large values of λ_t. These spikes correspond to periods of unusually large disturbances, such as the oil shocks and the Great Inflation of the 1970s, the Volcker disinflation, and the Great Recession. The largest estimated λ_t is approximately 54, corresponding to 2008Q4, at the height of the recession. Under the scale-mixture representation, these large λ_t downweight the corresponding observations in the estimation of the regression coefficients. This downweighting is entirely data-driven: the model attenuates extreme observations without any *ad hoc* outlier rules or break dummies.

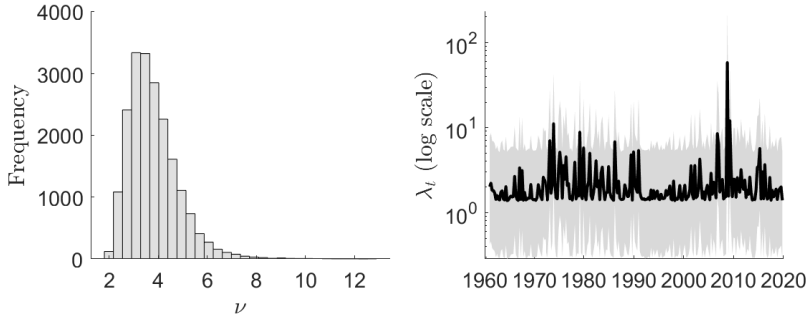

FIGURE 3.2: Posterior distribution of the degrees-of-freedom parameter ν and latent scale variables in an AR(2) model for US inflation with Student-t errors. The left panel shows the histogram of posterior draws of ν. The right panel plots the posterior mean of the latent scale λ_t together with a pointwise 95% credible band (log scale).

3.3 Linear Regression with Moving Average Errors

Thus far, we have assumed that the regression disturbances are serially independent. This assumption is frequently violated in macroeconomic and fi-

nancial applications, where the disturbances exhibit serial correlation. This section introduces a simple and widely used extension of the linear regression model that accommodates such dependence through moving average errors.

3.3.1 Model Specification and Likelihood

Consider the linear regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t,$$

where the error process $\{\varepsilon_t\}$ follows a moving average model of order one,

$$\varepsilon_t = u_t + \psi u_{t-1}. \tag{3.4}$$

We impose the invertibility restriction $|\psi| < 1$ and assume $u_0 = 0$ and $\{u_t\}$ is iid $\mathcal{N}(0, \sigma^2)$. When $\psi = 0$, the model reduces to the standard linear regression with independent errors. Hence, testing the hypothesis $\psi = 0$ provides a direct assessment of whether serial correlation is present in the disturbances.

This model fits naturally into the general framework introduced in Section 3.1. To apply the results derived there, it suffices to characterize the joint distribution of $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)'$. We rewrite (3.4) in matrix form as

$$\boldsymbol{\varepsilon} = \mathbf{H}_\psi \mathbf{u}, \tag{3.5}$$

where $(\mathbf{u} \,|\, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T)$, and

$$\mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \psi & 1 & 0 & \cdots & 0 \\ 0 & \psi & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \psi & 1 \end{pmatrix}.$$

The matrix $\mathbf{H}_\psi$ can be written compactly as $\mathbf{H}_\psi = \mathbf{I}_T + \psi \mathbf{S}_1$, where $\mathbf{S}_1$ is the first-lag shift matrix with ones on the first subdiagonal and zeros elsewhere. This shift-matrix representation recurs in later chapters, notably in the state space models of Chapter 9.

Several computational features of this representation are worth emphasizing, as they will play an important role in later chapters. First, $\mathbf{H}_\psi$ is a lower triangular $T \times T$ matrix with ones on the main diagonal. As a result, its determinant is equal to one, and the matrix is nonsingular for any finite ψ. Second, $\mathbf{H}_\psi$ is a band matrix: its nonzero elements are confined to the main diagonal and the first subdiagonal. This sparsity structure can be exploited to substantially reduce computational costs in likelihood evaluation and posterior sampling.

Although $\mathbf{H}_\psi^{-1}$ exists, it is dense. In what follows, expressions may involve $\mathbf{H}_\psi^{-1}$ implicitly, but this inverse is never computed directly in practice. Instead, all computations are carried out using the banded structure of $\mathbf{H}_\psi$ itself.

It follows immediately from (3.5) that $\boldsymbol{\varepsilon}$ is a linear transformation of a normal random vector. Consequently, $\boldsymbol{\varepsilon}$ itself is multivariate normal. We now derive its mean vector and covariance matrix.

Since $(\mathbf{u} \,|\, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T)$, the mean of $\boldsymbol{\varepsilon}$ is

$$\mathbb{E}\,\boldsymbol{\varepsilon} = \mathbf{H}_\psi \mathbb{E}\,\mathbf{u} = \mathbf{0}_T.$$

The covariance matrix is given by

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{H}_\psi \mathrm{Cov}(\mathbf{u}) \mathbf{H}_\psi' = \sigma^2 \mathbf{H}_\psi \mathbf{H}_\psi'.$$

Because $\mathbf{H}_\psi$ has nonzero elements only on the main diagonal and the first subdiagonal, the covariance matrix $\sigma^2 \mathbf{H}_\psi \mathbf{H}_\psi'$ is a *tridiagonal* matrix. This structure implies that each error term is correlated only with its immediate neighbors, reflecting the one-period dependence inherent in the MA(1) specification. It follows that the conditional distribution of the data is

$$(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2, \psi) \sim \mathcal{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{H}_\psi \mathbf{H}_\psi'\right). \tag{3.6}$$

The linear regression model with MA(1) errors therefore fits squarely into the general framework of Section 3.1, with $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2 \mathbf{H}_\psi \mathbf{H}_\psi'$. We maintain the same independent priors for $\boldsymbol{\beta}$ and σ^2 as in previous sections. For the moving-average parameter, we adopt a uniform prior on $(-1, 1)$, $\psi \sim \mathcal{U}(-1, 1)$, which enforces the invertibility condition.

Before describing the Gibbs sampler, we first outline an efficient method for evaluating the likelihood function of the MA(1) regression model. A traditional approach is to cast the model into a linear Gaussian state space form (see Chapter 9) and evaluate the likelihood via the Kalman filter. Here we instead follow the direct approach of Chan (2013), which exploits fast band-matrix algorithms and is particularly convenient in regression settings.

Specifically, since $\det(\mathbf{H}_\psi) = 1$, we have $\det(\mathbf{H}_\psi \mathbf{H}_\psi') = 1$, and the likelihood implied by (3.6) can be written as

$$p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2, \psi) = (2\pi\sigma^2)^{-T/2} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}.$$

Thus, likelihood evaluation reduces to computing the quadratic form in the exponent.

To do so efficiently, we avoid forming $(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1}$ explicitly. Instead, define

$$\mathbf{u} = \mathbf{H}_\psi^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

which can be obtained by solving the banded linear system

$$\mathbf{H}_\psi \mathbf{z} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

for $\mathbf{z}$. Since $\mathbf{H}_\psi$ is lower triangular with bandwidth one, this system can be solved quickly using standard routines—e.g., the backslash operator \ in MATLAB. The quadratic form then simplifies to $\mathbf{u}'\mathbf{u}$, yielding a fast and numerically stable evaluation of the likelihood.

The MA(1) specification is adopted here for simplicity, and the framework extends immediately to a general moving average model of order q. In particular, for an MA(q) error process

$$\varepsilon_t = u_t + \psi_1 u_{t-1} + \cdots + \psi_q u_{t-q},$$

the matrix $\mathbf{H}_\psi$ in (3.5) is replaced by a lower triangular band matrix with $q+1$ nonzero diagonals, where the main diagonal contains ones and the jth subdiagonal contains ψ_j, $j = 1, \ldots, q$. All results above continue to hold with this redefinition of $\mathbf{H}_\psi$. In particular, $\mathbf{H}_\psi$ remains sparse, its determinant equals one, and likelihood evaluation and posterior sampling can be carried out using the same band-matrix algorithms. For $q > 1$, however, the invertibility region is no longer simply $|\psi_j| < 1$, and the prior for $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_q)'$ should be restricted to this region accordingly.

Having established the likelihood and its efficient evaluation, we now turn to posterior sampling for the regression model with MA errors.

3.3.2 Posterior Sampling via Metropolis-within-Gibbs

Posterior simulation for the MA(1) regression model proceeds by combining Gibbs updates for parameters with conjugate conditional distributions and a Metropolis–Hastings (MH) step for the moving average parameter. Specifically, we iteratively draw from the conditional distributions of $p(\boldsymbol{\beta} \,|\, \mathbf{y}, \psi, \sigma^2)$ and $p(\sigma^2 \,|\, \mathbf{y}, \psi, \boldsymbol{\beta})$, and update ψ using an MH step.

The first step of sampling from $p(\boldsymbol{\beta} \,|\, \mathbf{y}, \psi, \sigma^2)$ is straightforward. Conditional on (ψ, σ^2), the model falls within the general framework of Section 3.1. Applying Theorem 3.1 yields

$$(\boldsymbol{\beta} \,|\, \mathbf{y}, \psi, \sigma^2) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}}), \tag{3.7}$$

where

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\sigma^2} \mathbf{X}'(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1} \mathbf{X} \right)^{-1},$$
$$\widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}} \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}_0 + \frac{1}{\sigma^2} \mathbf{X}'(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1} \mathbf{y} \right).$$

As in likelihood evaluation, direct computation of $(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1}$ is neither necessary nor desirable, since $\mathbf{H}_\psi^{-1}$ is dense. Instead, we exploit the banded structure of $\mathbf{H}_\psi$ by working with transformed variables. Specifically, define

$$\widetilde{\mathbf{X}} = \mathbf{H}_\psi^{-1}\mathbf{X}, \quad \widetilde{\mathbf{y}} = \mathbf{H}_\psi^{-1}\mathbf{y},$$

which can be obtained by solving the banded linear systems

$$\mathbf{H}_\psi \mathbf{Z} = \mathbf{X}, \qquad \mathbf{H}_\psi \mathbf{z} = \mathbf{y}.$$

Once $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{y}}$ are available, the posterior moments reduce to

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\sigma^2}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}, \quad \widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}\left(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \frac{1}{\sigma^2}\widetilde{\mathbf{X}}'\widetilde{\mathbf{y}}\right),$$

thereby avoiding the formation or inversion of any dense $T \times T$ matrices.

An equivalent interpretation of this procedure is to view it as *whitening* the regression errors. Specifically, left-multiplying the regression equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ by $\mathbf{H}_\psi^{-1}$ to obtain

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{H}_\psi^{-1}\boldsymbol{\varepsilon} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{u},$$

where $\widetilde{\mathbf{y}} = \mathbf{H}_\psi^{-1}\mathbf{y}$, $\widetilde{\mathbf{X}} = \mathbf{H}_\psi^{-1}\mathbf{X}$, and $(\mathbf{u} \,|\, \sigma^2) \sim \mathcal{N}(\mathbf{0}_T, \sigma^2 \mathbf{I}_T)$. The transformed regression therefore has serially independent Gaussian errors.

Because the whitened model is an ordinary linear regression with homoskedastic disturbances, the standard conjugate results from Chapter 2 apply directly and lead to the same conditional posterior distribution for $\boldsymbol{\beta}$ as in (3.7).

The idea of whitening correlated disturbances in Bayesian regression dates back at least to Chib and Greenberg (1994). Their implementation, however, differs from the one used here: instead of explicitly applying $\mathbf{H}_\psi^{-1}$, they compute the elements of $\widetilde{\mathbf{y}}$ and $\widetilde{\mathbf{X}}$ recursively via systems of equations. The matrix-based approach adopted here is typically faster, as it exploits vectorized operations and avoids explicit loops.

The update for σ^2 is also straightforward. Conditional on $(\boldsymbol{\beta}, \psi)$, the model is Gaussian with known covariance structure, and conjugacy implies that the conditional posterior distribution of σ^2 is inverse-gamma. Specifically, it can be shown that

$$(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}, \psi) \sim \mathcal{IG}\left(\nu_0 + \frac{T}{2}, S_0 + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{H}_\psi \mathbf{H}_\psi')^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

The derivation closely parallels that of the homoskedastic Gaussian regression and is omitted for brevity.

Next, given the uniform prior $\psi \sim \mathcal{U}(-1, 1)$, the conditional distribution of ψ is

$$p(\psi \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \psi)\, \mathbf{1}\{|\psi| < 1\},$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function and $p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \psi)$ is the likelihood function. This density does not belong to a standard family, so direct sampling is not available. Although a Griddy-Gibbs step could be used, it requires repeated likelihood evaluations on a fine grid and does not scale well to higher-order MA models. These considerations motivate the use of a Metropolis–Hastings update.

The MH algorithm constructs a Markov chain whose stationary distribution coincides with a target density f, known up to a normalizing constant. Starting from the current state $\boldsymbol{\theta}$, a proposal $\boldsymbol{\theta}^*$ is drawn from a proposal density $q(\cdot \mid \boldsymbol{\theta})$ and accepted with probability

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*)\, q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)}{f(\boldsymbol{\theta})\, q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})}\right\}.$$

Algorithm 3.3 Metropolis–Hastings Algorithm

Input: Target density $f(\boldsymbol{\theta})$, initial value $\boldsymbol{\theta}^{(0)}$, proposal density $q(\cdot \mid \boldsymbol{\theta})$, and number of iterations R.

1. For $r = 1, \ldots, R$:
   (a) Draw a proposal $\boldsymbol{\theta}^* \sim q(\cdot \mid \boldsymbol{\theta}^{(r-1)})$.
   (b) Compute the acceptance probability
   $$\alpha(\boldsymbol{\theta}^{(r-1)}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*)\, q(\boldsymbol{\theta}^{(r-1)} \mid \boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}^{(r-1)})\, q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(r-1)})}\right\}.$$
   (c) With probability α, set $\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^*$; otherwise set $\boldsymbol{\theta}^{(r)} = \boldsymbol{\theta}^{(r-1)}$.

One of the most widely used MH variants is the *random-walk sampler*, in which proposals are generated according to

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} + \mathbf{v}, \quad \mathbf{v} \sim g,$$

where the proposal distribution g is typically, though not necessarily, symmetric about zero. When g is symmetric, we have $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)$, so the proposal densities cancel in the acceptance ratio. As a result, the acceptance probability simplifies to

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta})}\right\}.$$

The performance of the algorithm depends critically on the scale of the random-walk proposal. Larger proposal steps promote broader exploration of the parameter space but typically result in lower acceptance rates, while very small steps yield high acceptance probabilities at the cost of slow mixing and highly persistent draws. In practice, the proposal variance must be tuned to balance these effects. A commonly used rule of thumb in univariate problems is to target an acceptance rate between roughly 20% and 40%. Further discussion of optimal scaling results and acceptance-rate guidelines is provided in Chapter 6. The random-walk sampler is straightforward to implement and often effective in low-dimensional settings, though its efficiency depends critically on careful tuning of the proposal scale.

Another important MH variant is the *independence-chain sampler*, obtained by choosing a proposal density that does not depend on the current state, $q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = g(\boldsymbol{\theta}^*)$. In this case, the acceptance probability becomes

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*)\, g(\boldsymbol{\theta})}{f(\boldsymbol{\theta})\, g(\boldsymbol{\theta}^*)}\right\}.$$

The performance of the independence-chain sampler hinges on how well the proposal density g approximates the target density f.

In contrast to the random-walk Metropolis algorithm, there is no intrinsic trade-off between step size and acceptance probability in the independence-chain sampler. When g provides a good global approximation to f, proposed draws are frequently accepted, and acceptance rates close to one are desirable and indicative of efficient sampling. In practice, however, constructing such a proposal often requires additional optimization and numerical differentiation. For this reason, especially in low-dimensional settings, a well-tuned random-walk Metropolis step is often preferred for its simplicity, even if it operates at a lower acceptance rate.

In the MA(1) regression, we update the moving average parameter ψ using an MH step targeting its conditional density

$$f(\psi) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \psi)\mathbf{1}\{|\psi| < 1\},$$

treating $(\boldsymbol{\beta}, \sigma^2)$ as fixed at their current Gibbs draws.

We now turn to an empirical illustration in which this Metropolis–Hastings update is embedded within a Gibbs sampler for an autoregressive moving average model.

3.3.3 Application: Modeling PCE Inflation via an ARMA Model

To illustrate the benefits of allowing for serially correlated disturbances, we revisit the PCE inflation example previously analyzed under the assumption of

independent errors. Using quarterly US PCE inflation over the sample period 1960Q1–2019Q4, we retain the same AR(2) specification for the conditional mean but replace the assumption of independent disturbances with an MA(1) error structure. This specification is a special case of the autoregressive moving average (ARMA) family of models.

Definition 3.1 (ARMA(p, q) Model)**.** *A univariate time series* $\{y_t\}$ *follows an* autoregressive moving average model *of order* (p, q)*, denoted ARMA*(p, q)*, if it satisfies*

$$y_t = \beta_1 + \beta_2 y_{t-1} + \cdots + \beta_{p+1} y_{t-p} + \varepsilon_t, \quad \varepsilon_t = u_t + \psi_1 u_{t-1} + \cdots + \psi_q u_{t-q},$$

where $\{u_t\}$ *is a sequence of iid random variables with mean zero and variance* σ^2*.*

While autoregressive terms capture persistence in the conditional mean, they do not in general account for all forms of serial dependence in a time series. In practice, approximating a covariance-stationary process using only autoregressive terms may require a large number of lags. By combining AR and MA components, ARMA models provide a more parsimonious and flexible approximation to a wide class of covariance-stationary time series.

To estimate the ARMA(2,1) model, we implement the three-block Metropolis-within-Gibbs sampler described earlier. We begin by introducing a supporting function, `loglike_MA1.m`, which evaluates the log-likelihood of the regression model with MA(1) errors.

The function takes three inputs: the scalar `psi`, corresponding to the MA(1) parameter ψ; a $T \times 1$ vector `e` containing the regression residuals $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ evaluated at the current draw of $\boldsymbol{\beta}$; and the innovation variance `sig2` $= \sigma^2$. Given these inputs, the function constructs the MA(1) band matrix $\mathbf{H}_\psi$ via

```
Hpsi = spdiags([psi*ones(T,1), ones(T,1)], [-1, 0], T, T);
```

which places ones on the main diagonal and ψ on the first subdiagonal, yielding a sparse, lower-triangular bidiagonal matrix that encodes the MA(1) dependence structure. The whitened residuals $\mathbf{u} = \mathbf{H}_\psi^{-1}\boldsymbol{\varepsilon}$ are then obtained by solving a banded linear system using the backslash operator, thereby avoiding explicit matrix inversion. The function returns both the resulting Gaussian log-likelihood and the vector $\mathbf{u}$.

```
function [loglik, u] = loglike_MA1(psi, e, sig2)
T = length(e);
Hpsi = spdiags([psi*ones(T,1), ones(T,1)], [-1, 0], T, T);
u = Hpsi\e;
loglik = -0.5*T*log(2*pi*sig2) - 0.5*(u'*u)/sig2;
```

```
end
```

Listing 3.4: Log-likelihood evaluation for the MA(1) regression model (`loglike_MA1.m`)

We next implement a random-walk MH update for the MA(1) parameter ψ in the function `sample_psi_RW.m`. A candidate value is generated by perturbing the current draw with Gaussian noise, and the proposal is accepted or rejected using the MH rule. The scalar tuning parameter `g_var` controls the variance of the random-walk proposal and therefore governs the size of the proposed moves. The invertibility constraint $|\psi| < 1$ is enforced by assigning zero acceptance probability to proposals outside the admissible region. Because the random-walk proposal is symmetric, the proposal densities cancel in the acceptance ratio, and the acceptance probability depends only on the change in the log-likelihood, which is evaluated using `loglike_MA1.m`.

```
function [psi, accept] = sample_psi_RW(psi, e, sig2, g_var)
psi_c = psi + sqrt(g_var)*randn; % propose candidate

% check invertibility (0.999 is a numerical buffer)
if abs(psi_c) < 0.999
    log_alpha = loglike_MA1(psi_c, e, sig2) ...
        - loglike_MA1(psi, e, sig2);
else
    log_alpha = -inf;
end

% accept/reject step
accept = (log(rand) < log_alpha);
if accept
    psi = psi_c;
end
end
```

Listing 3.5: Random-Walk Metropolis–Hastings update for the MA(1) parameter ψ (`sample_psi_RW.m`).

