Chapter 12
Bayesian Methods for Credit Risk Modeling

INTRODUCTION

Bayesian statistics is an alternative perspective of statistics compared to the classical frequentist approach. A good introduction on Bayesian statistics is given in Greenberg (2014). Simply speaking, the frequentist or classical statistician interprets probabilities as limits of empirical frequencies of realizations of random events when the number of repetitions of the random experiment goes to infinity. Hence, probabilities are objective. Model estimation is performed by maximizing the likelihood using sample information. In contrast, the Bayesian approach makes use of subjective probabilities in addition to information from the likelihood. The subjective probabilities make up the prior distribution of the events under consideration. In other words, information from the sample data is complemented by prior (subjective) information and assumptions. It can be shown that the estimation results using the Bayesian approach become more and more similar to the estimates from the classical approach when larger information is available. On the other hand, when only little information from the data is available, the prior approach is more dominant. Bayesian statistics can provide powerful tools, particularly for sparse data such as short time series or credit portfolios with small numbers of defaults. However, the prior assumptions have to be chosen wisely, as these may dominate the posterior outcome.

As an example, consider data y from a number of n random coin tosses. The frequentist approach treats the parameter (i.e., the probability π of "head") as an unknown parameter (i.e., a fixed, but unknown number) that can be estimated from the sample data. In contrast, the Bayesian approach considers this parameter to be unknown *and* a random variable itself and assigns a prior distribution $f(\pi), 0 \leq \pi \leq 1$ to it. It then links the prior distribution with observable data from the realized coin tosses and generates a posterior distribution $f(\pi|y)$, that is, the conditional distribution of π, given the data y. The goal is then to learn about the parameter from the data and update the prior distribution. The Bayesian view is therefore different from the frequentist approach in many perspectives. Parameter estimators are derived under the frequentist approach claiming criteria such as unbiasedness or consistency. The Bayesian approach obtains point estimates for parameters as minima of specific loss functions. The frequentist approach constructs confidence intervals and computes p-values, while the Bayesian method computes credibility intervals directly from the posterior distribution and compares different models (similar to hypotheses) via marginal likelihoods. The following tablecontrasts the main differences between the frequentist and the Bayesian approach.

	Frequentist	***Bayesian***
Probability view	Objective	Subjective
Data	Data are a repeatable random sample	Data are observed from a realized sample
Parameters	Parameters are unknown but fixed	Parameters are unknown but random
Estimation criterion	Unbiasedness, consistency	Minimizing a loss function
Parameter estimate	Point estimate from estimation approach	Location measure (e.g., mean) of posterior distribution
Interval	Confidence interval	Credibility interval from posterior distribution
Hypothesis testing	Via p -value	Via marginal likelihoods of various models

THE BAYESIAN APPROACH TO STATISTICS

To understand the Bayesian approach more formally, we start with Bayes' theorem. Consider two random events A and B. Bayes' theorem states that the conditional probability of A given B is

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Now, if we consider random data Y (e.g., coin tosses) and a parameter θ (e.g., the probability of a head occurring), Bayes' theorem gives the posterior distribution

$$f(\theta \mid y) = \frac{f(y \mid \theta)f(\theta)}{f(y)}$$

where $f(y) = \int f(y \mid \theta)f(\theta)\, d\theta$ is the unconditional distribution of the data and $f(y \mid \theta)$ is the likelihood function. This basically indicates that the conditional (posterior) distribution of the parameter θ given some realized data y is the ratio of the product of the likelihood and the prior distribution $f(\theta)$ of the parameter divided by the unconditional distribution of the data. The prior distribution may consist of further parameters (such as expectation and variance of the normal distribution), which are called hyper parameters and can also be modeled as random variables with hyper prior distributions and so forth.

The likelihood function is given by a model and the data as in the frequentist approach. The prior distribution is prespecified by the Bayesian statistician according to her or his beliefs and prior experiences or from other data sources. Unless we assume special cases for the likelihood and the prior, the posterior distribution cannot be evaluated analytically. Rather, the evaluation can be performed via a special Monte Carlo technique, the Markov chain–Monte Carlo (MCMC) method, which generates a large number of simulation trials via special sampling algorithms in order to approximate the posterior. In SAS, this is implemented in PROC MCMC.