In the main MATLAB script `linreg_ma1.m`, we implement the three-block Metropolis-within-Gibbs sampler described in the previous section. Each MCMC iteration cycles through the following steps. First, conditional on (ψ, σ^2), we whiten the regression using the band matrix $\mathbf{H}_\psi$ and draw $\boldsymbol{\beta}$ from its Gaussian conditional posterior. Second, conditional on $(\boldsymbol{\beta}, \psi)$, we compute the whitened residuals $\mathbf{u} = \mathbf{H}_\psi^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and sample σ^2 from its inverse-gamma conditional distribution. Third, conditional on $(\boldsymbol{\beta}, \sigma^2)$, we update the MA(1) parameter ψ using a random-walk MH step with proposal variance `g_var`. After updating ψ, we reconstruct $\mathbf{H}_\psi$ so that the next iteration uses a covariance structure consistent with the current draw.

```
nsim = 50000; burnin = 1000;

% load data
data = readmatrix('USPCE.csv', 'Range', 'B2:B241');
y0 = data(1:4);  % initial conditions
y = data(5:end); % sample used for estimation
T = size(y,1);

% regressors for AR(2): [1, y_{t-1}, y_{t-2}]
xlag1 = [y0(4); y(1:end-1)];
xlag2 = [y0(3:4); y(1:end-2)];
X = [ones(T,1), xlag1, xlag2];
k = size(X,2);   % number of regressors

% prior hyperparameters (indep normal and inverse-gamma)
beta0 = zeros(k,1);  iVbeta = 1/100*speye(k);
nu0 = 4;  S0 = 1;

% initialize the Markov chain
psi = 0;
beta = (X'*X)\(X'*y);
sig2 = sum((y-X*beta).^2)/T;

Hpsi = spdiags([psi*ones(T,1),ones(T,1)],[-1, 0],T,T);
store_theta = zeros(nsim,k+2);  % [beta', sig2, psi]
count_psi = 0;
g_var = 0.02;  % proposal variance for RW-MH step for psi

for isim = 1:nsim + burnin
    % sample beta
    X_tilde = Hpsi\X;
    y_tilde = Hpsi\y;
    Dbeta = (iVbeta + X_tilde'*X_tilde/sig2)\speye(k);
    beta_hat = Dbeta*(iVbeta*beta0 + X_tilde'*y_tilde/sig2);
    beta = beta_hat + chol(Dbeta,'lower')*randn(k,1);

    % sample sig2
    e = y - X*beta;
    u = Hpsi\e;
    sig2 = 1/gamrnd(nu0+T/2,1/(S0 + u'*u/2));

    % sample psi
    [psi, accept] = sample_psi_RW(psi,e,sig2,g_var);
    Hpsi = spdiags([psi*ones(T,1),ones(T,1)],[-1, 0],T,T);

    % store the parameters
    if isim > burnin
        isave = isim - burnin;
        count_psi = count_psi + accept; % post-burnin
        store_theta(isave,:) = [beta', sig2, psi];
    end
end
% acceptance rate
accept_rate  = count_psi/nsim;
fprintf('Acceptance rate for psi: %.2f\n', accept_rate);
```

```
% Posterior summaries
theta_mean = mean(store_theta,1);
theta_lo   = quantile(store_theta,.025,1)';
theta_hi   = quantile(store_theta,.975,1)';
```

Listing 3.6: Posterior simulation of the ARMA(2,1) model (`linreg_ma1.m`).

Using $50,000$ posterior draws, the posterior means of the first- and second-order autoregressive coefficients are 1.11 and -0.165, respectively. The moving-average parameter ψ is clearly negative: its posterior mean is approximately -0.51, with a 95% credible interval of $(-0.78, -0.13)$ that excludes zero, and $\mathbb{P}(\psi < 0 \,|\, \mathbf{y}) = 0.991$. The left panel of Figure 3.3 displays the histogram of posterior draws of ψ. Because $\psi = 0$ corresponds to the benchmark model with conditionally independent regression errors, these results provide strong evidence against the independence assumption for this dataset. In Chapter 5, we formally compare the MA(1) specification with the standard linear regression model with independent errors using Bayesian model comparison tools.

Allowing for a moving-average component also affects the estimated dynamics of the autoregressive part. Relative to an AR(2) model with independent errors, incorporating MA(1) disturbances leads to a more persistent estimated AR component. Omitting the MA term forces the autoregressive coefficients to absorb this dependence, distorting the estimated persistence.

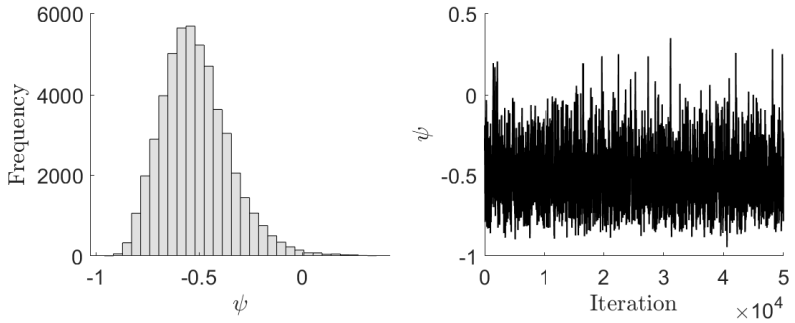

FIGURE 3.3: Posterior distribution and trace plot for the MA(1) parameter ψ. The left panel displays the histogram of posterior draws of ψ from the ARMA(2,1) model, while the right panel shows the corresponding trace plot from the MH update within the Gibbs sampler.

The random-walk MH update for ψ also exhibits satisfactory numerical performance. With the proposal variance `g_var` set to 0.02, the post-burn-in acceptance rate is approximately 42%, which is close to the commonly recommended range of 20%–40% for univariate random-walk samplers. The trace plot in Figure 3.3 shows no evidence of long stretches in which the chain

remains at a constant value, indicating that the sampler does not become stuck and mixes reasonably well.

3.4 Further Reading

The idea of introducing latent or missing data to simplify inference dates back to the Expectation-Maximization algorithm of Dempster, Laird and Rubin (1977), who show that augmenting the observed data can greatly facilitate maximum likelihood estimation through iterative optimization. In a Bayesian setting, Tanner and Wong (1987) formalize this principle under the name data augmentation and demonstrate how latent variables can be exploited to construct efficient posterior samplers.

The impact of this idea expanded rapidly with the advent of Gibbs sampling, most notably in the work of Albert and Chib (1993) for probit models and Geweke (1993) for linear regression with Student-t errors. These contributions firmly establish data-augmented Gibbs samplers as a central tool in modern Bayesian econometrics. An overview is provided by van Dyk and Meng (2001), who synthesize the development of data augmentation since Tanner and Wong (1987), including marginal augmentation and working-parameter strategies that improve algorithmic efficiency.

Further examples of expressing non-Gaussian models as conditionally Gaussian via mixture representations are discussed in Chapter 4. Bayesian analysis of linear regression models with autoregressive moving average errors is developed by Chib and Greenberg (1994), who introduce the idea of "whitening" the error process to obtain conditionally independent disturbances, implemented via recursive filtering. Computational refinements based on band-matrix methods are later proposed by Chan (2013).

Exercises

Exercise 3.1 (Scale-Mixture Representation of the Student-t Distribution)**.** Let

$$(\varepsilon \mid \lambda) \sim \mathcal{N}(0, \lambda\,\sigma^2), \quad \lambda \sim \mathcal{IG}(\nu/2, \nu/2).$$

Derive the joint density of (ε, λ) and, by integrating out λ, show that the marginal density of ε is the Student-t distribution $\mathcal{T}_\nu(0, \sigma^2)$.

Exercise 3.2 (Full Conditional Distributions in the Student-t Regression)**.** Consider the linear regression $y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$ with Student-t errors in the scale-mixture form $(\varepsilon_t \,|\, \lambda_t, \sigma^2) \sim \mathcal{N}(0, \lambda_t\sigma^2)$ and $(\lambda_t \,|\, \nu) \sim \mathcal{IG}(\nu/2, \nu/2)$, together with the priors $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$. Let $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_T)$.

(a) By combining the conditionally Gaussian likelihood with the inverse-gamma prior and collecting terms in σ^2, show that

$$(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\lambda}) \sim \mathcal{IG}\left(\nu_0 + \frac{T}{2},\, S_0 + \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Lambda}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

(b) Show that the latent scales are conditionally independent across t, with

$$(\lambda_t \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2) \sim \mathcal{IG}\left(\frac{\nu+1}{2},\, \frac{1}{2}\left(\nu + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\sigma^2}\right)\right), \quad t = 1, \ldots, T.$$

(c) Using the result in part (b), show that

$$\mathbb{E}(\lambda_t^{-1} \,|\, \mathbf{y}, \boldsymbol{\beta}, \sigma^2) = \frac{\nu+1}{\nu + (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2/\sigma^2},$$

and explain how this expression formalizes the way a large standardized residual downweights observation t in the update of $\boldsymbol{\beta}$.

Exercise 3.3 (Inverse-Transform Sampling for a Discrete Distribution)**.** Let X be a discrete random variable with support $\{1, 2, \ldots, n\}$ and probabilities $\mathbb{P}(X = i) = p_i$, where $p_i > 0$ and $\sum_{i=1}^n p_i = 1$.

(a) Write a script that takes as inputs the probability vector $\mathbf{p} = (p_1, \ldots, p_n)'$ and an integer R, and returns R independent draws $x^{(1)}, \ldots, x^{(R)}$ from this distribution using the inverse-transform method.

(b) Verify your implementation by simulating a large number of draws and comparing the empirical frequencies $\widehat{\mathbb{P}}(X = i)$ with the target probabilities p_i.

Exercise 3.4 (Inverse-Transform Sampling for a Truncated Normal Distribution)**.** Consider the truncated normal random variable $X \sim \mathcal{TN}_{(a,b)}(\mu, \sigma^2)$ restricted to the interval (a, b), where $-\infty \le a < b \le \infty$.

(a) Show that the cdf of X can be written as

$$F_X(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad a < x < b,$$

where $\Phi(\cdot)$ denotes the standard normal cdf.

(b) Derive an inverse-transform sampling rule for X. In particular, show that if $U \sim \mathcal{U}(0,1)$ and

$$V = \Phi\left(\frac{a-\mu}{\sigma}\right) + U\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right],$$

then

$$X = \mu + \sigma\,\Phi^{-1}(V)$$

has the truncated normal distribution on (a, b).

(c) Write a script that generates R independent draws from the truncated normal distribution using the inverse-transform method. Your script should accept (μ, σ^2, a, b, R) as inputs and return the simulated draws.

(d) For the special case $\mu = 0$, $\sigma^2 = 1$, $a = 0$, and $b = \infty$, simulate draws and compare the empirical mean and variance with their theoretical values.

Exercise 3.5 (Density Forecast under Student-t versus Gaussian Errors)**.** Consider the AR(2) model for quarterly US PCE inflation

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \varepsilon_t,$$

estimated on the sample 1960Q1–2019Q4.

(a) Using posterior draws from the Student-t model obtained with `linreg_t.m`, construct a one-step-ahead density forecast for the inflation in 2020Q1 by Monte Carlo simulation. Plot the resulting predictive density.

(b) Construct the corresponding density forecast under the Gaussian AR(2) model estimated in Section 2.3.2. Compare the two predictive densities, focusing on their dispersion and tail behavior, and briefly discuss the implications for assessing downside risk.

Exercise 3.6 (Independence-Chain MH Update for ψ in an ARMA$(2,1)$ Model)**.** Consider the ARMA$(2,1)$ model for quarterly PCE inflation studied in Section 3.3.3. Conditional on current Gibbs draws of $(\boldsymbol{\beta}, \sigma^2)$, let

$$f(\psi) \propto p(\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2, \psi)\, \mathbf{1}\{|\psi| < 1\}$$

denote the conditional target density for the MA parameter ψ.

(a) Write a script to obtain the mode $\widehat{\psi}$ of $\log f(\psi)$ numerically over $(-1, 1)$ by minimizing the negative log-density (e.g., using `fminbnd`). Approximate the second derivative of $\log f(\psi)$ at $\widehat{\psi}$ using a centered finite-difference method with step size $\Delta > 0$, and use it to construct a Gaussian proposal density with mean $\widehat{\psi}$ and variance

$$D_\psi = -\left(\frac{\partial^2}{\partial \psi^2} \log f(\psi)\Big|_{\psi=\widehat{\psi}}\right)^{-1}.$$

(b) Using $g(\psi)$ as an independence proposal, implement a Metropolis–Hastings update for ψ, enforcing the invertibility restriction $|\psi| < 1$.

(c) Incorporate this independence-chain MH step into the Gibbs sampler for the ARMA$(2, 1)$ model and replicate the PCE inflation application. Report posterior summaries for ψ, the acceptance rate of the MH step, and briefly compare the numerical performance with the random-walk MH sampler used in Section 3.3.3.

Exercise 3.7 (Posterior Sampling for an MA(2) Regression)**.** This exercise extends the MA(1) regression in Section 3.3 to a higher-order moving average error process and applies the resulting sampler to US PCE inflation. Consider

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2},$$

where the u_t are iid $\mathcal{N}(0, \sigma^2)$. Assume a uniform prior on $\boldsymbol{\psi} = (\psi_1, \psi_2)'$ over the invertibility region of the MA(2) model, given by $-1 < \psi_2 < 1$, $\psi_1 + \psi_2 > -1$, and $\psi_2 - \psi_1 > -1$.

(a) Write the model in matrix form $\boldsymbol{\varepsilon} = \mathbf{H}_{\boldsymbol{\psi}}\mathbf{u}$, where $\mathbf{H}_{\boldsymbol{\psi}}$ is a lower-triangular band matrix with three nonzero diagonals. Explain how to evaluate the likelihood efficiently.

(b) Construct a Metropolis-within-Gibbs sampler for $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\psi})$, using conjugate updates for $(\boldsymbol{\beta}, \sigma^2)$ and an MH step for $\boldsymbol{\psi}$.

(c) Fit both ARMA$(2, 1)$ and ARMA$(2, 2)$ models to quarterly US PCE inflation, 1960Q1–2019Q4, and briefly discuss which specification is more appropriate.

4

Mixture Models

So far we have focused on linear regression models with Gaussian innovations, where posterior computation is straightforward. Economic and financial data, however, often exhibit non-Gaussian features such as heavy tails, skewness, or latent heterogeneity. In this chapter we introduce two important classes of flexible specifications—*location-scale mixtures of normals* and *finite mixtures of normals*—which greatly broaden the range of distributions we can model. Both admit data-augmentation representations that make the models conditionally Gaussian, so estimation requires only minor modifications of the Gibbs samplers developed earlier.

4.1 Location-Scale Mixture of Normals

We begin by introducing the basic constructions underlying scale and location-scale mixtures of normals and illustrating them with several canonical examples.

4.1.1 Basic Constructions and Examples

Many familiar non-Gaussian distributions can be expressed as *scale mixtures of normals*. Formally, we have the following definition.

Definition 4.1 (Scale Mixture of Normals)**.** *A random variable X is said to follow a* scale mixture of normals *if it admits the hierarchical representation*

$$\begin{aligned}(X \mid \lambda, \sigma^2) &\sim \mathcal{N}(0, \sigma^2\lambda),\\ \lambda &\sim G,\end{aligned}$$

where G is a distribution supported on the positive real line $\mathbb{R}_+$.

We encountered one such example in Chapter 3, where a Student-t distribution was represented as a scale mixture of normals with $(\lambda \mid \nu) \sim$

$\mathcal{IG}(\nu/2, \nu/2)$. Scale mixtures provide a simple and powerful mechanism for introducing heavy tails while preserving conditional Gaussianity.

Example 4.1 (Laplace Distribution as a Scale Mixture of Normals)**.** As another illustration, suppose the scale-mixing variable λ follows a gamma distribution $\mathcal{G}(1, 1/2)$ (equivalently, an exponential distribution with rate parameter $1/2$). In this case, the marginal distribution of X is the *Laplace* (or *double-exponential*) distribution with scale parameter $\sigma > 0$, whose density is given by

$$f(x) = \frac{1}{2\sigma} \mathrm{e}^{-\frac{|x|}{\sigma}}. \tag{4.1}$$

The derivation of this result is given in Exercise 4.2.

Figure 4.1 compares the density of the Laplace distribution with those of the standard normal and Student-t distributions. All three distributions are standardized to have mean zero and unit variance, and the degrees-of-freedom parameter for the Student-t distribution is set to $\nu = 3$.

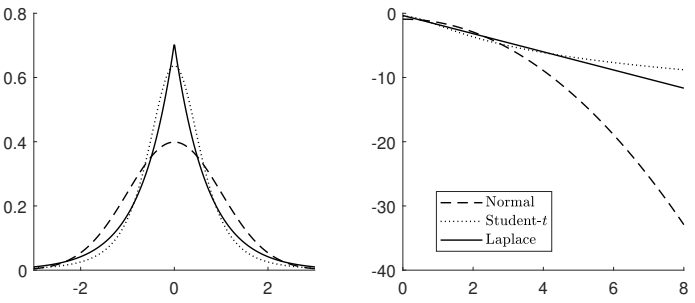

FIGURE 4.1: Density functions (left panel) and log densities (right panel) of the standard normal distribution, the Student-t distribution with $\nu = 3$, and the Laplace distribution.

As shown in the left panel of Figure 4.1, relative to the normal distribution, the Laplace distribution assigns substantially more mass near zero. At the same time, its tails decay more slowly than those of the normal distribution but more rapidly than those of the Student-t distribution. This reflects the differing tail behaviors: the Laplace density decays at rate $\exp(-c_1|x|)$, compared with $\exp(-c_2x^2)$ for the normal distribution and $|x|^{-c_3}$ for the Student-t distribution, where c_1, c_2, and c_3 are positive constants.

Scale mixtures generate heavy tails (and often substantial mass near zero), but they cannot produce *skewness* because the conditional Gaussian compo-

nent is centered at zero. A natural extension is obtained by allowing the latent variable to affect not only the conditional variance but also the conditional mean. This leads to the class of *location-scale mixtures of normals*, also known as *normal mean-variance mixtures*.

Definition 4.2 (Location-Scale Mixture of Normals)**.** *A random variable X is said to follow a* location-scale mixture of normals *if it admits the hierarchical representation*

$$\begin{aligned}(X \mid \boldsymbol{\lambda}) &\sim \mathcal{N}\big(m(\boldsymbol{\lambda}), s^2(\boldsymbol{\lambda})\big),\\ \boldsymbol{\lambda} &\sim G,\end{aligned}$$

where $\boldsymbol{\lambda}$ is a latent random vector, G is a mixing distribution, and the functions $m(\cdot)$ and $s^2(\cdot)$ specify the conditional mean and variance, respectively.

The scale-mixture representation is obtained as a special case when the conditional mean function satisfies $m(\boldsymbol{\lambda}) \equiv 0$. More generally, allowing $m(\boldsymbol{\lambda})$ to vary with $\boldsymbol{\lambda}$ introduces asymmetry into the marginal distribution of X. Despite this added flexibility, the model remains conditionally Gaussian given the latent variables $\boldsymbol{\lambda}$.

Example 4.2 (Skew-Normal Distribution as a Location Mixture of Normals)**.** One convenient way to introduce skewness is through a latent truncated normal variable. Let $\lambda \sim \mathcal{TN}_{(0,\infty)}(0,1)$, that is, λ follows a standard normal distribution truncated to $(0,\infty)$. Conditional on λ, define

$$(X \mid \lambda) \sim \mathcal{N}(\delta\lambda,\ 1-\delta^2),$$

where $\delta \in (-1,1)$ controls the degree of skewness. Marginally, X follows a skew-normal distribution in the sense of Azzalini (1985).

This representation is a *location mixture* of normals: the latent variable λ shifts the conditional mean while the conditional variance remains constant. As a result, the marginal distribution of X is asymmetric, while the model retains conditional Gaussianity.

Example 4.3 (Skew-t Distribution as a Location-Scale Mixture of Normals)**.** Heavy tails can be introduced on top of skewness by combining a location-mixture representation with a scale-mixture representation. Let

$$\begin{aligned}\lambda_1 &\sim \mathcal{TN}_{(0,\infty)}(0,1),\\ (\lambda_2 \mid \nu) &\sim \mathcal{IG}(\nu/2, \nu/2),\end{aligned}$$

and define the conditional distribution

$$(X \mid \lambda_1, \lambda_2) \sim \mathcal{N}\big(\delta\,\lambda_1\sqrt{\lambda_2}, (1-\delta^2)\lambda_2\big),$$

for $\delta \in (-1, 1)$. Marginally, X follows a skew-t distribution in the sense of Azzalini and Capitanio (2003). In this formulation the latent vector (λ_1, λ_2) governs both the conditional mean and conditional variance, and therefore defines a location-scale mixture of normals.

4.1.2 Bayesian Quantile Regression

Location-scale mixtures are commonly used to model the innovations in linear regression models. For example, Section 3.2.4 analyzes PCE inflation using an AR model with Student-t innovations. In such settings, for a linear regression model with outcome y_t and regressors $\mathbf{x}_t$, the linear predictor $\mathbf{x}_t'\boldsymbol{\beta}$ is interpreted as the conditional mean of y_t given $\mathbf{x}_t$.

In many applications, however, interest lies in the tails of the conditional distribution rather than its center. Quantities such as downside risk or distributional asymmetry are more naturally described by conditional quantiles than by the conditional mean. This motivates *quantile regression*, which replaces the conditional-mean restriction with a restriction on a specified conditional quantile.

Formally, for a fixed $\tau \in (0, 1)$, quantile regression specifies the conditional τ-quantile of y_t given $\mathbf{x}_t$ as $Q_\tau(y_t \,|\, \mathbf{x}_t, \boldsymbol{\beta}_\tau) = \mathbf{x}_t'\boldsymbol{\beta}_\tau$. Here the coefficient vector $\boldsymbol{\beta}_\tau$ is allowed to vary with τ, thereby capturing heterogeneous covariate effects across different regions of the conditional distribution.

Quantile regression can be viewed as a change in the target of inference rather than in the functional form of the linear predictor. Instead of imposing an error distribution centered at zero in the mean, one specifies an error distribution whose τ-quantile is equal to zero. If this condition holds, then in the linear regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t,$$

the conditional τ-quantile of y_t given $\mathbf{x}_t$ is $Q_\tau(y_t \,|\, \mathbf{x}_t, \boldsymbol{\beta}) = \mathbf{x}_t'\boldsymbol{\beta}$. In principle, any asymmetric distribution satisfying this property could be used as a likelihood for Bayesian quantile regression. This is convenient in a Bayesian setting, where likelihood-based inference requires an explicit distributional assumption for the regression innovations.

A convenient and widely used choice is the *asymmetric Laplace distribution*, whose use as a likelihood for Bayesian quantile regression is introduced by Yu and Moyeed (2001). We write $\varepsilon \sim \mathcal{AL}(0, \sigma^2, \tau)$ if ε has density

$$\begin{aligned} f(\varepsilon) &= \frac{\tau(1-\tau)}{\sigma^2} \mathrm{e}^{-\frac{1}{\sigma^2}\rho_\tau(\varepsilon)}, \\ \rho_\tau(\varepsilon) &= \varepsilon\big(\tau - \mathbf{1}\{\varepsilon < 0\}\big), \end{aligned} \tag{4.2}$$

where $\tau \in (0, 1)$ is the quantile index and $\sigma^2 > 0$ is a scale parameter. The

function $\rho_\tau(\cdot)$ is the *check loss* from classical quantile regression. As verified in Exercise 4.3, the τ-quantile of $\mathcal{AL}(0, \sigma^2, \tau)$ is zero, which makes it a natural likelihood for quantile regression at level τ.

The asymmetric Laplace distribution offers two further advantages. First, its log-likelihood is proportional to $-\sum_{t=1}^{T} \rho_\tau(y_t - \mathbf{x}_t'\boldsymbol{\beta}_\tau)$, so that under a flat prior the posterior mode of $\boldsymbol{\beta}_\tau$ coincides with the Koenker–Bassett quantile regression estimator (Koenker and Bassett, 1978). Second, it admits a location-scale mixture of normals representation, which restores conditional conjugacy and renders posterior computation particularly simple. As a result, Bayesian quantile regression under the asymmetric Laplace distribution can be formulated as a conditionally Gaussian model with latent variables, fitting squarely into the mixture framework developed earlier in this chapter.

Definition 4.3 (Bayesian Quantile Regression)**.** *For a fixed quantile index $\tau \in (0, 1)$, the* Bayesian quantile regression model *is defined by*

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_\tau + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{AL}(0, \sigma^2, \tau),$$

where $\boldsymbol{\beta}_\tau$ is the vector of quantile-specific regression coefficients and $\sigma^2 > 0$ is the scale parameter of the asymmetric Laplace distribution.

Direct posterior sampling under the asymmetric Laplace likelihood is not straightforward. To exploit the mixture representation, we follow Kozumi and Kobayashi (2011) and define the constants

$$\vartheta_\tau = \frac{1 - 2\tau}{\tau(1 - \tau)}, \quad \varphi_\tau = \frac{2}{\tau(1 - \tau)}. \tag{4.3}$$

Consider the hierarchical model

$$\begin{aligned} (\varepsilon \,|\, \lambda, \sigma^2) &\sim \mathcal{N}\big(\vartheta_\tau \lambda,\ \varphi_\tau \sigma^2 \lambda\big), \\ (\lambda \,|\, \sigma^2) &\sim \mathcal{G}(1, 1/\sigma^2), \end{aligned} \tag{4.4}$$

where $\mathcal{G}$ denotes a gamma distribution. Marginally of λ, ε follows the asymmetric Laplace distribution defined in (4.2). When $\tau = 1/2$, $\vartheta_\tau = 0$, and the asymmetric Laplace distribution reduces—up to a reparameterization of the scale—to the symmetric Laplace distribution illustrated in Example 4.1.

We now adopt the quantile regression model in Definition 4.3 and employ the latent-variable representation of the asymmetric Laplace distribution in (4.4):

$$\begin{aligned} (y_t \,|\, \lambda_t, \boldsymbol{\beta}_\tau, \sigma^2) &\sim \mathcal{N}\big(\mathbf{x}_t'\boldsymbol{\beta}_\tau + \vartheta_\tau \lambda_t, \varphi_\tau \sigma^2 \lambda_t\big), \\ (\lambda_t \,|\, \sigma^2) &\sim \mathcal{G}(1, 1/\sigma^2). \end{aligned}$$

We assume independent priors $\boldsymbol{\beta}_\tau \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma^2 \sim \mathcal{IG}(\nu_0, S_0)$. Let

$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_T)$. Define the shifted response $\widetilde{y}_t = y_t - \vartheta_\tau \lambda_t$, and stack $\widetilde{\mathbf{y}} = \mathbf{y} - \vartheta_\tau \boldsymbol{\lambda}$. Conditional on $(\boldsymbol{\lambda}, \sigma^2)$, we then have

$$(\widetilde{\mathbf{y}} \,|\, \boldsymbol{\beta}_\tau, \boldsymbol{\lambda}, \sigma^2) \sim \mathcal{N}\big(\mathbf{X}\boldsymbol{\beta}_\tau, \varphi_\tau \sigma^2 \boldsymbol{\Lambda}\big).$$

This is a heteroskedastic Gaussian regression model and therefore admits a straightforward three-block Gibbs sampler. The resulting Gibbs sampler cycles through the following three steps.

First, applying Theorem 3.1 yields

$$(\boldsymbol{\beta}_\tau \,|\, \mathbf{y}, \boldsymbol{\lambda}, \sigma^2) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}_\tau, \mathbf{D}_{\boldsymbol{\beta}}),$$

where

$$\mathbf{D}_{\boldsymbol{\beta}} = \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \frac{1}{\varphi_\tau \sigma^2} \mathbf{X}' \boldsymbol{\Lambda}^{-1} \mathbf{X} \right)^{-1}, \quad \widehat{\boldsymbol{\beta}}_\tau = \mathbf{D}_{\boldsymbol{\beta}} \left(\mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\beta}_0 + \frac{1}{\varphi_\tau \sigma^2} \mathbf{X}' \boldsymbol{\Lambda}^{-1} \widetilde{\mathbf{y}} \right).$$

Next, combining the Gaussian likelihood with the gamma mixing density yields an inverse-gamma full conditional distribution for σ^2:

$$(\sigma^2 \,|\, \mathbf{y}, \boldsymbol{\beta}_\tau, \boldsymbol{\lambda}) \sim \mathcal{IG}\left(\nu_0 + \frac{3T}{2}, \widehat{S} \right),$$

where $\widehat{S} = S_0 + \sum_{t=1}^{T} \lambda_t + \frac{1}{2\varphi_\tau} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_\tau - \vartheta_\tau \boldsymbol{\lambda})' \boldsymbol{\Lambda}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_\tau - \vartheta_\tau \boldsymbol{\lambda})$.

Finally, $\{\lambda_t\}$ are conditionally independent given $(\mathbf{y}, \boldsymbol{\beta}_\tau, \sigma^2)$. The full conditional kernel for λ_t can be written as

$$p(\lambda_t \,|\, \mathbf{y}, \boldsymbol{\beta}_\tau, \sigma^2) \propto \lambda_t^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(a_\tau \lambda_t + \frac{b_{t,\tau}}{\lambda_t} \right)},$$

where

$$a_\tau = \frac{\vartheta_\tau^2 + 2\varphi_\tau}{\varphi_\tau \sigma^2}, \quad b_{t,\tau} = \frac{(y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau)^2}{\varphi_\tau \sigma^2}. \tag{4.5}$$

Although this kernel does not correspond to a standard distribution for λ_t, a simple transformation yields a convenient form. In particular, the inverse λ_t^{-1} follows an inverse Gaussian distribution; see Exercise 4.4 for the derivation and details. Accordingly, we draw $z_t \sim \mathcal{IGAUSS}(a_\tau, \mu_{t,\tau})$ and set $\lambda_t = z_t^{-1}$, where

$$\mu_{t,\tau} = \sqrt{\frac{a_\tau}{b_{t,\tau}}} = \frac{\sqrt{\vartheta_\tau^2 + 2\varphi_\tau}}{|y_t - \mathbf{x}_t' \boldsymbol{\beta}_\tau|}.$$

An efficient routine for sampling from the inverse Gaussian distribution is provided in the MATLAB function `igaussrnd.m`, which can be downloaded from the companion website for this book and reused directly here.

4.1.3 Application: Growth-at-Risk

We illustrate Bayesian quantile regression in an application to *growth-at-risk* (GaR), as proposed by Adrian, Boyarchenko and Giannone (2019). The objective is to quantify how financial conditions affect downside risks to future real GDP growth. Following Adrian, Boyarchenko and Giannone (2019), we focus on two key US macro-financial series: real GDP growth and the Chicago Fed National Financial Conditions Index (NFCI). The NFCI summarizes information from a large set of credit, funding, and risk-premia indicators; positive values indicate tighter-than-average financial conditions, whereas negative values indicate looser-than-average conditions.

The sample period begins in 1971Q1, the first quarter for which the NFCI is available, and ends in 2024Q4. Real GDP growth is observed at a quarterly frequency, whereas the NFCI is a weekly time series. To align frequencies, we compute the average NFCI within each quarter.

Growth-at-risk is defined as a low conditional quantile of future real GDP growth and is a quantile-based forecast. Fix a forecast horizon h (e.g., $h = 4$ quarters), and let g_{t+h} denote quarterly annualized real GDP growth observed h quarters ahead. Given predictors observed at time t, collected in $\mathbf{x}_t$, we define

$$\mathsf{GaR}_t(h;\tau) = Q_\tau(g_{t+h} \,|\, \mathbf{x}_t), \quad \tau \in (0,1),$$

where $\tau = 0.05$ is commonly used as a measure of downside risk.

We consider the parsimonious quantile regression specification

$$g_{t+h} = \mathbf{x}_t'\boldsymbol{\beta}_{\tau,h} + \varepsilon_{t+h}, \quad \varepsilon_{t+h} \sim \mathcal{AL}(0, \sigma^2_{\tau,h}, \tau), \tag{4.6}$$

where $\mathbf{x}_t$ includes a constant, current real GDP growth, and the NFCI. Because the asymmetric Laplace distribution has τ-quantile equal to zero (see Exercise 4.3), the linear predictor $\mathbf{x}_t'\boldsymbol{\beta}_{\tau,h}$ coincides with the conditional τ-quantile of g_{t+h} given $\mathbf{x}_t$.

For each (τ, h) pair, we estimate (4.6) using the Gibbs sampler developed in Section 4.1.2, implemented in the script `linreg_bqr.m`. In particular, we use the function `igaussrnd.m` to sample from the inverse Gaussian distribution; this routine can be downloaded from the companion website for this book. For posterior draw r, the implied growth-at-risk is

$$\mathsf{GaR}_t^{(r)}(h;\tau) = \mathbf{x}_t'\boldsymbol{\beta}_{\tau,h}^{(r)}.$$

We summarize $\mathsf{GaR}_t(h;\tau)$ using posterior means and pointwise credible intervals constructed from $\{\mathsf{GaR}_t^{(r)}(h;\tau)\}_{r=1}^R$.

```
nsim = 20000; burnin = 1000;
tau = 0.05;  % quantile level (e.g., 0.05, 0.10, 0.50)
```

```
h = 4;      % forecast horizon in quarters (e.g., 1 or 4)

% AL mixture constants
vartheta = (1 - 2*tau) / (tau*(1 - tau));
varphi = 2 / (tau*(1 - tau));

% load GDP-NFCI merged data
%   DATE is a quarter-start date
%   GDP is quarterly (annualized) real GDP growth
%   NFCI is quarterly average of weekly NFCI
opts = detectImportOptions('GDP_NFCI_merged.csv');
opts = setvaropts(opts,'Date','InputFormat','MM/dd/uuuu');
TBL = readtable('GDP_NFCI_merged.csv', opts);

% keep observations with NFCI available
idx = ~isnan(TBL.NFCI) & ~isnan(TBL.GDP);
TBL = TBL(idx,:);
gdp  = TBL.GDP;
nfci = TBL.NFCI;

% Construct regression for Growth-at-Risk:
%   y_{t+h} = beta0 + beta1*gdp_t + beta2*nfci_t + eps_{t+h}
T0 = length(gdp);
T  = T0 - h;
y = gdp(1+h:end); % y_{t+h}
X = [ones(T,1), gdp(1:end-h), nfci(1:end-h)];
k = size(X,2);

% prior hyperparameters (indep normal and inverse-gamma)
beta0 = zeros(k,1);
iVbeta = speye(k)/100;  % prior precision
nu0 = 3;
S0 = 1*(nu0 - 1);

% initialize the Markov chain
beta = (X'*X)\(X'*y);
sig2 = mean((y - X*beta).^2);
lam = gamrnd(1, sig2, T, 1);
iLam   = sparse(1:T,1:T,1./lam);
store_theta = zeros(nsim, k+1);   % [beta' sig2]

% Gibbs sampler starts here
for isim = 1:nsim + burnin
    % sample beta
    ytilde = y - vartheta*lam;
    Dbeta = (iVbeta + (X'*iLam*X)/(varphi*sig2))\speye(k);
    beta_hat = Dbeta*(iVbeta*beta0 ...
        + (X'*iLam*ytilde)/(varphi*sig2));
    C = chol(Dbeta,'lower');
    beta = beta_hat + C*randn(k,1);

    % sample sig2
    e = y - X*beta - vartheta*lam;
    S_hat = S0 + sum(lam) + e'*iLam*e/(2*varphi);
    sig2  = 1/gamrnd(nu0 + 3*T/2, 1/S_hat);

    % sample lambda
```

```
    a_tau = (vartheta^2 + 2*varphi)/(varphi*sig2);
    mu = sqrt(vartheta^2 + 2*varphi) ./ abs(y - X*beta);
    lam = 1./igaussrnd(a_tau*ones(T,1), mu);
    iLam = sparse(1:T,1:T,1./lam);

    if isim > burnin
        isave = isim - burnin;
        store_theta(isave,:) = [beta' sig2];
    end
end
theta_mean = mean(store_theta,1)';
theta_CI  = quantile(store_theta,[.025 .975]);

disp('Posterior mean (beta; sig2):');
disp(theta_mean);
disp('Posterior 95% CI (rows: 2.5%, 97.5%):');
disp(theta_CI);

% compute GaR_t(h; tau)
B = store_theta(:,1:k);
Xgar = X; % regressors at time t (aligned with y_{t+h})
GaR_draws = Xgar * B';  %  dimension is T x nsim
GaR_mean = mean(GaR_draws, 2);
```

Listing 4.1: Gibbs sampler for Bayesian quantile regression in the growth-at-risk application with horizon h (`linreg_bqr.m`).

The growth-at-risk estimates provide a direct measure of macroeconomic vulnerability. In particular, when the NFCI is elevated, indicating tight financial conditions, the lower conditional quantiles of future GDP growth shift downward. Figure 4.2 plots the estimated growth-at-risk for $h = 4$ and $\tau = 0.05$.

Plotted at the forecast date $t + h$, the estimated growth-at-risk series declines sharply during periods of financial stress and NBER recessions, whereas the median growth forecast remains comparatively stable. At the four-quarter horizon, the 5th-percentile growth-at-risk typically lies between -1% and -2% in tranquil periods, but deteriorates markedly during downturns. For example, it falls below -13% in the 1974–75 recession, remains in the range of -8% to -10% in the early 1980s, declines to about -9% during the Great Recession, and drops to roughly -6% at the onset of the COVID-19 recession.

By contrast, the median four-quarter-ahead growth forecast stays close to 2.5%–3% throughout the sample. This implies that the gap between the center and the left tail of the predictive distribution widens substantially during periods of financial stress. This pronounced asymmetry suggests that financial conditions primarily affect downside risk rather than typical growth outcomes, consistent with the vulnerable-growth mechanism emphasized by Adrian, Boyarchenko and Giannone (2019).

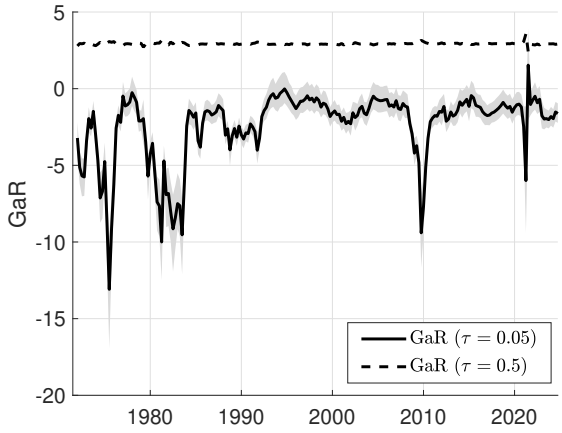

FIGURE 4.2: Growth-at-risk estimates at horizon $h = 4$ for $\tau = 0.05$ and $\tau = 0.50$. Shaded regions denote 95% pointwise credible intervals for the lower-tail estimate.

Table 4.1 reports posterior estimates of the Bayesian quantile regression coefficients at horizon $h = 4$. The results exhibit a clear asymmetry across quantiles. At the lower tail of the distribution ($\tau = 0.05$), the coefficient on current GDP growth is small and statistically insignificant, while the coefficient on the NFCI is large, negative, and precisely estimated, with a posterior mean of -2.13 and a 95% credible interval that excludes zero. A similar pattern holds at $\tau = 0.10$, indicating that financial conditions robustly affect downside risk across low quantiles.

TABLE 4.1: Posterior estimates from Bayesian quantile regression at horizon $h = 4$.

	GDP growth		Financial conditions (NFCI)	
τ	Posterior mean	95% C.I.	Posterior mean	95% C.I.
0.05	0.11	$(-0.02,\ 0.19)$	-2.13	$(-2.99,\ -1.47)$
0.10	-0.01	$(-0.19,\ 0.15)$	-1.91	$(-2.58,\ -1.28)$
0.50	-0.02	$(-0.12,\ 0.09)$	-0.01	$(-0.58,\ 0.54)$

By contrast, at the median ($\tau = 0.50$), both coefficients are economically small and statistically insignificant. In particular, the NFCI has essentially no effect on typical growth outcomes at the four-quarter horizon.

4.2 Finite Mixture of Normals

In the previous section we studied scale and location-scale mixtures of normals. In those models, continuous latent variables modify the mean and/or variance of a single Gaussian distribution to generate heavy tails or skewness. A different approach is the *finite mixture of normal distributions*. Instead of introducing continuous latent variation within one component, a finite normal mixture assumes that the data arise from a finite set of unobserved regimes or subpopulations, each governed by its own normal component. These components are combined through mixture weights. This discrete form of heterogeneity is useful for modeling outliers, latent subpopulations, and clustered dynamics in macroeconomic and financial time series.

To keep the discussion concrete, consider the standard linear regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \quad (\varepsilon_t \,|\, \sigma^2) \sim \mathcal{N}(0, \sigma^2).$$

Under this specification, each observation y_t is drawn from a single Gaussian distribution with mean $\mathbf{x}_t'\boldsymbol{\beta}$ and variance σ^2. A natural generalization is to allow observations to be generated from more than one Gaussian component. This leads to the following definition.

Definition 4.4 (Finite Mixture of Normals)**.** *Let M be a fixed positive integer. A random variable y_t is said to follow a* finite mixture of normals *if its conditional density can be written as*

$$p(y_t \,|\, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) = \sum_{m=1}^{M} w_m f_{\mathcal{N}}(y_t; \mathbf{x}_t'\boldsymbol{\beta}_m, \sigma_m^2), \tag{4.7}$$

where $\boldsymbol{\beta}_m$ and σ_m^2 are the regression coefficients and variance of component m, $\mathbf{w} = (w_1, \ldots, w_M)'$ is the vector of mixture weights satisfying $w_m > 0$ and $\sum_{m=1}^{M} w_m = 1$, and $f_{\mathcal{N}}(\cdot\,; a, b^2)$ denotes the Gaussian density with mean a and variance b^2. Under this representation, y_t is drawn from the mth normal component $\mathcal{N}(\mathbf{x}_t'\boldsymbol{\beta}_m, \sigma_m^2)$ with probability w_m.