Once the posterior distribution is generated, analogously to the frequentist approach, estimates of the moments of the posterior and so-called *credibility intervals* can be derived. However, instead of claiming specific properties for the estimators such as unbiasedness and consistency, the Bayesian criterion for creating an estimator is the minimization of a "loss function." Consider a loss function $L(\hat{\theta}, \theta)$, which specifies the loss incurred by using estimate $\hat{\theta}$ instead of the true θ. Examples are

- Absolute value loss function:

  $L_1(\hat{\theta}, \theta) = | \hat{\theta} - \theta |$

- Quadratic loss function:

  $L_2(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$

- Bilinear loss function:

$$L_3(\hat{\theta}, \theta) = \begin{cases} a \,|\, \hat{\theta} - \theta \,|\, \text{ for } \theta > \hat{\theta} \\ b \,|\, \hat{\theta} - \theta \,|\, \text{ for } \theta \leq \hat{\theta} \end{cases}$$

  where $a, b > 0$

The *Bayes estimator* for θ is the value $\hat{\theta}$ that minimizes the expected value of the loss, where expectation is taken over the posterior distribution; that is, $\hat{\theta}$ is chosen to minimize

$$E\left[L(\hat{\theta}, \theta)\right] = \int L(\hat{\theta}, \theta) f(\theta \mid y)\, d\theta$$

For example, under quadratic loss we minimize

$$E\left[L(\hat{\theta}, \theta)\right] = \int (\hat{\theta} - \theta)^2 f(\theta \mid y)\, d\theta$$

We then differentiate with respect to $\hat{\theta}$ and set the result equal to zero:

$$2 \int (\hat{\theta} - \theta) f(\theta \mid y)\, d\theta = 0$$

$$\hat{\theta} = \int \theta f(\theta \mid y)\, d\theta$$

Hence, under the quadratic loss function the Bayes estimator is $\hat{\theta} = E(\theta \mid y)$ (i.e., the expectation of the posterior distribution). For computing credibility intervals, the Bayesian approach reports an interval estimate of the form

$$P(\theta_L \leq \theta \leq \theta_U) = 0.95$$

In other words, it requires that the probability of the parameter θ being between a lower bound θ_L and an upper bound θ_U equals some (high) probability (e.g., 95% or 99%).

Given the posterior distribution, forecasts for out-of-sample (e.g., future) observations can be derived. Let y_f be some forecast, then its density, given the observed data y and the posterior distribution is given by

$$f(y_f \mid y) = \int f(y_f \mid \theta, y) f(\theta \mid y)\, d\theta$$

This implies that one computes the probability or density of y given the data and given some parameter value, and mixes over the (posterior) density of all potential parameter values.

PD ESTIMATION WITH BAYESIAN STATISTICS

Probit Analysis

We start introducing the Bayesian approach with PD estimation and compare it to the classical approach using a probit model. As the Bayesian approach generally requires evaluating the posterior distribution by MCMC, it might have a longer running time. For illustration purposes, we therefore draw a random sample from the entire data set of size 1 percent and the reader is encouraged to evaluate the estimation for the entire data set or larger subsamples. The sample is drawn using only those observations for which a randomly drawn uniformly distributed random variable is lower than 0.01. The probit model is then estimated as in the former chapter on PD estimation using three covariates.

Following our chapter on continuous-time hazard models, we prepare the data and draw a 1 percent random sample (see Exhibit 12.1):

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	−1.4138	0.4730	8.9331	0.0028
FICO_orig_time	1	−0.00224	0.000504	19.7709	<.0001
LTV_orig_time	1	0.0119	0.00360	10.9189	0.0010
gdp_time	1	−0.0676	0.0163	17.1275	<.0001

Exhibit 12.1 Probit Model with PROC LOGISTIC

```
/*Data preprocessing and selection of 1 percent random sample*/
DATA sample;
SET data.mortgage;
time1 = time-first_time;
time2 = time-first_time+1;
IF RANUNI(12345) < 0.01;
RUN;
```

```
/*Probit Model with PROC LOGISTIC*/
ODS GRAPHICS ON;
PROC LOGISTIC DATA=sample DESCENDING;
MODEL default_time = FICO_orig_time
LTV_orig_time gdp_time
 / OUTROC=roc_logistic RSQUARE LINK=Probit;
RUN;
ODS GRAPHICS OFF;
```

Next, we estimate the Bayesian model using PROC MCMC with the following code. As usual in this book, for computational details of this procedure we refer the reader to the SAS manual; see SAS Institute Inc. (2015). The options after the PROC MCMC statement refer, among others, to the number of MCMC simulation steps (NMC). The model has four parameters ($\beta_0, \beta_1, \beta_2, \beta_3$) for which a normal distribution each with mean zero and variance of 1,000 is assumed. As the variance is very high compared to the parameters, this prior reflects only diffuse information and is therefore called a diffuse or uninformative prior. The model is specified in a similar way as in PROC LOGISTIC. However, the PD as a function of the covariates and the distribution of defaults is explicitly given. Note that the parameters are scaled by factors 10 and 100, respectively, for numerical reasons.

```
/*Probit Model with MCMC*/
ODS LISTING CLOSE;
ODS LATEX;
ODS GRAPHICS ON / IMAGEFMT=PDF  IMAGENAME='bayesprobit';
PROC MCMC DATA = sample Statistics=ALL Diagnostics=ALL
SCALE=5 MINTUNE=5 NBI=1000 NMC=10000 THIN=2 PROPCOV=QUANEW SEED=12345;
PARMS beta0 beta1 beta2 beta3;
PRIOR beta0 beta1 beta2 beta3 ~ normal(0, var = 1000);
pd = PROBNORM(beta0 + beta1 / 10 * FICO_orig_time +
beta2 / 100 * LTV_orig_time + beta3 /100 * gdp_time);
MODEL default_time ~ binomial(1,pd);
RUN;
ODS GRAPHICS OFF;
```

The first table (Exhibit 12.2) shows the number of observations and the prior assumptions about the parameters.

The MCMC Procedure

Number of Observations Read	6333	Number of Observations Used	6333

Parameters				
Block	**Parameter**	**Sampling Method**	**Initial Value**	**Prior Distribution**
1	beta0	*N*–Metropolis	0	normal(0, var = 1000)
	beta1		0	normal(0, var = 1000)
	beta2		0	normal(0, var = 1000)
	beta3		0	normal(0, var = 1000)

Exhibit 12.2 MCMC Parameter Information for Probit Model

The next table (Exhibit 12.3) gives the summaries for the simulated posterior distributions of the parameters. Especially the means of the posteriors, the standard deviations, and some percentiles are of interest. Here, the means have the expected signs and their values are very similar to those of the frequentist model (after scaling is taken into account). Also note that the standard deviations are similar to the standard errors of PROC LOGISTIC.

The MCMC Procedure

Posterior Summaries						
Parameter	***N***	**Mean**	**Standard Deviation**	**Percentiles**		
				25	**50**	**75**
beta0	5000	−1.4239	0.4820	−1.7587	−1.4253	−1.0884
beta1	5000	−0.0226	0.00510	−0.0262	−0.0225	−0.0191
beta2	5000	1.2069	0.3746	0.9513	1.2011	1.4690
beta3	5000	−6.7046	1.6374	−7.7341	−6.6887	−5.5764

Exhibit 12.3 MCMC Parameter Summaries for Probit Model

In contrast to the frequentist approach, the Bayesian approach computed credibility intervals directly from the posterior distribution, which are shown in the next table (Exhibit 12.4). If one computes equal-tail and highest posterior density (HPD) 95 percent posterior credibility intervals as in the output, it can be seen that the value of zero is outside the interval for FICO and GDP, which is analogous to their significance in the frequentist approach.

Posterior Intervals					
Parameter	**Alpha**	**Equal-Tail Interval**		**HPD Interval**	
beta0	0.050	−2.3432	−0.4655	−2.3373	−0.4625
beta1	0.050	−0.0323	−0.0124	−0.0325	−0.0128
beta2	0.050	0.4878	1.9404	0.4827	1.9224
beta3	0.050	−9.9586	−3.6301	−9.8259	−3.5376

Exhibit 12.4 MCMC Procedure Output for Probit Model

Next, the correlation matrix of the parameters is given in Exhibit 12.5. Except for the constant, it shows only small correlations, which means that there is no issue with correlated samples.

Posterior Correlation Matrix				
Parameter	**beta0**	**beta1**	**beta2**	**beta3**
beta0	1.0000	−0.7772	−0.7161	−0.0939
beta1	−0.7772	1.0000	0.1248	0.1045
beta2	−0.7161	0.1248	1.0000	−0.0138
beta3	−0.0939	0.1045	−0.0138	1.0000

Exhibit 12.5 MCMC Procedure Output for Probit Model

In order to check convergence of the simulation, autocorrelations are important and are shown in the next table (Exhibit 12.6). High correlations between long lags indicate poor mixing. As can be seen, the autocorrelation is high for small lags and decreases sharply for longer lags, which is the desired property.

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.7380	0.1887	0.0316	–0.0246
beta1	0.7528	0.2400	0.0749	–0.0196
beta2	0.7235	0.1796	0.0012	–0.0512
beta3	0.7639	0.2724	0.0555	0.0005

Exhibit 12.6 MCMC Procedure Output for Probit Model

For a closer check of convergence, some diagnostics are computed, named Geweke, Raftery-Lewis, and Heidelberger-Welch diagnostics, with accompanying tests as reported in the next table (Exhibit 12.7). The Geweke diagnostics test whether the mean estimates have converged by comparing the means from the early and latter part of the Markov chain. It comes with a two-sided test based on a z-score statistic. Large absolute z values indicate rejection. In our example the convergence is obviously satisfactory.

Geweke Diagnostics		
Parameter	z	Pr > |z|
beta0	0.1421	0.8870
beta1	–1.1085	0.2676
beta2	0.9280	0.3534
beta3	0.3963	0.6919

Exhibit 12.7 MCMC Diagnostics

The Raftery-Lewis diagnostics evaluate the accuracy of the estimated (desired) percentiles by reporting the number of samples needed to reach the desired accuracy of the percentiles. Failure could indicate that a longer Markov chain is needed. If the total samples needed are fewer than the Markov chain sample, this indicates rejection. Here, the total sample is larger and therefore the accuracy should be satisfactory.