By allowing the regression coefficients and variances to differ across components, finite mixtures can capture discrete shifts and outliers, as well as distinct regimes such as expansions versus recessions or periods of low versus high volatility. Because the component indicators in (4.7) are independent across time, however, the model captures discrete heterogeneity rather than persistent regimes; capturing persistence requires additional structure, such as Markov-switching component indicators with serially dependent transition probabilities.

In the formulation (4.7), the number of mixture components M is assumed

to be known. This choice can be guided by Bayesian model comparison, for example using marginal likelihoods (see Chapter 5). Alternatively, M may be treated as unknown and inferred using reversible jump MCMC, which permits transitions between models with different numbers of components (Richardson and Green, 1997). Chapter 8 discusses infinite mixture models, which eliminate the need to fix M in advance by allowing the data to determine the effective number of components.

4.2.1 Posterior Sampling via Data Augmentation

For estimation, the mixture density in (4.7) is more challenging to work with than a single-component Gaussian likelihood. Even when normal-inverse-gamma priors are placed on the component-specific regression coefficients and variances, the posterior distribution does not factor into standard forms because each observation may arise from any mixture component. As a result, direct posterior sampling is not straightforward.

To address this difficulty, we again employ a data-augmentation strategy. Specifically, we introduce a discrete latent variable $s_t \in \{1, \dots, M\}$ indicating the mixture component from which observation y_t is drawn. Conditional on the mixture weights $\mathbf{w}$, the latent indicators satisfy $\mathbb{P}(s_t = m \,|\, \mathbf{w}) = w_m$ and are independent across t. The variables s_t are commonly referred to as *component indicators* or *component labels*. Conditional on the label s_t, the observation y_t is generated from the corresponding normal distribution:

$$(y_t \,|\, s_t, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) \sim \mathcal{N}(\mathbf{x}_t' \boldsymbol{\beta}_{s_t}, \sigma^2_{s_t}),$$

where the shorthand $\boldsymbol{\beta}_{s_t}$ denotes $\boldsymbol{\beta}_m$ if $s_t = m$, and similarly for $\sigma^2_{s_t}$.

Conditional on the full sequence of labels $s_1, \dots, s_T$, the sample partitions naturally into M subsets, corresponding to observations with $s_t = 1, \dots, M$. The component-specific parameters $\boldsymbol{\beta}_m$ and σ^2_m for $m = 1, \dots, M$ can then be estimated using only the observations assigned to their respective components. Conditional on the latent scales, the model is Gaussian, and all full conditional distributions are standard.

To verify that the latent-variable representation is equivalent to the original mixture model, we integrate out the component label s_t. By construction, $\mathbb{P}(s_t = m \,|\, \mathbf{w}) = w_m$, so the marginal density of y_t can be written as

$$p(y_t \,|\, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) = \sum_{m=1}^{M} w_m \, p(y_t \,|\, s_t = m, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) = \sum_{m=1}^{M} w_m \, f_{\mathcal{N}}(y_t;\, \mathbf{x}_t' \boldsymbol{\beta}_m, \sigma^2_m).$$

Thus, integrating out the latent indicator s_t recovers the observed-data likelihood in (4.7).

Finite mixture models suffer from a well-known identifiability issue, com-

monly referred to as the *label switching problem*. For an M-component mixture, the likelihood is invariant to permutations of the component labels. When exchangeable priors (i.e., priors invariant to permutations of the component labels) are placed on component-specific parameters, the posterior inherits this symmetry and contains $M!$ equivalent modes corresponding to relabelings of the same mixture.

Label switching can complicate posterior sampling and interpretation. MCMC algorithms may have difficulty moving between symmetric modes, leading to poor mixing, or may traverse multiple modes, in which case component-specific posterior summaries become uninformative. When interest is restricted to label-invariant quantities—such as the overall mixture density or predictive distributions—label switching poses no substantive difficulty, and explicit relabeling may be unnecessary (e.g., Celeux et al., 2000; Frühwirth-Schnatter, 2001).

Difficulties arise when component-level interpretation is required, such as identifying high-variance or low-mean regimes. In these cases, some convention for distinguishing components is needed. Hard identifiability constraints or asymmetric priors can impose a fixed labeling, though such choices may be arbitrary. A more principled approach is to sample under an exchangeable prior and relabel the MCMC output *post hoc*. Simple rules, such as ordering components by their means or weights, are effective when components are well separated, while more general methods align each draw with a reference configuration by minimizing a discrepancy measure over permutations (Stephens, 2000). After relabeling, component labels should be interpreted as post-processing conventions rather than inherent features of the model.

We now consider the baseline case with an exchangeable prior across mixture components. For each component $m = 1, \ldots, M$, we adopt a normal-inverse-gamma prior on the regression coefficients and error variance, $(\boldsymbol{\beta}_m \,|\, \sigma_m^2) \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma_m^2 \mathbf{V}_{\boldsymbol{\beta}})$ and $\sigma_m^2 \sim \mathcal{IG}(\nu_0, S_0)$. For the mixture weights $\mathbf{w} = (w_1, \ldots, w_M)'$, which are positive and sum to one, we use a Dirichlet prior, $\mathbf{w} \sim \mathcal{D}(\mathbf{a})$, with concentration parameters $\mathbf{a} = (a_1, \ldots, a_M)'$ and $a_m > 0$. The Dirichlet distribution and its properties are summarized in Appendix A. Finally, the prior for the latent indicators $\mathbf{s} = (s_1, \ldots, s_T)'$ is specified as above: conditional on $\mathbf{w}$, the component labels are independent with $\mathbb{P}(s_t = m \,|\, \mathbf{w}) = w_m$.

Posterior inference for the M-component normal mixture model is performed using a three-block Gibbs sampler that cycles through draws of the component-specific parameters $(\boldsymbol{\beta}_m, \sigma_m^2)$, the latent component indicators $\mathbf{s}$, and the mixture weights $\mathbf{w}$.

To implement Step 1, note that conditional on the component indicators $\mathbf{s}$, only observations with $s_t = m$ are informative about the component-specific parameters $\boldsymbol{\beta}_m$ and σ_m^2 for $m = 1, \ldots, M$. The sample therefore partitions

naturally into M disjoint subsets indexed by the labels s_t. Conditional on $\mathbf{s}$, the M parameter pairs $(\boldsymbol{\beta}_m, \sigma_m^2)$ can be sampled independently across components.

Let $\mathbf{y}_m$ denote the vector of observations y_t with $s_t = m$, and let $\mathbf{X}_m$ denote the corresponding matrix of regressors. Applying the conjugacy results for the normal-inverse-gamma prior established in Theorem 2.1 in Chapter 2, we obtain

$$(\boldsymbol{\beta}_m, \sigma_m^2 \,|\, \mathbf{y}, \mathbf{s}, \mathbf{w}) \sim \mathcal{NIG}(\widehat{\boldsymbol{\beta}}_m, \mathbf{D}_{\boldsymbol{\beta}_m}, \nu_0 + T_m/2, \widehat{S}_m),$$

where $T_m = \sum_{t=1}^{T} \mathbf{1}\{s_t = m\}$ is the number of observations allocated to component m,

$$\mathbf{D}_{\boldsymbol{\beta}_m} = (\mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}_m'\mathbf{X}_m)^{-1}, \quad \widehat{\boldsymbol{\beta}}_m = \mathbf{D}_{\boldsymbol{\beta}_m}(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \mathbf{X}_m'\mathbf{y}_m),$$
$$\widehat{S}_m = S_0 + \frac{1}{2}(\mathbf{y}_m'\mathbf{y}_m + \boldsymbol{\beta}_0'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_m'\mathbf{D}_{\boldsymbol{\beta}_m}^{-1}\widehat{\boldsymbol{\beta}}_m).$$

If a component is empty at a given iteration ($T_m = 0$), the conditional posterior above reduces to the prior: no observations are currently assigned to component m, so $(\boldsymbol{\beta}_m, \sigma_m^2)$ is drawn from its normal-inverse-gamma prior, and the component may be repopulated in later iterations. Frequently empty components in the MCMC output are a practical signal that M may be set too large.

Next, to implement Step 2, we sample $\mathbf{s} = (s_1, \ldots, s_T)'$; first note that the component labels are conditionally independent given the data and parameters. In particular, the joint conditional density factorizes as:

$$p(\mathbf{s} \,|\, \mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) \propto \prod_{t=1}^{T} p(y_t \,|\, s_t, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) p(s_t \,|\, \mathbf{w}),$$

so each s_t can be sampled separately. Because s_t takes value in $\{1, 2, \ldots, M\}$, we only need to compute the conditional probabilities $\mathbb{P}(s_t = m \,|\, y_t, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w})$. By Bayes' theorem,

$$\mathbb{P}(s_t = m \,|\, y_t, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) = c_t f_{\mathcal{N}}(y_t; \mathbf{x}_t'\boldsymbol{\beta}_m, \sigma_m^2) w_m,$$

where c_t is the normalizing constant. Using the fact that the probabilities must sum to one, we obtain $c_t = 1/(\sum_{j=1}^{M} f_{\mathcal{N}}(y_t; \mathbf{x}_t'\boldsymbol{\beta}_j, \sigma_j^2) w_j)$. Thus,

$$\mathbb{P}(s_t = m \,|\, y_t, \boldsymbol{\beta}, \boldsymbol{\sigma}^2, \mathbf{w}) = \frac{f_{\mathcal{N}}(y_t; \mathbf{x}_t'\boldsymbol{\beta}_m, \sigma_m^2) w_m}{\sum_{j=1}^{M} f_{\mathcal{N}}(y_t; \mathbf{x}_t'\boldsymbol{\beta}_j, \sigma_j^2) w_j}.$$

Each s_t can then be drawn from this discrete distribution using the inverse-transform method; see Exercise 3.3.

Finally, to implement Step 3, we derive the full conditional distribution of

$\mathbf{w}$. Note that $\mathbf{w}$ appears only in its prior and the prior probabilities $\mathbb{P}(s_t = m \,|\, \mathbf{w}) = w_m$ for $m = 1, \ldots, M$. The contribution of a single label s_t to the conditional distribution of $\mathbf{w}$ can be written succinctly as

$$p(s_t \,|\, \mathbf{w}) = \prod_{m=1}^{M} w_m^{\mathbf{1}\{s_t = m\}}.$$

Therefore,

$$\begin{aligned} p(\mathbf{w} \,|\, \mathbf{y}, \mathbf{s}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) &\propto p(\mathbf{w}) \times \prod_{t=1}^{T} p(s_t \,|\, \mathbf{w}) \propto \prod_{m=1}^{M} w_m^{a_m - 1} \times \prod_{t=1}^{T} \prod_{m=1}^{M} w_m^{\mathbf{1}\{s_t = m\}} \\ &= \prod_{m=1}^{M} w_m^{a_m + T_m - 1}, \end{aligned}$$

where $T_m = \sum_{t=1}^{T} \mathbf{1}\{s_t = m\}$ is the number of observations allocated to component m. Thus, updating $\mathbf{w}$ amounts to counting the occurrences of each component. The resulting full conditional is $(\mathbf{w} \,|\, \mathbf{y}, \mathbf{s}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2) \sim \mathcal{D}(a_1 + T_1, \ldots, a_M + T_M)$.

To complete Step 3, we require draws from the full conditional distribution of the mixture weights $\mathbf{w}$, which follows a Dirichlet distribution. Although MATLAB does not provide a built-in routine for sampling from a Dirichlet distribution, simulation is straightforward using a standard gamma-normalization construction. Because this device is useful in many finite mixture and Bayesian nonparametric models, we present it formally in Exercise 4.8. A practical implementation is provided in the MATLAB function `dirirnd.m`, which can be downloaded from the companion website.

4.2.2 Illustration: Approximating the $\log \chi_1^2$ Distribution

Finite normal mixtures can also be used as flexible approximations to unknown or non-Gaussian distributions. As an illustration, suppose we observe $T = 2,000$ independent draws from the $\log \chi_1^2$ distribution. This setting can be viewed as a special case of the linear regression mixture model introduced earlier, with the regressor vector $\mathbf{x}_t$ consisting only of an intercept. The resulting model is an M-component normal mixture for the marginal distribution of y_t,

$$p(y_t \,|\, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{w}) = \sum_{m=1}^{M} w_m \, f_{\mathcal{N}}(y_t; \mu_m, \sigma_m^2),$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)'$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_M^2)'$, and $\mathbf{w} = (w_1, \ldots, w_M)'$ with $w_m > 0$ and $\sum_{m=1}^{M} w_m = 1$.

Estimation proceeds using the data-augmentation framework developed

in Section 4.2.1. Introducing component indicators $s_t \in \{1, \ldots, M\}$ makes the model conditionally Gaussian, and all full conditional distributions follow directly from the results derived earlier. Because our interest here is the overall mixture density—which is invariant to permutations of the component labels—we adopt an exchangeable prior across components, $\sigma_m^2 \sim \mathcal{IG}(\nu_0, S_0)$, $(\mu_m \,|\, \sigma_m^2) \sim \mathcal{N}(\mu_0, V_\mu \sigma_m^2)$, and $\mathbf{w} \sim \mathcal{D}(\mathbf{a})$, and set $\nu_0 = 2$, $S_0 = 1$, $\mu_0 = 0$, $V_\mu = 100$, and $\mathbf{a} = 2 \cdot \mathbf{1}_M$.

We fit a four-component mixture ($M = 4$) using the Gibbs sampler described in Section 4.2.1. The MATLAB program `finmix_logchi2.m` implements this procedure, using the auxiliary routine `dirirnd.m` to sample the mixture weights from their Dirichlet full conditional. To accelerate convergence, the chain is initialized using K-means clustering of the data. Rather than storing draws of the component-specific parameters, we record posterior draws of the implied mixture density evaluated on a grid over $[-10, 5]$.

```
nsim = 20000; burnin = 1000;
M = 4; % # of components

% generate data
rng(42); T = 2000; df = 1;
y = log(chi2rnd(df,T,1));

% true log-chi^2_1 density
ftrue = @(x) 1/sqrt(2*pi) * exp(0.5*x - 0.5*exp(x));

% prior hyperparameters
nu0 = 2; S0 = 1;
mu0 = 0; Vmu = 100;
a0 = 2*ones(M,1);

ngrid = 500; xgrid = linspace(-10,5,ngrid)'; % grid
store_mixden = zeros(nsim,ngrid);
like = zeros(T,M);

[s, mu] = kmeans(y, M); % initialize using k-means
sig2 = zeros(M,1); alp  = zeros(M,1);
for m = 1:M
    idx = (s == m);
    sig2(m) = var(y(idx));
    alp(m)  = sum(idx)/T;
end

for isim = 1:(nsim + burnin)
    % sample (mu_m, sig2_m) for each component
    for m = 1:M
        idx = (s == m); Tm  = sum(idx); ym  = y(idx);
        Kmum   = 1/Vmu + Tm;
        mum_hat = (mu0/Vmu + sum(ym))/Kmum;
        Sm_hat  = S0 + 0.5*(ym'*ym + mu0^2/Vmu - mum_hat^2*Kmum);

        sig2(m) = 1/gamrnd(nu0 + Tm/2, 1/Sm_hat);
```

```
        mu(m)   = mum_hat + sqrt(sig2(m)/Kmum)*randn;
    end

    % sample s
    for m = 1:M
        like(:,m) = normpdf(y, mu(m), sqrt(sig2(m)));
    end
    joint_den = like .* repmat(alp', T, 1);
    prob = joint_den ./ repmat(sum(joint_den,2), 1, M);
    u = rand(T,1);
    cumprob = cumsum(prob, 2);
    s = 1 + sum(u > cumprob, 2);

    % sample alp
    ns = zeros(M,1);
    for m = 1:M
        ns(m) = sum(s == m);
    end
    alp = dirirnd(ns + a0, 1)';

    % store mixture density
    if isim > burnin
        isave = isim - burnin;
        mixden = normpdf(xgrid, mu', sqrt(sig2')) * alp;
        store_mixden(isave,:) = mixden';
    end
end
mixden_mean = mean(store_mixden,1)';
mixden_low  = prctile(store_mixden, 2.5, 1)';
mixden_high = prctile(store_mixden, 97.5, 1)';
```

Listing 4.2: Gibbs sampler for fitting a finite normal mixture to simulated data from the $\log \chi^2_1$ distribution (`finmix_logchi2.m`).

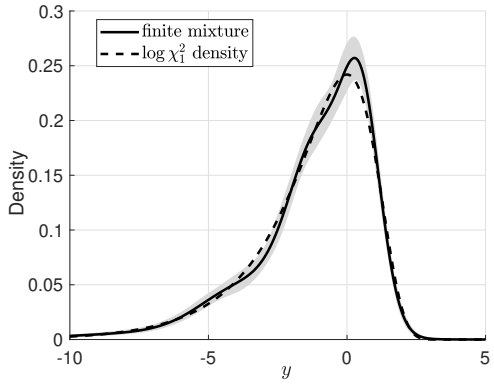

FIGURE 4.3: Estimated density from a four-component normal mixture: posterior mean and pointwise 95% credible bands, together with the true $\log \chi^2_1$ density.

Figure 4.3 compares the estimated mixture density with the true $\log \chi^2_1$ density. The normal mixture closely tracks the target distribution, and the pointwise 95% credible bands generally cover the true density. This example illustrates how finite mixtures of normals can provide accurate approximations to complex non-Gaussian distributions using a small number of components.

4.2.3 Application: Modeling PCE Inflation via an AR Model with an Outlier Component

Inflation dynamics vary across economic environments. Sustained expansions with stable monetary policy tend to feature low, persistent inflation, whereas recessions, supply disruptions, and periods of heightened uncertainty often bring large, volatile price movements. A single-component Gaussian autoregression may therefore be too restrictive for inflation, especially in the presence of rare but large shocks.

To accommodate such behavior, we consider an autoregressive model with a discrete *outlier component*, which can be viewed as a parsimonious special case of the finite mixture models introduced in Definition 4.4. Specifically, we model US PCE inflation using the following AR(2) specification with an outlier component:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + o_t\,\varepsilon_t, \quad (\varepsilon_t \mid \sigma^2) \sim \mathcal{N}(0, \sigma^2),$$

where $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})'$. The latent variable o_t is a discrete scale factor that captures occasional outlier realizations.

We assume that o_t takes values in a finite set of size $M = 3$, $o_t \in \{1, 5, 10\}$, with $\mathbb{P}(o_t = 1) = 1 - p_o$ corresponding to regular observations, and $\mathbb{P}(o_t = 5) = \mathbb{P}(o_t = 10) = p_o/2$ corresponding to moderate and large outliers, respectively. Conditional on o_t, the innovation variance is inflated by a factor o_t^2.

This specification fits naturally into the finite mixture framework introduced earlier. In particular, the latent scale o_t plays the role of a component indicator in a restricted finite mixture of normals with $M = 3$ components. Conditional on o_t, the observation y_t is Gaussian, while marginally the innovation distribution is a finite mixture of normals with a common conditional mean $\mathbf{x}_t'\boldsymbol{\beta}$ and component-specific variances σ^2, $25\sigma^2$, and $100\sigma^2$.

Relative to the general finite mixture autoregressive model, this outlier formulation imposes strong restrictions: the regression coefficients are shared across components, and the component variances are fixed multiples of a common scale. These restrictions yield a parsimonious and interpretable model in which mixture components correspond directly to rare large shocks rather than persistent regimes. The model can be viewed as a simplified version of the outlier specification proposed by Stock and Watson (2016).

We complete the model by assigning independent priors $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 100 \cdot \mathbf{I}_3)$, $\sigma^2 \sim \mathcal{IG}(3, 2)$, and a beta prior on the outlier probability,

$$p_o \sim \mathcal{B}\left(\frac{10}{4}, 40\left(1 - \frac{1}{16}\right)\right).$$

The beta prior is calibrated so that the prior mean implies one outlier, on average, every four years in quarterly data. Compared with the Dirichlet prior used in the general finite mixture model, this prior structure exploits the restricted form of the mixture weights implied by the $M = 3$ outlier specification and reduces the effective dimensionality of the parameter space.

We estimate the model on quarterly US PCE inflation from 1960Q1 to 2024Q4, extending the sample used in the earlier PCE examples through 2024 so that the COVID-19 period is included. The following MATLAB script `linreg_outlier.m` implements a four-block Gibbs sampler based on the data-augmentation representation of the outlier model. Conditional on the latent scale factors $\{o_t\}$, the model reduces to a Gaussian AR(2) regression with heteroskedastic errors, so the regression coefficients are drawn from a multivariate normal distribution and the innovation variance from an inverse-gamma distribution using standard conjugate formulas.