The Heidelberger-Welch diagnostics are divided into a stationarity test and a half-width test. The former tests whether the Markov chain is a covariance (or weakly) stationary process. Failure could indicate that a longer Markov chain is needed. It is a one-sided test based on a Cramer–von Mises statistic. Small p-values indicate rejection. In Exhibit 12.8 the p-values are large (with the exception for β_0) and therefore, the chain can be considered as covariance stationary. SAS automatically returns the information whether the test is passed or not. The half-width test reports whether the sample size is adequate to meet the required accuracy for the mean estimate. Failure could indicate that a longer Markov chain is needed. If a relative half-width statistic is greater than a predetermined accuracy measure, this indicates rejection. In Exhibit 12.9, the tests signal that the length of the Markov chain should be sufficient.

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/–0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
beta0	11	11486	3746	3.0662
beta1	18	21054	3746	5.6204
beta2	11	11311	3746	3.0195
beta3	11	12287	3746	3.2800

Exhibit 12.8 MCMC Diagnostics

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
beta0	0.0750	0.7216	Passed	0	0.0378	–1.4239	–0.0265	Passed
beta1	0.0803	0.6900	Passed	0	0.000480	–0.0226	–0.0213	Passed
beta2	0.1276	0.4656	Passed	0	0.0233	1.2069	0.0193	Passed
beta3	0.0738	0.7286	Passed	0	0.1212	–6.7046	–0.0181	Passed

Exhibit 12.9 MCMC Diagnostics

Finally, the effective sample size is computed (see Exhibit 12.10). It relates to autocorrelation and measures the mixing of the Markov chain. A large discrepancy between the effective sample size and the simulation sample size indicates poor mixing.

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
beta0	776.4	6.4399	0.1553
beta1	565.6	8.8397	0.1131
beta2	899.4	5.5594	0.1799
beta3	689.5	7.2520	0.1379

Exhibit 12.10 MCMC Diagnostics

In addition to the statistics, PROC MCMC also shows some diagnostic plots (see Exhibits 12.11 through 12.14). For each parameter, the simulation history is shown in the upper plot. If it looks like a “white-noise” process without drift (as is the case for our four parameters), it indicates good convergence. In the lower left, the autocorrelations are shown for lags up to 50. As in the table, the autocorrelations are high for short lags and close to zero for longer lags, which is the desired property. Finally, in the lower right the simulated posterior distribution is given for each parameter from which the point estimates (i.e., the means) and the credibility intervals are computed.

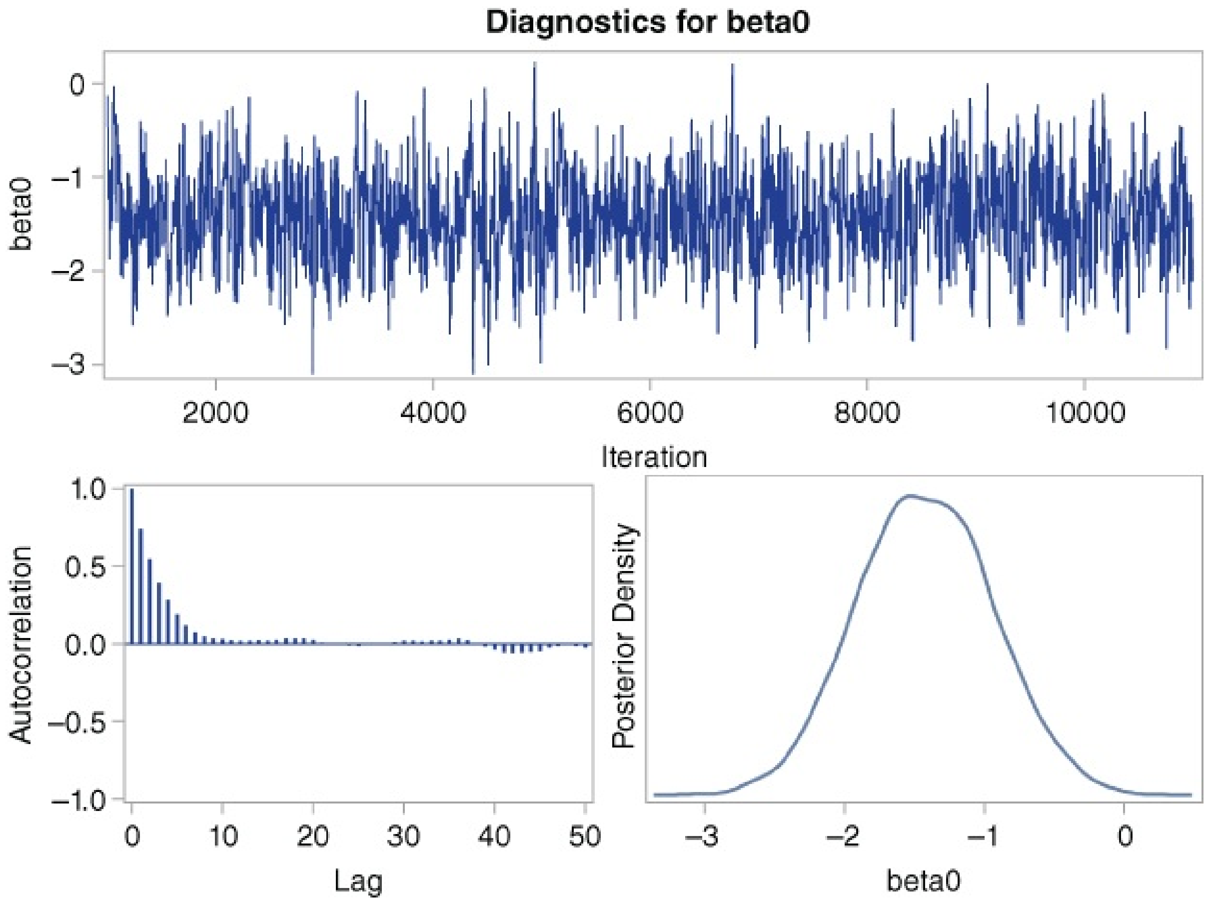

Exhibit 12.11 Diagnostic Plots

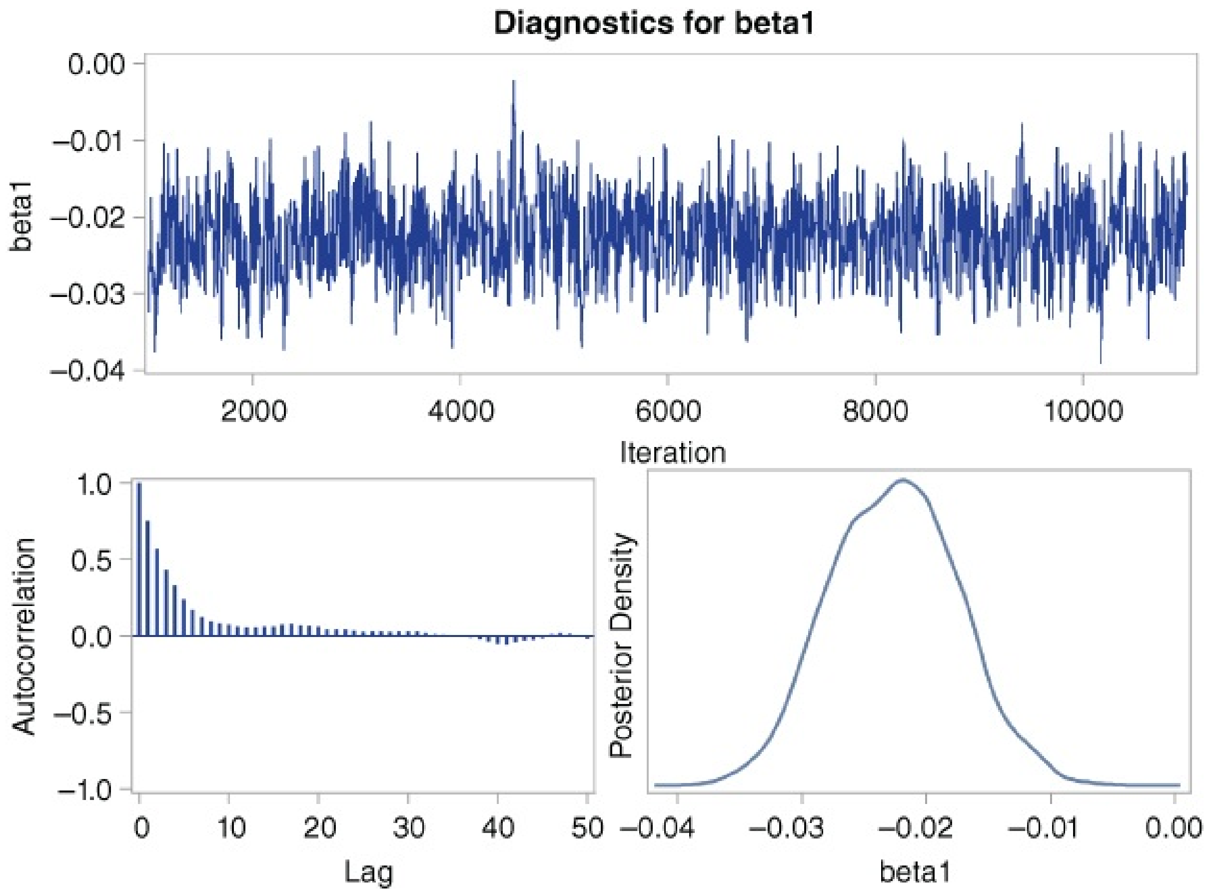

Exhibit 12.12 Diagnostic Plots

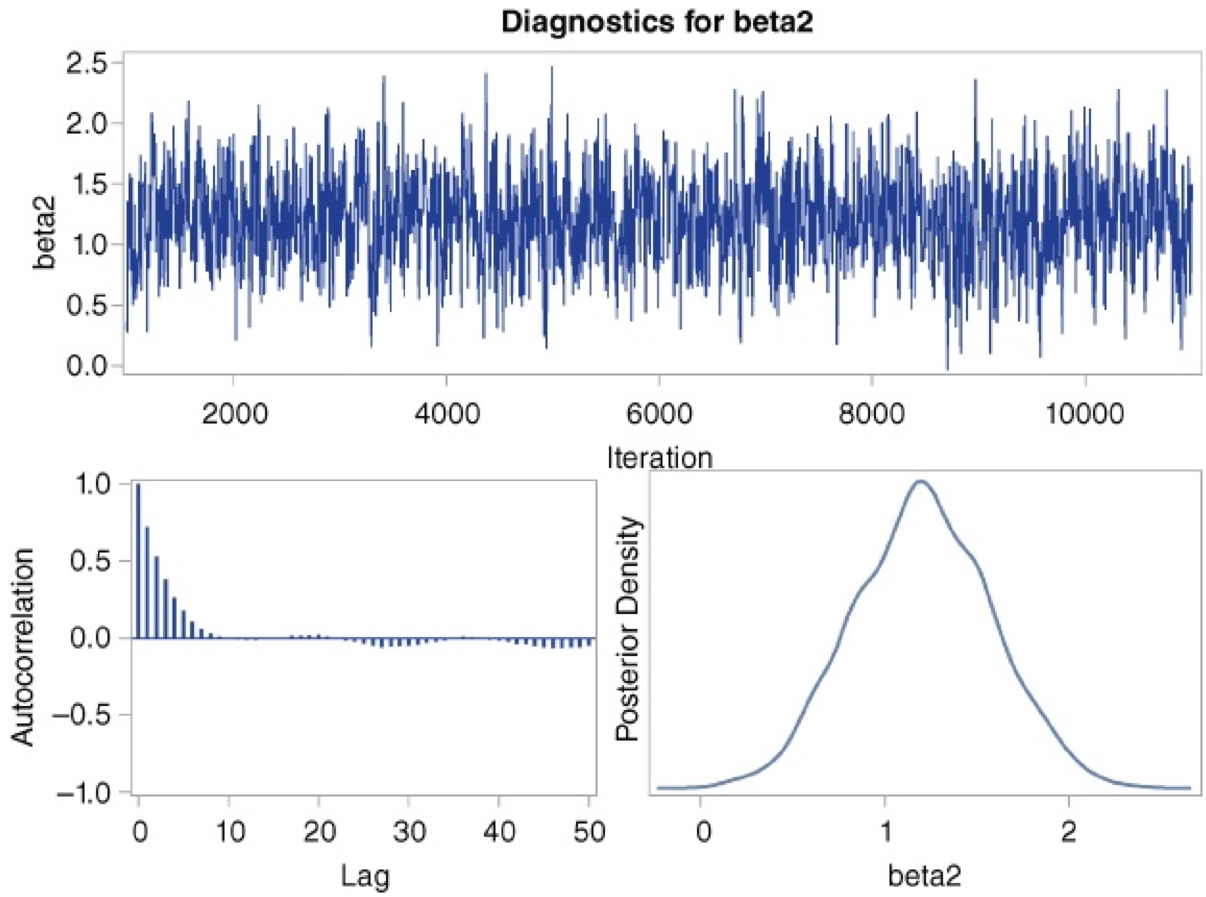

Exhibit 12.13 Diagnostic Plots

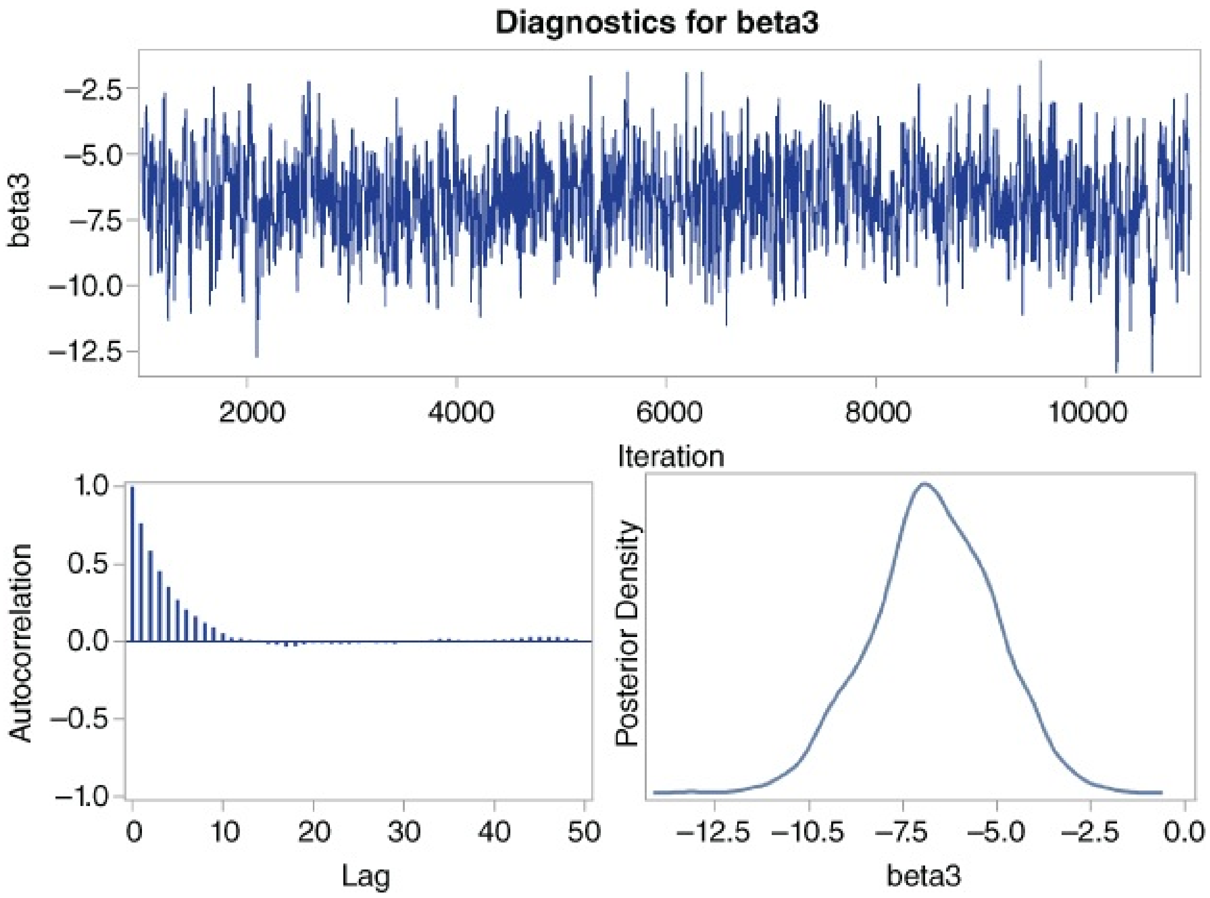

Exhibit 12.14 Diagnostic Plots

In this example, the MCMC results lead to similar conclusions about the parameters as in the classical approach. So, what is the difference between the two approaches, or, in other words, when is it really helpful to use Bayesian statistics? One important instance is the availability of useful prior information. Suppose there is an expert on mortgages who has the strong opinion that it is very important to consider loan-to-value (LTV) information and he or she insists that the coefficient should be a lot higher than the output of the classical approach