The outlier indicators $\{o_t\}$ are then updated individually from a discrete posterior over the finite support $\{1, 5, 10\}$, with posterior probabilities proportional to the Gaussian likelihood evaluated at variance $o_t^2\sigma^2$ and the prior weights implied by p_o. Finally, the outlier probability p_o is updated from its beta full conditional by counting the number of quarters classified as outliers.

```
nsim = 50000; burnin = 1000;

% load data
data = readmatrix('USPCE.csv', 'Range', 'B2:B261');
y0 = data(1:4); y  = data(5:end); T  = length(y);
xlag1 = [y0(4); y(1:end-1)]; xlag2 = [y0(3:4); y(1:end-2)];
X = [ones(T,1), xlag1, xlag2];
k = size(X,2);

% prior hyperparameters
beta0 = zeros(k,1); iVbeta = speye(k)/100;
nu0 = 3; S0 = 2;
p0a = 10/4; p0b = (1-1/16)*40;

% initialize the chain
beta = (X'*X)\(X'*y);
sig2 = sum((y-X*beta).^2)/T;
po = 1/16;
o = ones(T,1);

store_o = zeros(nsim,T);
store_theta = zeros(nsim,5);  % [beta', sig2, po]
```

```
o_grid = [1, 5, 10]'; % possible values for o_t
for isim = 1:nsim + burnin
        % sample beta
    iO = sparse(1:T,1:T,1./o.^2);
    Dbeta = (iVbeta + X'*iO*X/sig2)\speye(k);
    beta_hat = Dbeta*(iVbeta*beta0 + X'*iO*y/sig2);
    beta = beta_hat + chol(Dbeta,'lower')*randn(k,1);

        % sample sig2
    e = (y - X*beta)./o;
    sig2 = 1/gamrnd(nu0+T/2,1/(S0 + e'*e/2));

        % sample o
    o_lpri = log([1-po; po/2; po/2]); % log prior density
    u = (y - X*beta)/sqrt(sig2);
    for tt=1:T
        lliket = -log(o_grid) -.5*u(tt)^2./o_grid.^2;
        o_post = exp(lliket + o_lpri - max(lliket));
        o_post = o_post/sum(o_post);
        idx = find(rand<cumsum(o_post),1);
        o(tt) = o_grid(idx);
    end

        % sample po
    tmp = sum(o>1);
    po = betarnd(p0a + tmp, p0b + T-tmp);

        % store the parameters
    if isim > burnin
        isave = isim - burnin;
        store_theta(isave,:) = [beta', sig2, po];
        store_o(isave,:) = o';
    end
end
theta_mean = mean(store_theta);
theta_CI = quantile(store_theta,[.025 .975]);
o_mean = mean(store_o)';
```

Listing 4.3: Gibbs sampler for an AR(2) model with an outlier component (`linreg_outlier.m`).

The script stores posterior draws of the model parameters and the latent outlier indicators, which are used to construct posterior summaries and the outlier probabilities reported in Figure 4.4. For most of the sample, the posterior probability of an outlier, $\mathbb{P}(o_t > 1 \,|\, \mathbf{y})$, is close to zero, typically on the order of 0.5–1%. In contrast, the model assigns sharply elevated outlier probabilities to a small number of quarters that coincide with well-known historical episodes. For example, the outlier probability rises to 0.23 in 1973Q2 and to 0.80 in 1974Q1 during the first oil shock, and spikes to one in 2008Q4 at the height of the Great Recession. More recently, the model assigns elevated outlier probabilities of about 0.23–0.28 in 2020Q2–2020Q3, reflecting the unusually large inflation movements during the COVID-19 pandemic.

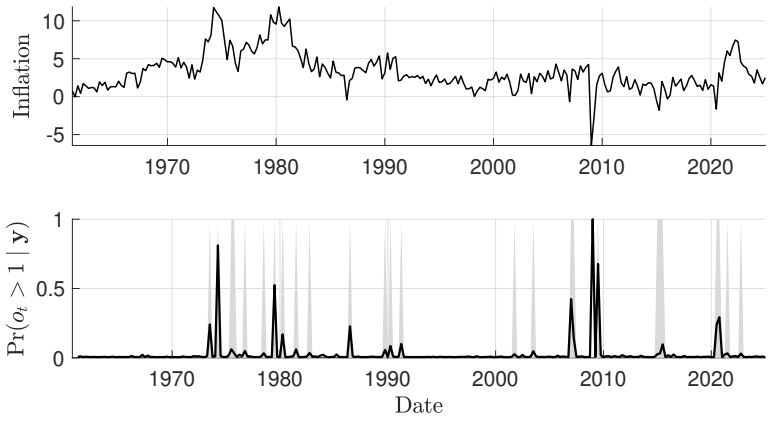

FIGURE 4.4: US PCE inflation and posterior outlier probabilities from the AR(2) outlier model. The top panel plots PCE inflation; the bottom panel plots the posterior probability of an outlier, $\mathbb{P}(o_t > 1 \mid \mathbf{y})$. Spikes correspond to quarters in which the model assigns high probability to variance-inflated inflation innovations.

Although Figure 4.4 reports only the posterior probability of an outlier, the posterior mean of the scale parameter o_t provides additional information about the *magnitude* of variance inflation. In tranquil periods, $\mathbb{E}(o_t \mid \mathbf{y})$ remains close to one by construction. During extreme episodes, however, the posterior mean rises sharply. For example, it reaches 5.75 in 1974Q1, corresponding to a variance inflation factor of roughly 33, and increases to 8.13 in 2008Q4, implying a variance inflation factor of about 66. These results indicate that the outlier model isolates rare but economically important quarters in which inflation innovations are far larger than can be accommodated by the baseline Gaussian AR(2) specification.

It is useful to contrast this behavior with the AR(2) model with Student-t innovations discussed in Section 3.2.4. Both models can be expressed as mixtures of normals. The Student-t specification corresponds to a *continuous* scale mixture that downweights extreme observations in a fully data-driven manner. The outlier model, by contrast, employs a *discrete* latent scale and isolates a small number of quarters as distinct high-variance episodes. Although both specifications generate heavy tails, the discreteness of the latent scale in the outlier model sharpens the distinction between "regular" and "outlier" periods and can facilitate economic interpretation.

4.3 Further Reading

Location-scale (or mean-variance) mixtures of normals have a long history in probability and statistics and are widely used in finance and macroeconomics. A foundational reference is Barndorff-Nielsen (1977), which introduces the generalized hyperbolic class through normal mean-variance mixing. A systematic treatment of normal mean-variance mixtures and related families is provided by Barndorff-Nielsen, Kent and Sørensen (1982). In empirical finance, generalized hyperbolic and related mixtures are used as flexible parametric models for asset returns and other macro-financial quantities; see, for example, Eberlein and Keller (1995). The normal inverse Gaussian distribution is an important special case, and Barndorff-Nielsen (1997) discusses its links to stochastic volatility modeling. Skewed families are likewise widely used; for a book-length treatment of skew-normal and skew-t families, including multivariate extensions and inferential aspects, see Azzalini and Capitanio (2014).

Beyond the basic asymmetric Laplace formulation, a growing literature develops Bayesian quantile regression methods that relax distributional assumptions and broaden the range of applications. In addition to Yu and Moyeed (2001) and the efficient conditionally Gaussian sampler in Kozumi and Kobayashi (2011), useful extensions include Bayesian regularized quantile regression (Li, Lin and Xi, 2010) and Bayesian approaches that enforce noncrossing quantiles in simultaneous linear quantile regression (Tokdar and Kadane, 2012). For financial and macroeconomic time series, Gerlach, Chen and Chan (2011) develop Bayesian time-varying quantile models for value-at-risk, incorporating dynamics in the conditional quantiles themselves. Likelihood alternatives that move beyond the asymmetric Laplace distribution are discussed by Yang and He (2012), who introduce Bayesian empirical likelihood methods for quantile regression. For background on classical quantile regression and broader methodological context, see Koenker (2005).

In macroeconomic applications, the growth-at-risk literature initiated by Adrian, Boyarchenko and Giannone (2019) has been extended in several directions, including a critical reassessment of out-of-sample predictability in Plagborg-Møller et al. (2020) and a Bayesian VAR-based measurement of macroeconomic tail risk in Carriero, Clark and Marcellino (2024).

Turning to the second broad class of models in this chapter, finite mixture models are treated comprehensively in McLachlan and Peel (2000), including non-Gaussian mixtures and estimation via the Expectation-Maximization algorithm. A Bayesian-focused monograph is Frühwirth-Schnatter (2006), which also covers marginal likelihood computation and Markov-switching extensions. Frühwirth-Schnatter and Kaufmann (2008) use finite mixtures to cluster mul-

tiple time series into groups that share common dynamics, selecting the number of groups by marginal likelihood.

Early Bayesian work on mixture posteriors and computation includes Diebolt and Robert (1994) and Celeux, Hurn and Robert (2000). The label switching problem and principled relabeling strategies are discussed in depth by Stephens (2000).

When the number of components is treated as unknown, it can be inferred jointly with the parameters by reversible jump MCMC, introduced by Green (1995) and applied to mixtures by Richardson and Green (1997). For a modern collection of perspectives and recent developments, see the edited volume by Frühwirth-Schnatter, Celeux and Robert (2019).

Bayesian outlier modeling via scale mixtures has a long history; the seminal contribution is West (1984), which formalizes the use of latent scale variables to downweight aberrant observations in a regression setting.

Exercises

Exercise 4.1 (Symmetry of Scale Mixtures of Normals)**.** Let $(X \mid \lambda) \sim \mathcal{N}(0, \sigma^2\lambda)$ and $\lambda \sim G$, where G is supported on $\mathbb{R}_+$.

(a) Show that the marginal density of X is symmetric: $f(x) = f(-x)$ for all $x \in \mathbb{R}$. Conclude that $\mathbb{P}(X \leq 0) = 1/2$ and hence that the median of X is zero.

(b) Show that all odd moments (when they exist) satisfy $\mathbb{E}(X^{2k+1}) = 0$.

(c) Give a brief explanation of why scale mixtures with zero conditional mean cannot generate skewness.

Exercise 4.2 (Laplace and Asymmetric Laplace as Normal Mixtures)**.** This exercise illustrates how the Laplace and asymmetric Laplace distributions arise as normal mixtures.

(a) *Laplace distribution.* Let $(X \mid \lambda) \sim \mathcal{N}(0, \sigma^2\lambda)$ and $\lambda \sim \mathcal{G}(1, 1/2)$. Derive the marginal density of X and show that it corresponds to the density of the Laplace distribution with scale parameter σ given in (4.1).

(b) *Asymmetric Laplace distribution.* Consider the hierarchical model $(X \mid \lambda) \sim \mathcal{N}\big(\vartheta_\tau \lambda, \varphi_\tau \sigma^2 \lambda\big)$ and $\lambda \sim \mathcal{G}(1, 1/\sigma^2)$, where ϑ_τ and φ_τ are defined in (4.3). Show that the marginal distribution of X is the asymmetric Laplace distribution $\mathcal{AL}(0, \sigma^2, \tau)$ with density given in (4.2).

Hint: Both parts rely on the identity

$$\int_0^\infty \lambda^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\left(a\lambda + \frac{b}{\lambda}\right)} \, \mathrm{d}\lambda = \sqrt{\frac{2\pi}{a}} \exp(-\sqrt{ab}), \quad a > 0, b \geq 0.$$

In part (b), you may find it helpful to consider the cases $x \geq 0$ and $x < 0$ separately.

Exercise 4.3 (τ-Quantile of the Asymmetric Laplace Distribution)**.** Let $X \sim \mathcal{AL}(0, \sigma^2, \tau)$, where the asymmetric Laplace density is given in (4.2).

(a) Show that the density can be written in the piecewise form

$$f(x \,|\, \sigma^2, \tau) = \begin{cases} \dfrac{\tau(1-\tau)}{\sigma^2} \mathrm{e}^{\frac{1-\tau}{\sigma^2} x}, & x < 0, \\ \dfrac{\tau(1-\tau)}{\sigma^2} \mathrm{e}^{-\frac{\tau}{\sigma^2} x}, & x \geq 0. \end{cases}$$

(b) Using the expression above, compute the cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$ for $x \leq 0$.

(c) Evaluate $F(0)$ and show that $\mathbb{P}(X \leq 0) = \tau$. Conclude that 0 is the τ-quantile of the asymmetric Laplace distribution.

Exercise 4.4 (Inverse Gaussian Representation for Latent Scales in Bayesian Quantile Regression)**.** This exercise derives the full conditional distribution used to sample the latent scale parameters in Bayesian quantile regression.

(a) *Inverse Gaussian distribution.* A random variable Z is said to follow an *inverse Gaussian* distribution with parameters (ψ, μ), written $Z \sim \mathcal{IGAUSS}(\psi, \mu)$, if its density is

$$f(z) \propto z^{-\frac{3}{2}} \mathrm{e}^{-\frac{\psi}{2}\left(\frac{(z-\mu)^2}{\mu^2 z}\right)}, \quad z > 0,$$

where $\psi > 0$ is a shape parameter and $\mu > 0$ is a mean parameter. Show that if $Z \sim \mathcal{IGAUSS}(\psi, \mu)$, then $W = Z^{-1}$ has density with kernel

$$f_W(w) \propto w^{-\frac{1}{2}} \mathrm{e}^{-\frac{\psi}{2}\left(w + \mu^{-2} w^{-1}\right)}.$$

(b) Using the kernel in part (a), show that the full conditional distribution of λ_t^{-1} in the Bayesian quantile regression model satisfies

$$(\lambda_t^{-1} \,|\, \mathbf{y}, \boldsymbol{\beta}_\tau, \sigma^2) \sim \mathcal{IGAUSS}\left(a_\tau, \ \mu_{t,\tau}\right),$$

where $\mu_{t,\tau} = \sqrt{a_\tau / b_{t,\tau}}$ with a_τ and $b_{t,\tau}$ being defined in (4.5).

Exercise 4.5 (Posterior Predictive Simulation for Bayesian Quantile Regression)**.** Adopt the Bayesian quantile regression model in Definition 4.3 and the mixture representation in (4.4). Let y^* denote a future outcome with regressors $\mathbf{x}^*$.

(a) Conditional on $(\boldsymbol{\beta}_\tau, \sigma^2)$, describe a simulation scheme to generate a draw from the predictive distribution of y^* using the latent scale $\lambda^* \sim \mathcal{G}(1, 1/\sigma^2)$:

$$y^* = \mathbf{x}^{*\prime}\boldsymbol{\beta}_\tau + \varepsilon^*, \quad (\varepsilon^* \mid \lambda^*, \sigma^2) \sim \mathcal{N}(\vartheta_\tau \lambda^*, \varphi_\tau \sigma^2 \lambda^*).$$

(b) Explain why the predictive τ-quantile satisfies $Q_\tau(y^* \mid \boldsymbol{\beta}_\tau, \sigma^2) = \mathbf{x}^{*\prime}\boldsymbol{\beta}_\tau$.

(c) Using posterior draws $\{\boldsymbol{\beta}_\tau^{(r)}, \sigma^{2(r)}\}_{r=1}^R$, propose a Monte Carlo estimate of $\mathbb{P}(y^* \leqslant c \mid \mathbf{y})$ for a fixed threshold c.

Exercise 4.6 (Skew-Normal Regression via a Latent Location Mixture)**.** Consider a regression model whose errors follow the skew-normal distribution of Example 4.2. In location-mixture form,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \psi\lambda_t + \varepsilon_t,$$

where $\lambda_t \sim \mathcal{TN}_{(0,\infty)}(0, 1)$ and $(\varepsilon_t \mid \omega^2) \sim \mathcal{N}(0, \omega^2)$. This is a reparameterization of the standardized skew-normal $(X \mid \lambda) \sim \mathcal{N}(\delta\lambda, 1 - \delta^2)$ of that example: the parameters $\psi = \delta\sigma$ and $\omega^2 = (1 - \delta^2)\sigma^2$ introduce a scale σ, so the error has total scale $\sigma^2 = \psi^2 + \omega^2$ and slant ψ/ω, with $\delta \in (-1, 1)$; here ψ governs the skewness and $\omega^2 > 0$ the variance. We observe $\{(\mathbf{x}_t, y_t)\}_{t=1}^T$ but not $\{\lambda_t\}$. Because the latent shift λ_t has a positive mean, the linear predictor $\mathbf{x}_t'\boldsymbol{\beta}$ is a location parameter rather than the conditional mean of y_t.

(a) Derive the marginal density $p(y_t \mid \boldsymbol{\beta}, \psi, \omega^2)$ by integrating out λ_t, and explain how ψ governs its skewness—in particular, why $\psi = 0$ gives a symmetric (Gaussian) error.

(b) Assume independent priors $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{V}_{\boldsymbol{\beta}})$, $\psi \sim \mathcal{N}(\psi_0, V_\psi)$, and $\omega^2 \sim \mathcal{IG}(\nu_0, S_0)$. Derive the Gibbs sampler for $(\boldsymbol{\beta}, \psi, \omega^2, \{\lambda_t\}_{t=1}^T)$. Verify that, because ψ enters the observation equation linearly, all four full conditionals are of standard form, so that no Metropolis–Hastings step is required.

Exercise 4.7 (Skew-Normal AR(2) Regression for US PCE)**.** Consider the skew-normal regression model introduced in Exercise 4.6, with $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})'$. Implement the four-block Gibbs sampler derived in that exercise to fit the US PCE dataset `USPCE.csv`. Adopt the following prior specification: $\boldsymbol{\beta}_0 = \mathbf{0}_3$, $\mathbf{V}_{\boldsymbol{\beta}} = 100 \cdot \mathbf{I}_3$, $\psi_0 = 0$, $V_\psi = 1$, $\nu_0 = 3$, and $S_0 = 2$.

Report the posterior mean $\mathbb{E}(\psi \mid \mathbf{y})$ and that of the implied skewness parameter $\delta = \psi/\sqrt{\psi^2 + \omega^2}$. Assess whether there is evidence of skewness in

the regression errors—for example, by checking whether the 95% credible interval for ψ excludes zero, or by computing a Savage–Dickey Bayes factor for the hypothesis $\psi = 0$—and comment on the direction and strength of the skewness.

Exercise 4.8 (Sampling from a Dirichlet Distribution)**.** Let $\mathbf{w} = (w_1, \dots, w_M)'$ follow a Dirichlet distribution with parameter vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)'$, where $\alpha_m > 0$ for all m.

(a) Show that if $g_m \sim \mathcal{G}(\alpha_m, 1)$ independently for $m = 1, \dots, M$, and $w_m = g_m / \sum_{j=1}^{M} g_j$, then $\mathbf{w}$ lies in the unit simplex; that is, $w_m > 0$ for all m and $\sum_{m=1}^{M} w_m = 1$.

(b) Prove that the resulting vector $\mathbf{w}$ has the Dirichlet distribution $\mathcal{D}(\boldsymbol{\alpha})$.

Exercise 4.9 (Finite-Mixture Gibbs Sampler)**.** This exercise asks you to implement the Gibbs sampler for a finite normal mixture and to interpret the fitted components, which need not be balanced. Consider the two-component regression mixture

$$y_t = \mathbf{x}_t' \boldsymbol{\beta}_{s_t} + \varepsilon_t, \quad (\varepsilon_t \,|\, s_t = m, \sigma_m^2) \sim \mathcal{N}(0, \sigma_m^2),$$

where $\mathbb{P}(s_t = m \,|\, \mathbf{w}) = w_m$, with exchangeable priors

$$(\boldsymbol{\beta}_m \,|\, \sigma_m^2) \sim \mathcal{N}(\boldsymbol{\beta}_0, \sigma_m^2 \mathbf{V}_{\boldsymbol{\beta}}), \quad \sigma_m^2 \sim \mathcal{IG}(\nu_0, S_0), \quad \mathbf{w} \sim \mathcal{D}(\mathbf{a}),$$

for $m = 1, 2$ and $\mathbf{a} = 2 \cdot \mathbf{1}_2$.

(a) Implement the three-block Gibbs sampler in Section 4.2.1 using an AR(2) design $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})'$, and fit the US PCE inflation series `USPCE.csv`. Use $\boldsymbol{\beta}_0 = \mathbf{0}_3$, $\mathbf{V}_{\boldsymbol{\beta}} = 100 \cdot \mathbf{I}_3$, $\nu_0 = 3$, and $S_0 = 2$. To resolve the labeling ambiguity, relabel the two components after each iteration so that $w_1 > w_2$.

(b) Report the posterior means of the weights $\mathbf{w}$ and the component parameters $(\boldsymbol{\beta}_m, \sigma_m^2)$. Compute the posterior classification probability

$$\widehat{\mathbb{P}}(s_t = 1 \,|\, \mathbf{y}) = \frac{1}{R} \sum_{r=1}^{R} \mathbf{1}\left\{ s_t^{(r)} = 1 \right\},$$

plot it over time together with inflation, and identify the observations assigned to the second component.

(c) Interpret the fitted mixture. Comment on the relative sizes of the two components and on what the smaller one captures. Explain why a mixture in which one component dominates acts as a flexible, fat-tailed alternative to the single-component Gaussian model, and relate this outlier-robust behavior to the Student-t regression of Chapter 3.

Part II

Bayesian Computation and High-Dimensional Methods

Part III

Bayesian Time-Series Models

Bibliography

Abadir, K. M. and Magnus, J. R. (2005), *Matrix Algebra*, Vol. 1 of *Econometric Exercises*, Cambridge University Press, Cambridge.

Adrian, T., Boyarchenko, N. and Giannone, D. (2019), 'Vulnerable growth', *American Economic Review* **109**(4), 1263–1289.

Aguilar, O. and West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **18**(3), 338–357.

Albert, J. and Chib, S. (1993), 'Bayesian analysis of binary and polychotomous response data', *Journal of the American Statistical Association* **88**, 669–679.

Álvarez, M. A., Rosasco, L. and Lawrence, N. D. (2012), 'Kernels for vector-valued functions: A review', *Foundations and Trends in Machine Learning* **4**(3), 195–266.

Amir-Ahmadi, P., Matthes, C. and Wang, M.-C. (2020), 'Choosing prior hyperparameters: With applications to time-varying parameter models', *Journal of Business and Economic Statistics* **38**(1), 124–136.

Amisano, G. and Geweke, J. (2017), 'Prediction using several macroeconomic models', *Review of Economics and Statistics* **99**(5), 912–925.

Anderson, T. W. and Rubin, H. (1956), Statistical inference in factor analysis, *in* 'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability', Vol. 5, pp. 111–150.

Andrieu, C., Doucet, A. and Holenstein, R. (2010), 'Particle markov chain monte carlo methods', *Journal of the Royal Statistical Society: Series B* **72**(3), 269–342.

Antolín-Díaz, J., Drechsel, T. and Petrella, I. (2024), 'Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails', *Journal of Econometrics* **238**(2), 105634.

Antolín-Díaz, J., Petrella, I. and Rubio-Ramírez, J. F. (2021), 'Structural scenario analysis with SVARs', *Journal of Monetary Economics* **117**, 798–815.

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018), 'Narrative sign restrictions for SVARs', *American Economic Review* **108**(10), 2802–2829.

Antoniak, C. E. (1974), 'Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems', *The Annals of Statistics* **2**(6), 1152–1174.

Ardia, D., Baştürk, N., Hoogerheide, L. and van Dijk, H. K. (2012), 'A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood', *Computational Statistics and Data Analysis* **56**(11), 3398–3414.

Arias, J. E., Caldara, D. and Rubio-Ramírez, J. F. (2019), 'The systematic component of monetary policy in SVARs: An agnostic identification procedure', *Journal of Monetary Economics* **101**, 1–13.

Arias, J. E., Rubio-Ramírez, J. F. and Shin, M. (2023), 'Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models', *Journal of Econometrics* **235**(2), 1054–1086.

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018), 'Inference based on

structural vector autoregressions identified with sign and zero restrictions: Theory and applications', *Econometrica* **86**(2), 685–720.

Aruoba, S. B., Diebold, F. X. and Scotti, C. (2009), 'Real-time measurement of business conditions', *Journal of Business and Economic Statistics* **27**(4), 417–427.

Asai, M. and McAleer, M. (2009), 'The structure of dynamic correlations in multivariate stochastic volatility models', *Journal of Econometrics* **150**(2), 182–192.

Asai, M., McAleer, M. and Yu, J. (2006), 'Multivariate stochastic volatility: A review', *Econometric Reviews* **25**(2-3), 145–175.

Azzalini, A. (1985), 'A class of distributions which includes the normal ones', *Scandinavian Journal of Statistics* **12**(2), 171–178.

Azzalini, A. and Capitanio, A. (2003), 'Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution', *Journal of the Royal Statistical Society: Series B* **65**(2), 367–389.

Azzalini, A. and Capitanio, A. (2014), *The Skew-Normal and Related Families*, Institute of Mathematical Statistics Monographs, Cambridge University Press, Cambridge.

Bańbura, M., Giannone, D. and Lenza, M. (2015), 'Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections', *International Journal of Forecasting* **31**(3), 739–756.

Bańbura, M., Giannone, D., Modugno, M. and Reichlin, L. (2013), Now-casting and the real-time data flow, *in* 'Handbook of Economic Forecasting', Vol. 2, Elsevier, pp. 195–237.

Bańbura, M., Giannone, D. and Reichlin, L. (2010), 'Large Bayesian vector auto regressions', *Journal of Applied Econometrics* **25**(1), 71–92.

Bańbura, M. and Modugno, M. (2014), 'Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data', *Journal of Applied Econometrics* **29**(1), 133–160.

Bańbura, M. and van Vlodrop, A. (2018), 'Forecasting with Bayesian vector autoregressions with time variation in the mean', *Tinbergen Institute Discussion Paper 2018-025/IV* .

Barndorff-Nielsen, O. E. (1977), 'Exponentially decreasing distributions for the logarithm of particle size', *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **353**(1674), 401–419.

Barndorff-Nielsen, O. E. (1997), 'Normal inverse Gaussian distributions and stochastic volatility modelling', *Scandinavian Journal of Statistics* **24**(1), 1–13.

Barndorff-Nielsen, O. E., Kent, J. and Sørensen, M. (1982), 'Normal variance-mean mixtures and z distributions', *International Statistical Review* **50**(2), 145–159.

Bassetti, F., Casarin, R. and Ravazzolo, F. (2018), 'Bayesian nonparametric calibration and combination of predictive distributions', *Journal of the American Statistical Association* **113**(522), 675–685.

Baumeister, C. and Hamilton, J. D. (2015), 'Sign restrictions, structural vector autoregressions, and useful prior information', *Econometrica* **83**(5), 1963–1999.

Baumeister, C., Korobilis, D. and Lee, T. K. (2022), 'Energy markets and global economic conditions', *Review of Economics and Statistics* **104**(4), 828–844.

Bauwens, L., Lubrano, M. and Richard, J. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, New York.

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

Benati, L. (2008), 'The "great moderation" in the United Kingdom', *Journal of Money, Credit and Banking* **40**(1), 121–147.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, second edn, Springer, New York.

Berger, T., Everaert, G. and Vierke, H. (2016), 'Testing for time variation in an unobserved components model for the U.S. economy', *Journal of Economic Dynamics and Control* **69**, 179–208.

Bernanke, B., Boivin, J. and Eliasz, P. S. (2005), 'Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach', *The Quarterly Journal of Economics* **120**(1), 387–422.

Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley & Sons, Chichester.

Bertsche, D. and Braun, R. (2022), 'Identification of structural vector autoregressions by stochastic volatility', *Journal of Business and Economic Statistics* **40**(1), 328–341.

Berument, H., Yalcin, Y. and Yildirim, J. (2009), 'The effect of inflation uncertainty on inflation: Stochastic volatility in mean model within a dynamic framework', *Economic Modelling* **26**(6), 1201–1207.

Betancourt, M. (2017), A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

Bhadra, A., Datta, J., Polson, N. G. and Willard, B. (2017), 'The horseshoe+ estimator of ultra-sparse signals', *Bayesian Analysis* **12**(4), 1105–1131.

Bhattacharya, A. and Dunson, D. (2011), 'Sparse Bayesian infinite factor models', *Biometrika* **98**(2), 291–306.

Bhattacharya, A., Pati, D., Pillai, N. S. and Dunson, D. B. (2015), 'Dirichlet-Laplace priors for optimal shrinkage', *Journal of the American Statistical Association* **110**(512), 1479–1490.

Bianchi, F. (2013), 'Regime switches, agents' beliefs, and post-World War II U.S. macroeconomic dynamics', *Review of Economic Studies* **80**(2), 463–490.

Blackwell, D. and MacQueen, J. B. (1973), 'Ferguson distributions via pólya urn schemes', *The Annals of Statistics* **1**(2), 353–355.

Blanchard, O. J. and Quah, D. (1989), 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review* **79**(4), 655–673.

Bognanni, M. (2018), 'A class of time-varying parameter structural VARs for inference under exact or set identification', *FRB of Cleveland Working Paper* .

Bognanni, M. (2022), 'Comment on "large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"', *Journal of Econometrics* **227**(2), 498–505.

Boivin, J., Giannoni, M. P. and Mihov, I. (2009), 'Sticky prices and monetary policy: Evidence from disaggregated US data', *American Economic Review* **99**(1), 350–384.

Boss, J., Datta, J., Wang, X., Park, S. K., Kang, J. and Mukherjee, B. (2024), 'Group inverse-gamma gamma shrinkage for sparse linear models with block-correlated regressors', *Bayesian Analysis* **19**(3), 785–814.

Botev, Z. I. (2017), 'The normal law under linear restrictions: simulation and estimation via minimax tilting', *Journal of the Royal Statistical Society: Series B* **79**(1), 125–148.

Braun, R. (2023), 'The importance of supply and demand for oil prices: evidence from non-gaussianity', *Quantitative Economics* **14**(4), 1163–1198.

Brooks, S., Gelman, A., Jones, G. and Meng, X.-L. (2011), *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC.

Caldara, D. and Herbst, E. P. (2019), 'Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs', *American Economic Journal: Macroeconomics* **11**(1), 157–192.

Canova, F. (1993), 'Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model', *Journal of Economic Dynamics and Control* **17**, 233–261.

Canova, F. and Ciccarelli, M. (2004), 'Forecasting and turning point predictions in a Bayesian panel VAR model', *Journal of Econometrics* **120**(2), 327–359.

Canova, F. and Ciccarelli, M. (2009), 'Estimating multicountry VAR models', *International Economic Review* **50**(3), 929–959.

Canova, F. and Ciccarelli, M. (2013), Panel vector autoregressive models: A survey, *in* T. B. Fomby, L. Kilian and A. Murphy, eds, 'VAR Models in Macroeconomics—New Developments and Applications: Essays in Honor of Christopher A. Sims', Vol. 32 of *Advances in Econometrics*, Emerald Group Publishing, pp. 205–246.

Canova, F. and De Nicoló, G. (2002), 'Monetary disturbances matter for business fluctuations in the G-7', *Journal of Monetary Economics* **49**(6), 1131–1159.

Carriero, A., Chan, J. C. C., Clark, T. E. and Marcellino, M. (2022), 'Corrigendum to "large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors"', *Journal of Econometrics* **227**(2), 506–512.

Carriero, A., Clark, T. E. and Marcellino, M. (2015*a*), 'Bayesian VARs: Specification choices and forecast accuracy', *Journal of Applied Econometrics* **30**(1), 46–73.

Carriero, A., Clark, T. E. and Marcellino, M. (2015*b*), 'Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility', *Journal of the Royal Statistical Society Series A: Statistics in Society* **178**(4), 837–862.

Carriero, A., Clark, T. E. and Marcellino, M. (2016), 'Common drifting volatility in large Bayesian VARs', *Journal of Business and Economic Statistics* **34**(3), 375–390.

Carriero, A., Clark, T. E. and Marcellino, M. (2018), 'Measuring uncertainty and its impact on the economy', *Review of Economics and Statistics* **100**(5), 799–815.

Carriero, A., Clark, T. E. and Marcellino, M. (2019), 'Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors', *Journal of Econometrics* **212**(1), 137–154.

Carriero, A., Clark, T. E. and Marcellino, M. (2024), 'Capturing macro-economic tail risks with Bayesian vector autoregressions', *Journal of Money, Credit and Banking* **56**(5), 1099–1127.

Carriero, A., Clark, T. E., Marcellino, M. and Mertens, E. (2024), 'Addressing COVID-19 outliers in BVARs with stochastic volatility', *Review of Economics and Statistics* **106**(5), 1403–1421.

Carriero, A., Kapetanios, G. and Marcellino, M. (2009), 'Forecasting exchange rates with a large Bayesian VAR', *International Journal of Forecasting* **25**(2), 400–417.

Carriero, A., Kapetanios, G. and Marcellino, M. (2011), 'Forecasting large datasets with Bayesian reduced rank multivariate models', *Journal of Applied Econometrics* **26**(5), 735–761.

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

Carvalho, C. M., Polson, N. G. and Scott, J. G. (2010), 'The horseshoe estimator for sparse signals', *Biometrika* **97**(2), 465–480.

Castelnuovo, E. (2023), 'Uncertainty before and during COVID-19: A survey', *Journal of Economic Surveys* **37**(3), 821–864.

Celeux, G., Forbes, F., Robert, C. P. and Titterington, D. M. (2006), 'Deviance information criteria for missing data models', *Bayesian Analysis* **1**(4), 651–673.

Celeux, G., Hurn, M. and Robert, C. P. (2000), 'Computational and inferential difficulties with mixture posterior distributions', *Journal of the American Statistical Association* **95**(451), 957–970.

Chan, J. C. C. (2013), 'Moving average stochastic volatility models with application to inflation forecast', *Journal of Econometrics* **176**(2), 162–172.

Chan, J. C. C. (2017), 'The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling', *Journal of Business and Economic Statistics* **35**(1), 17–28.

Chan, J. C. C. (2018), 'Specification tests for time-varying parameter models with stochastic volatility', *Econometric Reviews* **37**(8), 807–823.

Chan, J. C. C. (2020), 'Large Bayesian VARs: A flexible Kronecker error covariance structure', *Journal of Business and Economic Statistics* **38**(1), 68–79.

Chan, J. C. C. (2021), 'Minnesota-type adaptive hierarchical priors for large Bayesian VARs', *International Journal of Forecasting* **37**(3), 1212–1226.

Chan, J. C. C. (2022), 'Asymmetric conjugate priors for large Bayesian VARs', *Quantitative Economics* **13**(3), 1145–1169.

Chan, J. C. C. (2023*a*), 'Comparing stochastic volatility specifications for large Bayesian VARs', *Journal of Econometrics* **235**(2), 1419–1446.

Chan, J. C. C. (2023*b*), 'Large hybrid time-varying parameter VARs', *Journal of Business and Economic Statistics* **41**(3), 890–905.

Chan, J. C. C. (2024), BVARs and stochastic volatility, *in* M. P. Clements and A. B. Galvão, eds, 'Handbook of Research Methods and Applications in Macroeconomic Forecasting', Edward Elgar Publishing, pp. 43–67.

Chan, J. C. C., Clark, T. E. and Koop, G. (2018), 'A new model of inflation, trend inflation, and long-run inflation expectations', *Journal of Money, Credit and Banking* **50**, 5–53.

Chan, J. C. C., Doucet, A., León-González, R. and Strachan, R. W. (2025), 'Multivariate stochastic volatility with co-heteroscedasticity', *Studies in Nonlinear Dynamics and Econometrics* **29**(3), 265–300.

Chan, J. C. C. and Eisenstat, E. (2015), 'Marginal likelihood estimation with the Cross-Entropy method', *Econometric Reviews* **34**(3), 256–285.

Chan, J. C. C. and Eisenstat, E. (2018), 'Bayesian model comparison for time-varying parameter VARs with stochastic volatility', *Journal of Applied Econometrics* **33**(4), 509–532.

Chan, J. C. C., Eisenstat, E. and Strachan, R. W. (2020), 'Reducing the state space dimension in a large TVP-VAR', *Journal of Econometrics* **218**(1), 105–118.

Chan, J. C. C. and Grant, A. L. (2015), 'Pitfalls of estimating the marginal likelihood using the modified harmonic mean', *Economics Letters* **131**, 29–33.

Chan, J. C. C. and Grant, A. L. (2016*a*), 'Fast computation of the deviance information criterion for latent variable models', *Computational Statistics and Data Analysis* **100**, 847–859.

Chan, J. C. C. and Grant, A. L. (2016*b*), 'Modeling energy price dynamics: GARCH versus stochastic volatility', *Energy Economics* **54**, 182–189.

Chan, J. C. C. and Grant, A. L. (2016*c*), 'On the observed-data deviance information

criterion for volatility modeling', *Journal of Financial Econometrics* **14**(4), 772–802.

Chan, J. C. C. and Jeliazkov, I. (2009), 'Efficient simulation and integrated likelihood estimation in state space models', *International Journal of Mathematical Modelling and Numerical Optimisation* **1**(1/2), 101–120.

Chan, J. C. C., Koop, G., Poirier, D. J. and Tobias, J. L. (2019), *Bayesian Econometric Methods*, second edn, Cambridge University Press.

Chan, J. C. C., Koop, G. and Potter, S. M. (2013), 'A new model of trend inflation', *Journal of Business and Economic Statistics* **31**(1), 94–106.

Chan, J. C. C., Koop, G. and Potter, S. M. (2016), 'A bounded model of time variation in trend inflation, NAIRU and the Phillips curve', *Journal of Applied Econometrics* **31**(3), 551–565.

Chan, J. C. C., Koop, G. and Yu, X. (2024), 'Large order-invariant Bayesian VARs with stochastic volatility', *Journal of Business and Economic Statistics* **42**(2), 825–837.

Chan, J. C. C. and Kroese, D. P. (2012), 'Improved cross-entropy method for estimation', *Statistics and Computing* **22**(5), 1031–1040.

Chan, J. C. C. and Kroese, D. P. (2025), *Statistical Modeling and Computation*, second edn, Springer.

Chan, J. C. C., Leon-Gonzalez, R. and Strachan, R. W. (2018), 'Invariant inference and efficient computation in the static factor model', *Journal of the American Statistical Association* **113**(522), 819–828.

Chan, J. C. C., Poon, A. and Zhu, D. (2023), 'High-dimensional conditionally Gaussian state space models with missing data', *Journal of Econometrics* **236**(1), 105468.

Chan, J. C. C. and Qi, Y. (2025), Large Bayesian tensor VARs with stochastic volatility, *in* S. Mazur and P. Österholm, eds, 'Recent Developments in Bayesian Econometrics and Their Applications: Festschrift in Honour of Sune Karlsson', Springer, Cham, pp. 23–45.

Chan, J. C. C. and Qi, Y. (2026), 'Large Bayesian matrix autoregressions', *Journal of Econometrics* **256**, 105955.

Chan, J. C. C. and Yu, X. (2022), 'Fast and accurate variational inference for large Bayesian VARs with stochastic volatility', *Journal of Economic Dynamics and Control* **143**, 104505.

Chauvet, M. and Piger, J. (2008), 'A comparison of the real-time performance of business cycle dating methods', *Journal of Business and Economic Statistics* **26**(1), 42–49.

Chib, S. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–1321.

Chib, S. and Greenberg, E. (1994), 'Bayes inference in regression models with ARMA(p, q) errors', *Journal of Econometrics* **64**(1–2), 183–206.

Chib, S. and Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**(4), 327–335.

Chib, S. and Jeliazkov, I. (2001), 'Marginal likelihood from the Metropolis-Hastings output', *Journal of the American Statistical Association* **96**, 270–281.

Chib, S., Nardari, F. and Shephard, N. (2002), 'Markov chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics* **108**(2), 281–316.

Chib, S., Nardari, F. and Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**(2), 341–371.

Chiu, C. J., Mumtaz, H. and Pinter, G. (2017), 'Forecasting with VAR models: Fat tails and stochastic volatility', *International Journal of Forecasting* **33**(4), 1124–1143.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F. and Sokol, A. (2022), 'Nowcasting with large Bayesian vector autoregressions', *Journal of Econometrics* **231**(2), 500–519.

Clark, P. K. (1987), 'The cyclical component of US economic activity', *The Quarterly Journal of Economics* **102**(4), 797–814.

Clark, T. E. (2011), 'Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility', *Journal of Business and Economic Statistics* **29**(3), 327–341.

Clark, T. E. and Doh, T. (2014), 'A Bayesian evaluation of alternative models of trend inflation', *International Journal of Forecasting* **30**(3), 426–448.

Clark, T. E., Huber, F., Koop, G. and Marcellino, M. (2024), 'Forecasting U.S. inflation using Bayesian nonparametric models', *The Annals of Applied Statistics* **18**(2), 1421–1444.

Clark, T. E., Huber, F., Koop, G., Marcellino, M. and Pfarrhofer, M. (2023), 'Tail forecasting with multivariate Bayesian additive regression trees', *International Economic Review* **64**(3), 979–1022.

Clark, T. E. and Mertens, E. (2023), Stochastic volatility in Bayesian vector autoregressions, *in* 'Oxford Research Encyclopedia of Economics and Finance', Oxford University Press.

Clark, T. E. and Ravazzolo, F. (2015), 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics* **30**(4), 551–575.

Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010), 'Inflation-gap persistence in the U.S.', *American Economic Journal: Macroeconomics* **2**, 43–69.

Cogley, T. and Sargent, T. J. (2001), 'Evolving post-world war II US inflation dynamics', *NBER Macroeconomics Annual* **16**, 331–388.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: Monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics* **8**(2), 262–302.

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J. and Piatek, R. (2014), 'Bayesian exploratory factor analysis', *Journal of Econometrics* **183**(1), 31–57.

Crawford, L., Wood, K. C., Zhou, X. and Mukherjee, S. (2018), 'Bayesian approximate kernel regression with variable selection', *Journal of the American Statistical Association* **113**(524), 1710–1721.