suggests. The expert also remembers an earlier study where the coefficient was "around 3 or so." This can now be included as prior information. Instead of using a diffuse prior with mean 0 and variance of 1,000, one may now specify a normal prior with mean 3 and variance 0.5 as shown in the following code.

```
/*Probit Model with MCMC*/
ODS LISTING CLOSE;
ODS LATEX;
ODS GRAPHICS ON / IMAGEFMT=PDF  IMAGENAME='bayes2probit';
PROC MCMC DATA = sample STATISTICS=ALL DIAGNOSTICS=ALL
SCALE=5 MINTUNE=5 NBI=1000 NMC=10000 THIN=2 PROPCOV=QUANEW SEED=12345;
PARMS beta0 beta1 beta2 beta3;
PRIOR beta0 beta1 beta3 ~ normal(0, var = 1000);
PRIOR beta2 ~ normal (3, var = 0.5);
pd = PROBNORM(beta0 + beta1 / 10 * FICO_orig_time +
beta2 / 100 * LTV_orig_time + beta3 /100 * gdp_time);
MODEL default_time ~ BINOMIAL(1,pd);
RUN;
```

```
ODS GRAPHICS OFF;
```

As can be seen from the output (Exhibits 12.15 through 12.19), the mean coefficient for LTV has now shifted upward as a result of mixing data information with the expert's prior opinions (and the constant has moved in the opposite direction). The means and intervals for the other parameters are similar to what they were before and the simulations have converged well. This demonstrates the importance and usefulness of expert information, but also that one should include expert information wisely as it may impose restrictions on the parameters. The reader is encouraged to use other specifications of the priors and other options for PROC MCMC to become more familiar with its features and outputs. For example, an informative prior could be used with low variance and other mean values in order to gauge how the prior information changes the posterior.