Cross, J. L., Hou, C., Koop, G. and Poon, A. (2023), 'Large stochastic volatility in mean VARs', *Journal of Econometrics* **236**(1), 105469.

Cross, J. L., Hou, C. and Poon, A. (2020), 'Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity', *International Journal of Forecasting* **36**(3), 899–915.

Cross, J. L. and Poon, A. (2016), 'Forecasting structural change and fat-tailed events in Australian macroeconomic variables', *Economic Modelling* **58**, 34–51.

Crump, R. K., Eusepi, S., Giannone, D., Qian, E. and Sbordone, A. M. (2025), 'A large Bayesian VAR of the U.S. economy', *International Journal of Central Banking* **21**(2), 351–409.

Cukierman, A. and Meltzer, A. H. (1986), 'A theory of ambiguity, credibility, and inflation under discretion and asymmetric information', *Econometrica* **54**(5), 1099–1128.

D'Agostino, A., Gambetti, L. and Giannone, D. (2013), 'Macroeconomic forecasting and structural change', *Journal of Applied Econometrics* **28**, 82–101.

de Jong, P. and Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**, 339–350.

De Mol, C., Giannone, D. and Reichlin, L. (2008), 'Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?', *Journal of Econometrics* **146**(2), 318–328.

Del Moral, P., Doucet, A. and Jasra, A. (2006), 'Sequential Monte Carlo samplers', *Journal of the Royal Statistical Society: Series B* **68**(3), 411–436.

Del Negro, M. and Primiceri, G. E. (2015), 'Time-varying structural vector autoregressions and monetary policy: A corrigendum', *Review of Economic Studies* **82**(4), 1342–1345.

Del Negro, M. and Schorfheide, F. (2004), 'Priors from general equilibrium models for VARs', *International Economic Review* **45**(2), 643–673.

Del Negro, M. and Schorfheide, F. (2011), Bayesian macroeconometrics, *in* J. Geweke, G. Koop and H. van Dijk, eds, 'The Oxford Handbook of Bayesian Econometrics', Oxford University Press, pp. 293–389.

Demirer, M., Diebold, F. X., Liu, L. and Yilmaz, K. (2018), 'Estimating global bank network connectedness', *Journal of Applied Econometrics* **33**(1), 1–15.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society: Series B* **39**(1), 1–38.

Diebolt, J. and Robert, C. P. (1994), 'Estimation of finite mixture distributions through Bayesian sampling', *Journal of the Royal Statistical Society: Series B* **56**(2), 363–375.

Doan, T., Litterman, R. and Sims, C. A. (1984), 'Forecasting and conditional projection using realistic prior distributions', *Econometric Reviews* **3**(1), 1–100.

Doucet, A., De Freitas, N. and Gordon, N., eds (2001), *Sequential Monte Carlo Methods in Practice*, Springer, New York.

Doucet, A. and Johansen, A. M. (2011), A tutorial on particle filtering and smoothing: Fifteen years later, *in* D. Crisan and B. Rozovskii, eds, 'The Oxford Handbook of Nonlinear Filtering', Oxford University Press, Oxford, pp. 656–704.

Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987), 'Hybrid Monte Carlo', *Physics Letters B* **195**(2), 216–222.

Dufour, J.-M. and Pelletier, D. (2022), 'Practical methods for modeling weak VARMA processes: Identification, estimation and specification with a macroeconomic application', *Journal of Business and Economic Statistics* **40**(3), 1140–1152.

Durbin, J. and Koopman, S. J. (1997), 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika* **84**, 669–684.

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

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

Eberlein, E. and Keller, U. (1995), 'Hyperbolic distributions in finance', *Bernoulli* **1**(3), 281–299.

Eisenstat, E., Chan, J. C. C. and Strachan, R. W. (2016), 'Stochastic model specification search for time-varying parameter VARs', *Econometric Reviews* **35**(8-10), 1638–1665.

Ellahie, A. and Ricco, G. (2017), 'Government purchases reloaded: Informational insufficiency and heterogeneity in fiscal VARs', *Journal of Monetary Economics* **90**, 13–27.

Eltoft, T., Kim, T. and Lee, T. (2006), 'On the multivariate Laplace distribution', *Signal Processing Letters, IEEE* **13**(5), 300–303.

Engle, R. F., Lilien, D. M. and Robins, R. P. (1987), 'Estimating time varying risk premia in the term structure: the ARCH-M model', *Econometrica* **55**(2), 391–407.

Escobar, M. D. and West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.

Faust, J. (1998), The robustness of identified VAR conclusions about money, *in* 'Carnegie-Rochester Conference Series on Public Policy', Vol. 49, pp. 207–244.

Ferguson, T. S. (1973), 'A Bayesian analysis of some nonparametric problems', *The Annals of Statistics* **1**, 209–230.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007), 'Estimating macroeconomic models: A likelihood approach', *Review of Economic Studies* **74**(4), 1059–1087.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003), 'Do financial variables help forecasting inflation and real activity in the euro area?', *Journal of Monetary Economics* **50**(6), 1243–1255.

Friel, N. and Pettitt, A. N. (2008), 'Marginal likelihood estimation via power posteriors', *Journal of the Royal Statistical Society: Series B* **70**(3), 589–607.

Frühwirth-Schnatter, S. (1994), 'Data augmentation and dynamic linear models', *Journal of Time Series Analysis* **15**, 183–202.

Frühwirth-Schnatter, S. (2001), 'Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models', *Journal of the American Statistical Association* **96**(453), 194–209.

Frühwirth-Schnatter, S. (2006), *Finite Mixture and Markov Switching Models*, Springer, New York.

Frühwirth-Schnatter, S., Celeux, G. and Robert, C. P., eds (2019), *Handbook of Mixture Analysis*, Chapman & Hall/CRC.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007), 'Auxiliary mixture sampling with applications to logistic models', *Computational Statistics and Data Analysis* **51**(7), 3509–3528.

Frühwirth-Schnatter, S., Hosszejni, D. and Lopes, H. F. (2023), 'When it counts—econometric identification of the basic factor model based on GLT structures', *Econometrics* **11**(4), 1–30.

Frühwirth-Schnatter, S., Hosszejni, D. and Lopes, H. F. (2025), 'Sparse Bayesian factor analysis when the number of factors is unknown', *Bayesian Analysis* **20**(1), 213–344.

Frühwirth-Schnatter, S. and Kaufmann, S. (2008), 'Model-based clustering of multiple time series', *Journal of Business and Economic Statistics* **26**(1), 78–89.

Frühwirth-Schnatter, S. and Malsiner-Walli, G. (2019), 'From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering', *Advances in Data Analysis and Classification* **13**(1), 33–64.

Frühwirth-Schnatter, S. and Wagner, H. (2006), 'Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling', *Biometrika* **93**, 827–841.

Frühwirth-Schnatter, S. and Wagner, H. (2008), 'Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling', *Computational Statistics and Data Analysis* **52**(10), 4608–4624.

Frühwirth-Schnatter, S. and Wagner, H. (2010), 'Stochastic model specification search for Gaussian and partial non-Gaussian state space models', *Journal of Econometrics* **154**, 85–100.

Galí, J. (1999), 'Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?', *American Economic Review* **89**(1), 249–271.

Galí, J. and Gambetti, L. (2009), 'On the sources of the Great Moderation', *American Economic Journal: Macroeconomics* **1**(1), 26–57.

Garreau, D., Jitkrittum, W. and Kanagawa, M. (2017), Large sample analysis of the median heuristic. arXiv preprint arXiv:1707.07269.

Gefang, D. and Strachan, R. W. (2009), 'Nonlinear impacts of international business cycles on the U.K.—a Bayesian smooth transition VAR approach', *Studies in Nonlinear Dynamics & Econometrics* **14**(1), 1–33.

Gelfand, A. E. and Dey, D. K. (1994), 'Bayesian model choice: Asymptotics and exact calculations', *Journal of the Royal Statistical Society: Series B* **56**(3), 501–514.

Gelfand, A. E. and Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**(410), 398–409.

Gelman, A. (1996), Inference and monitoring convergence, *in* W. R. Gilks, S. Richardson and D. Spiegelhalter, eds, 'Markov Chain Monte Carlo in Practice', Chapman & Hall/CRC, Boca Raton, pp. 131–143.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian Data Analysis*, third edn, Chapman & Hall/CRC, Boca Raton.

Gelman, A. and Meng, X.-L. (1998), 'Simulating normalizing constants: From importance sampling to bridge sampling to path sampling', *Statistical Science* **13**(2), 163–185.

Gelman, A. and Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical Science* **7**(4), 457–472.

Geman, S. and Geman, D. (1984), 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6), 721–741.

George, E. I. and McCulloch, R. E. (1993), 'Variable selection via Gibbs sampling', *Journal of the American Statistical Association* **88**(423), 881–889.

Gerlach, R. H., Chen, C. W. S. and Chan, N. Y. C. (2011), 'Bayesian time-varying quantile forecasting for value-at-risk in financial markets', *Journal of Business and Economic Statistics* **29**(4), 481–492.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, pp. 169–193.

Geweke, J. (1993), 'Bayesian treatment of the independent Student-t linear model', *Journal of Applied Econometrics* **8**, S19–S40.

Geweke, J. (1996), 'Bayesian reduced rank regression in econometrics', *Journal of Econometrics* **75**(1), 121–146.

Geweke, J. (1999), 'Using simulation methods for Bayesian econometric models: inference, development, and communication', *Econometric Reviews* **18**(1), 1–73.

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, John Wiley & Sons, Hoboken, NJ.

Geweke, J. and Amisano, G. (2011), 'Optimal prediction pools', *Journal of Econometrics* **164**(1), 130–141.

Geweke, J. and Zhou, G. (1996), 'Measuring the pricing error of the arbitrage pricing theory', *The Review of Financial Studies* **9**, 557–587.

Ghosal, S. and van der Vaart, A. W. (2017), *Fundamentals of Nonparametric Bayesian Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Giannone, D., Lenza, M. and Primiceri, G. E. (2019), 'Priors for the long run', *Journal of the American Statistical Association* **114**(526), 565–580.

Giannone, D., Lenza, M. and Primiceri, G. E. (2021), 'Economic predictions with big data: The illusion of sparsity', *Econometrica* **89**(5), 2409–2437.

Giannone, D., Reichlin, L. and Small, D. (2008), 'Nowcasting: The real-time informational content of macroeconomic data', *Journal of Monetary Economics* **55**(4), 665–676.

Golub, G. H. and Van Loan, C. F. (2013), *Matrix Computations*, 4th edn, Johns Hopkins University Press, Baltimore.

Götz, T. and Hauzenberger, K. (2021), 'Large mixed-frequency VARs with a parsimonious time-varying parameter structure', *The Econometrics Journal* **24**(3), 442–461.

Grant, A. L. and Chan, J. C. C. (2017*a*), 'A Bayesian model comparison for trend-cycle decompositions of output', *Journal of Money, Credit and Banking* **49**(2-3), 525–552.

Grant, A. L. and Chan, J. C. C. (2017*b*), 'Reconciling output gaps: Unobserved components model and Hodrick-Prescott filter', *Journal of Economic Dynamics and Control* **75**, 114–121.

Green, P. J. (1995), 'Reversible jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika* **82**(4), 711–732.

Greenberg, E. (2013), *Introduction to Bayesian Econometrics*, second edn, Cambridge University Press.

Griffin, J. E. and Brown, P. (2010), 'Inference with normal-gamma prior distributions in regression problems', *Bayesian Analysis* **5**(1), 171–188.

Griffin, J. E., Łatuszyński, K. and Steel, M. F. J. (2021), 'In search of lost mixing time: Adaptive Markov chain Monte Carlo schemes for Bayesian variable selection with very large p', *Biometrika* **108**(1), 53–69.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.

Han, C. and Carlin, B. P. (2001), 'Markov chain Monte Carlo methods for computing Bayes factors: a comparative review', *Journal of the American Statistical Association* **96**, 1122–1132.

Hartwig, B. (2024), 'Bayesian VARs and prior calibration in times of COVID-19', *Studies in Nonlinear Dynamics and Econometrics* **28**(1), 1–24.

Harvey, A. C. (1985), 'Trends and cycles in macroeconomic time series', *Journal of Business and Economic Statistics* **3**(3), 216–227.

Harvey, A. C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge university press.

Harvey, A. C. and Jaeger, A. (1993), 'Detrending, stylized facts and the business cycle', *Journal of Applied Econometrics* **8**(3), 231–247.

Harvey, A. C. and Todd, P. H. J. (1983), 'Forecasting economic time series with structural and Box-Jenkins models: A case study', *Journal of Business and Economic Statistics* **1**(4), 299–307.

Harville, D. A. (1997), *Matrix Algebra From a Statistician's Perspective*, Springer, New York.

Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.

Hautsch, N. and Voigt, S. (2019), 'Large-scale portfolio allocation under transaction costs and model uncertainty', *Journal of Econometrics* **212**(1), 221–240.

Hauzenberger, N., Huber, F., Koop, G. and Mitchell, J. (2022), Bayesian modeling of time-varying parameters using regression trees. arXiv preprint arXiv:2209.11970.

Hauzenberger, N., Huber, F., Marcellino, M. and Petz, N. (2025), 'Gaussian process vector autoregressions and macroeconomic uncertainty', *Journal of Business and Economic Statistics* **43**(1), 27–43.

Hauzenberger, N., Marcellino, M., Pfarrhofer, M. and Stelzer, A. (2026), Direct Gaussian process predictive regressions with mixed frequency data. CEPR Discussion Paper 21214.

Herbst, E. P. and Schorfheide, F. (2014), 'Sequential Monte Carlo sampling for DSGE models', *Journal of Applied Econometrics* **29**(7), 1073–1098.

Herbst, E. P. and Schorfheide, F. (2016), *Bayesian estimation of DSGE models*, Princeton University Press.

Hesterberg, T. (1995), 'Weighted average importance sampling and defensive mixture distributions', *Technometrics* **37**(2), 185–194.

Hiraki, D., Chib, S. and Omori, Y. (2026), 'Stochastic volatility in mean: Efficient analysis by a generalized mixture sampler', *Journal of Econometrics* **256**, 105949.

Hodrick, R. J. and Prescott, E. C. (1980), 'Postwar US business cycles: An empirical investigation', *Carnegie Mellon University discussion paper* **451**.

Hodrick, R. J. and Prescott, E. C. (1997), 'Postwar US business cycles: An empirical investigation', *Journal of Money, Credit and Banking* **29**(1), 1–16.

Hoffman, M. D. and Gelman, A. (2014), 'The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo', *Journal of Machine Learning Research* **15**(47), 1593–1623.

Holland, A. S. (1995), 'Inflation and uncertainty: Tests for temporal ordering', *Journal of Money, Credit and Banking* **27**(3), 827–837.

Hou, C. (2017), 'Infinite hidden Markov switching VARs with application to macroeconomic forecast', *International Journal of Forecasting* **33**(4), 1025–1043.

Huber, F. and Feldkircher, M. (2019), 'Adaptive shrinkage in Bayesian vector autoregressive models', *Journal of Business and Economic Statistics* **37**(1), 27–39.

Huber, F., Koop, G. and Onorante, L. (2021), 'Inducing sparsity and shrinkage in time-varying parameter models', *Journal of Business and Economic Statistics* **39**(3), 669–683.

Huber, F. and Rossini, L. (2022), 'Inference in Bayesian additive vector autoregressive tree models', *The Annals of Applied Statistics* **16**(1), 104–123.

Hyvärinen, A., Zhang, K., Shimizu, S. and Hoyer, P. O. (2010), 'Estimation of a

structural vector autoregression model using non-Gaussianity', *Journal of Machine Learning Research* **11**, 1709–1731.

Ishwaran, H. and James, L. F. (2001), 'Gibbs sampling methods for stick-breaking priors', *Journal of the American Statistical Association* **96**(453), 161–173.

Jarociński, M. and Karadi, P. (2020), 'Deconstructing monetary policy surprises—the role of information shocks', *American Economic Journal: Macroeconomics* **12**(2), 1–43.

Jensen, M. J. and Maheu, J. M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**(2), 306–316.

Jin, X., Maheu, J. M. and Yang, Q. (2019), 'Bayesian parametric and semiparametric factor models for large realized covariance matrices', *Journal of Applied Econometrics* **34**(5), 641–660.

Jochmann, M. (2015), 'Modeling U.S. inflation dynamics: A Bayesian nonparametric approach', *Econometric Reviews* **34**(5), 537–558.

Jordà, Ò. (2005), 'Estimation and inference of impulse responses by local projections', *American Economic Review* **95**(1), 161–182.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kadiyala, K. and Karlsson, S. (1997), 'Numerical methods for estimation and inference in Bayesian VAR-models', *Journal of Applied Econometrics* **12**(2), 99–132.

Kalli, M., Griffin, J. E. and Walker, S. G. (2011), 'Slice sampling mixture models', *Statistics and Computing* **21**(1), 93–105.

Karlsson, S. (2013), Forecasting with Bayesian vector autoregressions, *in* G. Elliott and A. Timmermann, eds, 'Handbook of Economic Forecasting', Vol. 2 of *Handbook of Economic Forecasting*, Elsevier, pp. 791–897.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kastner, G. (2019), 'Sparse Bayesian time-varying covariance estimation in many dimensions', *Journal of Econometrics* **210**(1), 98–115.

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

Kastner, G. and Huber, F. (2020), 'Sparse Bayesian vector autoregressions in huge dimensions', *Journal of Forecasting* **39**(7), 1142–1165.

Kaufmann, S. and Schumacher, C. (2017), 'Identifying relevant and irrelevant variables in sparse factor models', *Journal of Applied Econometrics* **32**(6), 1123–1144.

Kaufmann, S. and Schumacher, C. (2019), 'Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification', *Journal of Econometrics* **210**(1), 116–134.

Kilian, L. (2009), 'Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market', *American Economic Review* **99**(3), 1053–1069.

Kilian, L. and Lütkepohl, H. (2017), *Structural Vector Autoregressive Analysis*, Cambridge University Press.

Kim, C. J. and Nelson, C. R. (1998), 'Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching', *Review of Economics and Statistics* **80**(2), 188–201.

Kim, C. J. and Nelson, C. R. (1999), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press, Cambridge.

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

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991), 'Stochastic trends and economic fluctuations', *American Economic Review* **81**(4), 819–840.

Kleibergen, F. and Paap, R. (2002), 'Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration', *Journal of Econometrics* **111**(2), 223–249.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons, New York.

Koop, G. (2013), 'Forecasting with medium and large Bayesian VARs', *Journal of Applied Econometrics* **28**(2), 177–203.

Koop, G. and Korobilis, D. (2010), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**(4), 267–358.

Koop, G. and Korobilis, D. (2013), 'Large time-varying parameter VARs', *Journal of Econometrics* **177**(2), 185–198.

Koop, G. and Korobilis, D. (2023), 'Bayesian dynamic variable selection in high dimensions', *International Economic Review* **64**(3), 1047–1074.

Koop, G., Korobilis, D. and Pettenuzzo, D. (2019), 'Bayesian compressed vector autoregressions', *Journal of Econometrics* **210**(1), 135–154.

Koop, G., Leon-Gonzalez, R. and Strachan, R. W. (2009), 'On the evolution of the monetary policy transmission mechanism', *Journal of Economic Dynamics and Control* **33**(4), 997–1017.

Koop, G., Leon-Gonzalez, R. and Strachan, R. W. (2010*a*), 'Dynamic probabilities of restrictions in state space models: an application to the Phillips curve', *Journal of Business and Economic Statistics* **28**(3), 370–379.

Koop, G., Leon-Gonzalez, R. and Strachan, R. W. (2010*b*), 'Efficient posterior simulation for cointegrated models with priors on the cointegration space', *Econometric Reviews* **29**(2), 224–242.

Koop, G., McIntyre, S., Mitchell, J. and Poon, A. (2020), 'Regional output growth in the United Kingdom: more timely and higher frequency estimates from 1970', *Journal of Applied Econometrics* **35**(2), 176–197.

Koop, G. and Potter, S. M. (1999), 'Bayes factors and nonlinearity: evidence from economic time series', *Journal of Econometrics* **88**(2), 251–281.

Koop, G., Strachan, R. W., van Dijk, H. K. and Villani, M. (2006), Bayesian approaches to cointegration, *in* T. C. Mills and K. Patterson, eds, 'The Palgrave Handbook of Econometrics, Volume 1: Econometric Theory', Palgrave Macmillan, Basingstoke, pp. 871–898.

Koopman, S. J. and Hol Uspensky, E. (2002), 'The stochastic volatility in mean model: Empirical evidence from international stock markets', *Journal of Applied Econometrics* **17**(6), 667–689.

Korobilis, D. and Pettenuzzo, D. (2019), 'Adaptive hierarchical priors for high-dimensional vector autoregressions', *Journal of Econometrics* **212**(1), 241–271.

Korobilis, D. and Shimizu, K. (2022), 'Bayesian approaches to shrinkage and sparse estimation', *Foundations and Trends in Econometrics* **11**(4), 230–354.

Kose, M. A., Otrok, C. and Whiteman, C. H. (2003), 'International business cycles: World, region, and country-specific factors', *American Economic Review* **93**(4), 1216–1239.

Kozumi, H. and Kobayashi, G. (2011), 'Gibbs sampling methods for Bayesian quantile regression', *Journal of Statistical Computation and Simulation* **81**(11), 1565–1578.

Kroese, D. P., Taimre, T. and Botev, Z. I. (2011), *Handbook of Monte Carlo Methods*, John Wiley & Sons, New York.

Kuttner, K. N. (1994), 'Estimating potential output as a latent variable', *Journal of Business and Economic Statistics* **12**(3), 361–368.

Kyung, M., Gill, J., Ghosh, M. and Casella, G. (2010), 'Penalized regression, standard errors, and Bayesian lassos', *Bayesian Analysis* **5**(2), 369–412.

Lancaster, T. (2004), *An Introduction to Modern Bayesian Econometrics*, Blackwell Publishing, Oxford.

Lanne, M., Meitz, M. and Saikkonen, P. (2017), 'Identification and estimation of non-Gaussian structural vector autoregressions', *Journal of Econometrics* **196**(2), 288–304.

Leng, C., Tran, M.-N. and Nott, D. (2014), 'Bayesian adaptive Lasso', *Annals of the Institute of Statistical Mathematics* **66**(2), 221–244.

Lenza, M. and Primiceri, G. E. (2022), 'How to estimate a VAR after March 2020', *Journal of Applied Econometrics* **37**(4), 688–699.

LeSage, J. P. and Hendrikz, D. (2019), 'Large Bayesian vector autoregressive forecasting for regions: A comparison of methods based on alternative disturbance structures', *The Annals of Regional Science* **62**(3), 563–599.

Lewis, D. J. (2021), 'Identifying shocks via time-varying volatility', *Review of Economic Studies* **88**(6), 3086–3124.

Li, M. and Scharth, M. (2022), 'Leverage, asymmetry, and heavy tails in the high-dimensional factor stochastic volatility model', *Journal of Business and Economic Statistics* **40**(1), 285–301.

Li, Q., Lin, N. and Xi, R. (2010), 'Bayesian regularized quantile regression', *Bayesian Analysis* **5**(3), 533–556.

Li, Y., Wang, N. and Yu, J. (2023), 'Improved marginal likelihood estimation via power posteriors and importance sampling', *Journal of Econometrics* **234**(1), 28–52.

Li, Y. and Yu, J. (2012), 'Bayesian hypothesis testing in latent variable models', *Journal of Econometrics* **166**(2), 237–246.

Li, Y., Yu, J. and Zeng, T. (2020), 'Deviance information criterion for latent variable models and misspecified models', *Journal of Econometrics* **216**(2), 450–493.

Li, Y., Zeng, T. and Yu, J. (2014), 'A new approach to Bayesian hypothesis testing', *Journal of Econometrics* **178**(3), 602–612.

Linero, A. R. (2025), Defensive model expansion for robust Bayesian inference. arXiv preprint arXiv:2510.09598.

Litterman, R. (1986), 'Forecasting with Bayesian vector autoregressions — five years of experience', *Journal of Business and Economic Statistics* **4**, 25–38.

Liu, J. S. (1994), 'The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem', *Journal of the American Statistical Association* **89**(427), 958–966.

Liu, Y. and Morley, J. C. (2014), 'Structural evolution of the postwar U.S. economy', *Journal of Economic Dynamics and Control* **42**, 50–68.

Llorente, F., Martino, L., Delgado, D. and López-Santiago, J. (2023), 'Marginal likelihood computation for model selection and hypothesis testing: An extensive review', *SIAM Review* **65**(1), 3–58.

Lopes, H. F., McCulloch, R. E. and Tsay, R. S. (2022), 'Parsimony inducing priors for large scale state-space models', *Journal of Econometrics* **230**(1), 39–61.

Louzis, D. P. (2019), 'Steady-state modeling and macroeconomic forecasting quality', *Journal of Applied Econometrics* **34**(2), 285–314.

Luo, S. and Startz, R. (2014), 'Is it one break or ongoing permanent shocks that explains US real GDP?', *Journal of Monetary Economics* **66**, 155–163.

Luo, Y. and Griffin, J. E. (2025), 'Bayesian inference of vector autoregressions with tensor decompositions', *Journal of Business and Economic Statistics* **43**(4), 941–955.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.

Magnus, J. R. and Neudecker, H. (2019), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edn, John Wiley & Sons, Chichester.

Maheu, J. M. and Yang, Q. (2016), 'An infinite hidden Markov model for short-term interest rates', *Journal of Empirical Finance* **38**, 202–220.

Makalic, E. and Schmidt, D. F. (2016), 'A simple sampler for the horseshoe estimator', *IEEE Signal Processing Letters* **23**(1), 179–182.

Mariano, R. S. and Murasawa, Y. (2003), 'A new coincident index of business cycles based on monthly and quarterly series', *Journal of Applied Econometrics* **18**(4), 427–443.

McCausland, W. J. (2012), 'The HESSIAN method: Highly efficient simulation smoothing, in a nutshell', *Journal of Econometrics* **168**(2), 189–206.

McCausland, W. J., Miller, S. and Pelletier, D. (2011), 'Simulation smoothing for state-space models: A computational efficiency analysis', *Computational Statistics and Data Analysis* **55**(1), 199–212.

McCausland, W. J., Miller, S. and Pelletier, D. (2021), 'Multivariate stochastic volatility using the HESSIAN method', *Econometrics and Statistics* **17**, 76–94.

McCracken, M. W. and Ng, S. (2016), 'FRED-MD: A monthly database for macroeconomic research', *Journal of Business and Economic Statistics* **34**(4), 574–589.

McLachlan, G. and Peel, D. (2000), *Finite Mixture Models*, John Wiley & Sons, New York.

Meng, X.-L. and Wong, W. H. (1996), 'Simulating ratios of normalizing constants via a simple identity: A theoretical exploration', *Statistica Sinica* **6**, 831–860.

Mertens, E. (2023), 'Precision-based sampling for state space models that have no measurement error', *Journal of Economic Dynamics and Control* **154**, 104720.

Mertens, K. and Ravn, M. O. (2013), 'The dynamic effects of personal and corporate income tax changes in the United States', *American Economic Review* **103**(4), 1212–1247.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

Millar, R. B. (2009), 'Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes factors', *Biometrics* **65**(3), 962–969.

Mitchell, T. J. and Beauchamp, J. J. (1988), 'Bayesian variable selection in linear regression', *Journal of the American Statistical Association* **83**(404), 1023–1032.

Morley, J. C., Nelson, C. R. and Zivot, E. (2003), 'Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?', *Review of Economics and Statistics* **85**(2), 235–243.

Mumtaz, H. (2016), 'The evolving transmission of uncertainty shocks in the United Kingdom', *Econometrics* **4**(1), 16.

Mumtaz, H. and Theodoridis, K. (2018), 'The changing transmission of uncertainty shocks in the U.S.', *Journal of Business and Economic Statistics* **36**(2), 239–252.

Mumtaz, H. and Zanetti, F. (2013), 'The impact of the volatility of monetary policy shocks', *Journal of Money, Credit and Banking* **45**(4), 535–558.

Nakajima, J. and Omori, Y. (2009), 'Leverage, heavy-tails and correlated jumps in stochastic volatility models', *Computational Statistics and Data Analysis* **53**(6), 2335–2353.

Nakajima, J. and Omori, Y. (2012), 'Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution', *Computational Statistics and Data Analysis* **56**(11), 3690–3704.

Nakajima, J. and West, M. (2013), 'Bayesian analysis of latent threshold dynamic models', *Journal of Business and Economic Statistics* **31**(2), 151–164.

Neal, R. M. (1996), *Bayesian Learning for Neural Networks*, Vol. 118 of *Lecture Notes in Statistics*, Springer, New York.

Neal, R. M. (2000), 'Markov chain sampling methods for Dirichlet process mixture models', *Journal of Computational and Graphical Statistics* **9**(2), 249–265.

Neal, R. M. (2003), 'Slice sampling', *The Annals of Statistics* **31**(3), 705–767.

Neal, R. M. (2011), MCMC using Hamiltonian dynamics, *in* S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds, 'Handbook of Markov Chain Monte Carlo', Chapman & Hall/CRC, Boca Raton, pp. 113–162.

Niemi, J. and West, M. (2010), 'Adaptive mixture modeling Metropolis methods for Bayesian analysis of nonlinear state-space models', *Journal of Computational and Graphical Statistics* **19**(2), 260–280.

Oh, K. H., Zivot, E. and Creal, D. (2008), 'The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics', *Journal of Econometrics* **146**(2), 207–219.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007), 'Stochastic volatility with leverage: Fast and efficient likelihood inference', *Journal of Econometrics* **140**(2), 425–449.

Park, T. and Casella, G. (2008), 'The Bayesian Lasso', *Journal of the American Statistical Association* **103**(482), 681–686.

Perron, P. and Wada, T. (2009), 'Let's take a break: Trends and cycles in US real GDP', *Journal of Monetary Economics* **56**(6), 749–765.

Philipov, A. and Glickman, M. (2006), 'Multivariate stochastic volatility via Wishart processes', *Journal of Business and Economic Statistics* **24**, 313–328.

Pitt, M. K. and Shephard, N. (1999), 'Time varying covariances: a factor stochastic volatility approach', *Bayesian Statistics* **6**, 547–570.

Plagborg-Møller, M., Reichlin, L., Ricco, G. and Hasenzagl, T. (2020), 'When is growth at risk?', *Brookings Papers on Economic Activity* **Spring**, 167–229.

Plagborg-Møller, M. and Wolf, C. K. (2021), 'Local projections and VARs estimate the same impulse responses', *Econometrica* **89**(2), 955–980.

Plumlee, M. and Joseph, V. R. (2018), 'Orthogonal Gaussian process models', *Statistica Sinica* **28**(2), 601–619.

Poirier, D. J. (1995), *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge, MA.

Polson, N. G. and Scott, J. G. (2011), 'Shrink globally, act locally: Sparse Bayesian regularization and prediction', *Bayesian Statistics* **9**, 501–538.

Polson, N. G., Scott, J. G. and Windle, J. (2013), 'Bayesian inference for logistic

models using Pólya-Gamma latent variables', *Journal of the American Statistical Association* **108**(504), 1339–1349.

Poon, A. (2018), 'Assessing the synchronicity and nature of Australian state business cycles', *Economic Record* **94**(307), 372–390.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

Prüser, J. and Huber, F. (2024), 'Nonlinearities in macroeconomic tail risk through the lens of big data quantile regressions', *Journal of Applied Econometrics* **39**(2), 269–291.

Quiñonero-Candela, J. and Rasmussen, C. E. (2005), 'A unifying view of sparse approximate Gaussian process regression', *Journal of Machine Learning Research* **6**, 1939–1959.

Raman, S., Fuchs, T. J., Wild, P. J., Dahl, E. and Roth, V. (2009), The Bayesian group-Lasso for analyzing contingency tables, *in* 'Proceedings of the 26th Annual International Conference on Machine Learning', ACM, pp. 881–888.

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.

Richardson, S. and Green, P. J. (1997), 'On Bayesian analysis of mixtures with an unknown number of components', *Journal of the Royal Statistical Society: Series B* **59**(4), 731–792.

Rigobon, R. (2003), 'Identification through heteroskedasticity', *Review of Economics and Statistics* **85**(4), 777–792.

Ritter, C. and Tanner, M. A. (1992), 'Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler', *Journal of the American Statistical Association* **87**(419), 861–868.

Robert, C. P. and Casella, G. (2004), *Monte Carlo Statistical Methods*, second edn, Springer, New York.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms', *Annals of Applied Probability* **7**(1), 110–120.

Roberts, G. O. and Rosenthal, J. S. (2001), 'Optimal scaling for various Metropolis-Hastings algorithms', *Statistical Science* **16**(4), 351–367.

Ročková, V. and George, E. I. (2018), 'The spike-and-slab Lasso', *Journal of the American Statistical Association* **113**(521), 431–444.

Romer, C. D. and Romer, D. H. (2004), 'A new measure of monetary shocks: Derivation and implications', *American Economic Review* **94**(4), 1055–1084.

Rubinstein, R. Y. (1997), 'Optimization of computer simulation models with rare events', *European Journal of Operational Research* **99**, 89–112.

Rubinstein, R. Y. (1999), 'The cross-entropy method for combinatorial and continuous optimization', *Methodology and Computing in Applied Probability* **1**(2), 127–190.

Rubinstein, R. Y. and Kroese, D. P. (2004), *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*, Springer, New York.

Rubio-Ramírez, J. F., Waggoner, D. F. and Zha, T. (2010), 'Structural vector autoregressions: Theory of identification and algorithms for inference', *Review of Economic Studies* **77**(2), 665–696.

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

Schorfheide, F. and Song, D. (2015), 'Real-time forecasting with a mixed-frequency VAR', *Journal of Business and Economic Statistics* **33**(3), 366–380.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.

Sentana, E. and Fiorentini, G. (2001), 'Identification, estimation and testing of conditionally heteroskedastic factor models', *Journal of Econometrics* **102**, 143–164.

Sethuraman, J. (1994), 'A constructive definition of Dirichlet priors', *Statistica Sinica* **4**, 639–650.

Shephard, N. and Pitt, M. K. (1997), 'Likelihood analysis of non-Gaussian measurement time series', *Biometrika* **84**, 653–667.

Shin, M. and Zhong, M. (2020), 'A new approach to identifying the real effects of uncertainty shocks', *Journal of Business and Economic Statistics* **38**(2), 367–379.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1–48.

Sims, C. A. (1993), A nine-variable probabilistic macroeconomic forecasting model, *in* J. H. Stock and M. W. Watson, eds, 'Business Cycles, Indicators, and Forecasting', University of Chicago Press, pp. 179–212.

Sims, C. A. (2001), 'Comment on Sargent and Cogley's 'evolving US postwar inflation dynamics", *NBER Macroeconomics Annual* **16**, 373–379.

Sims, C. A. and Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* **39**(4), 949–968.

Sims, C. A. and Zha, T. (2006), 'Were there regime switches in U.S. monetary policy?', *American Economic Review* **96**(1), 54–81.

Smith, M. and Kohn, R. (1996), 'Nonparametric regression using Bayesian variable selection', *Journal of Econometrics* **75**(2), 317–343.

Solin, A. and Särkkä, S. (2020), 'Hilbert space methods for reduced-rank Gaussian process regression', *Statistics and Computing* **30**(2), 419–446.

Song, Y. (2014), 'Modelling regime switching and structural breaks with an infinite hidden Markov model', *Journal of Applied Econometrics* **29**(5), 825–842.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B* **64**(4), 583–639.

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer, New York.

Stephens, M. (2000), 'Dealing with label switching in mixture models', *Journal of the Royal Statistical Society: Series B* **62**(4), 795–809.

Stock, J. H. and Watson, M. W. (2002), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and Economic Statistics* **20**(2), 147–162.

Stock, J. H. and Watson, M. W. (2007), 'Why has U.S. inflation become harder to forecast?', *Journal of Money, Credit and Banking* **39**(s1), 3–33.

Stock, J. H. and Watson, M. W. (2012), 'Disentangling the channels of the 2007–09 recession', *Brookings Papers on Economic Activity* **43**(1), 81–135.

Stock, J. H. and Watson, M. W. (2016), 'Core inflation and trend inflation', *Review of Economics and Statistics* **98**(4), 770–784.

Strachan, R. W. and Inder, B. (2004), 'Bayesian analysis of the error correction model', *Journal of Econometrics* **123**(2), 307–325.

Stroud, J. R., Müller, P. and Polson, N. G. (2003), 'Nonlinear state-space models with state-dependent variances', *Journal of the American Statistical Association* **98**(462), 377–386.

Tallman, E. W. and Zaman, S. (2020), 'Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy', *International Journal of Forecasting* **36**(2), 373–398.

Tanner, M. A. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation', *Journal of the American Statistical Association* **82**(398), 528–540.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the Lasso', *Journal of the Royal Statistical Society: Series B* **58**(1), 267–288.

Tierney, L. (1994), 'Markov chains for exploring posterior distributions', *The Annals of Statistics* **22**(4), 1701–1728.

Tokdar, S. T. and Kadane, J. B. (2012), 'Simultaneous linear quantile regression: A semiparametric Bayesian approach', *Bayesian Analysis* **7**(1), 51–72.

Uhlig, H. (1997), 'Bayesian vector autoregressions with stochastic volatility', *Econometrica* **65**(1), 59–73.

Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**(2), 381–419.

van der Vaart, A. W. and van Zanten, J. H. (2009), 'Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth', *The Annals of Statistics* **37**(5B), 2655–2675.

van Dyk, D. A. and Meng, X.-L. (2001), 'The art of data augmentation', *Journal of Computational and Graphical Statistics* **10**(1), 1–50.

van Dyk, D. A. and Park, T. (2008), 'Partially collapsed Gibbs samplers: Theory and methods', *Journal of the American Statistical Association* **103**(482), 790–796.

Vannucci, M. (2021), Discrete spike-and-slab priors: Models and computational aspects, *in* M. G. Tadesse and M. Vannucci, eds, 'Handbook of Bayesian Variable Selection', CRC Press, pp. 3–24.

Vehtari, A., Gelman, A. and Gabry, J. (2017), 'Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC', *Statistics and Computing* **27**(5), 1413–1432.

Verdinelli, I. and Wasserman, L. (1995), 'Computing Bayes factors using a generalization of the Savage-Dickey density ratio', *Journal of the American Statistical Association* **90**(430), 614–618.

Villani, M. (2005), 'Bayesian reference analysis of cointegration', *Econometric Theory* **21**(2), 326–357.

Villani, M. (2009), 'Steady-state priors for vector autoregressions', *Journal of Applied Econometrics* **24**(4), 630–650.

Waggoner, D. F. and Zha, T. (2003), 'A Gibbs sampler for structural vector autoregressions', *Journal of Economic Dynamics and Control* **28**(2), 349–366.

Wahba, G. (1990), *Spline Models for Observational Data*, Vol. 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia.

Walker, S. G. (2007), 'Sampling the Dirichlet mixture model with slices', *Communications in Statistics—Simulation and Computation* **36**(1), 45–54.

Wang, D., Zheng, Y., Lian, H. and Li, G. (2022), 'High-dimensional vector autoregressive time series modeling via tensor decomposition', *Journal of the American Statistical Association* **117**(539), 1338–1356.

Wang, J. J. J., Chan, J. S. K. and Choy, S. T. B. (2011), 'Stochastic volatility models with leverage and heavy-tailed distributions: A Bayesian approach using scale mixtures', *Computational Statistics and Data Analysis* **55**(1), 852–862.

Wang, J. J. J., Choy, S. T. B. and Chan, J. S. K. (2013), 'Modelling stochastic volatility using generalized t distribution', *Journal of Statistical Computation and Simulation* **83**(2), 340–354.

Watanabe, S. (2010), 'Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory', *Journal of Machine Learning Research* **11**, 3571–3594.

Watson, M. W. (1986), 'Univariate detrending methods with stochastic trends', *Journal of Monetary Economics* **18**(1), 49–75.

West, M. (1984), 'Outlier models and prior distributions in Bayesian linear regression', *Journal of the Royal Statistical Society: Series B* **46**(3), 431–439.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, 2 edn, Springer, New York.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L. and Chen, M.-H. (2011), 'Improving marginal likelihood estimation for Bayesian phylogenetic model selection', *Systematic Biology* **60**(2), 150–160.

Xu, Z., Schmidt, D. F., Makalic, E., Qian, G. and Hopper, J. L. (2016), Bayesian grouped horseshoe regression with application to additive models, *in* B. H. Kang and Q. Bai, eds, 'AI 2016: Advances in Artificial Intelligence', Springer, Cham, pp. 229–240.

Yang, Y. and He, X. (2012), 'Bayesian empirical likelihood for quantile regression', *The Annals of Statistics* **40**(2), 1102–1131.

Yu, J. (2005), 'On leverage in a stochastic volatility model', *Journal of Econometrics* **127**(2), 165–178.

Yu, K. and Moyeed, R. A. (2001), 'Bayesian quantile regression', *Statistics & Probability Letters* **54**(4), 437–447.

Zaman, S. (2025), 'A unified framework to estimate macroeconomic stars', *Review of Economics and Statistics* pp. 1–45.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons.

Zens, G., Böck, M. and Zörner, T. O. (2020), 'The heterogeneous impact of monetary policy on the US labor market', *Journal of Economic Dynamics and Control* **119**, 103989.

Zhang, Y. D., Naughton, B. P., Bondell, H. D. and Reich, B. J. (2022), 'Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior', *Journal of the American Statistical Association* **117**(538), 862–874.