The MCMC Procedure

Number of Observations Read	6,333	Number of Observations Used	6,333

Parameters				
Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	beta0	*N*–Metropolis	0	normal(0, var = 1000)
	beta1		0	normal(0, var = 1000)
	beta2		3.0000	normal(3, var = 0.5)
	beta3		0	normal(0, var = 1000)

The MCMC Procedure

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25	50	75
beta0	5000	−1.7798	0.4526	−2.0959	−1.7869	−1.4780
beta1	5000	−0.0216	0.00519	−0.0252	−0.0215	−0.0181
beta2	5000	1.5697	0.3242	1.3546	1.5684	1.7869
beta3	5000	−6.7201	1.7121	−7.9134	−6.6850	−5.5423

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
beta0	0.050	−2.6565	−0.8580	−2.7087	−0.9374
beta1	0.050	−0.0319	−0.0115	−0.0319	−0.0115
beta2	0.050	0.9207	2.1885	0.9647	2.2171
beta3	0.050	−9.9649	−3.3740	−9.8994	−3.3423

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.7300	0.2033	0.0151	−0.0104
beta1	0.7443	0.2444	0.0402	−0.0341
beta2	0.7418	0.2175	0.0294	−0.0201
beta3	0.7416	0.2036	0.0624	−0.0443

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
beta0	0.0766	0.7118	Passed	0	0.0290	−1.7798	−0.0163	Passed
beta1	0.1445	0.4068	Passed	0	0.000399	−0.0216	−0.0185	Passed
beta2	0.2136	0.2426	Passed	0	0.0231	1.5697	0.0147	Passed
beta3	0.1492	0.3920	Passed	0	0.1301	−6.7201	−0.0194	Passed

Exhibit 12.15 Summary Statistics

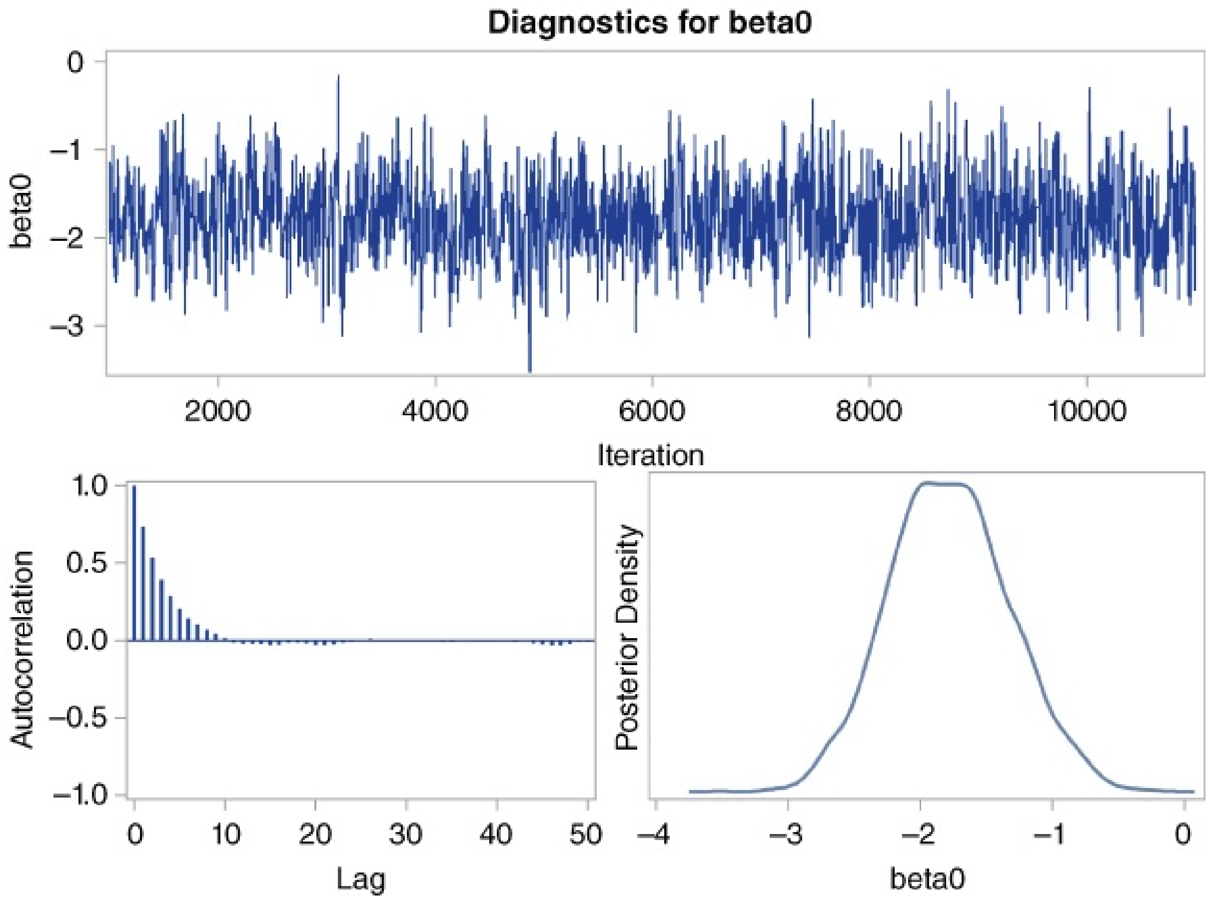

Exhibit 12.16 Diagnostic Plots

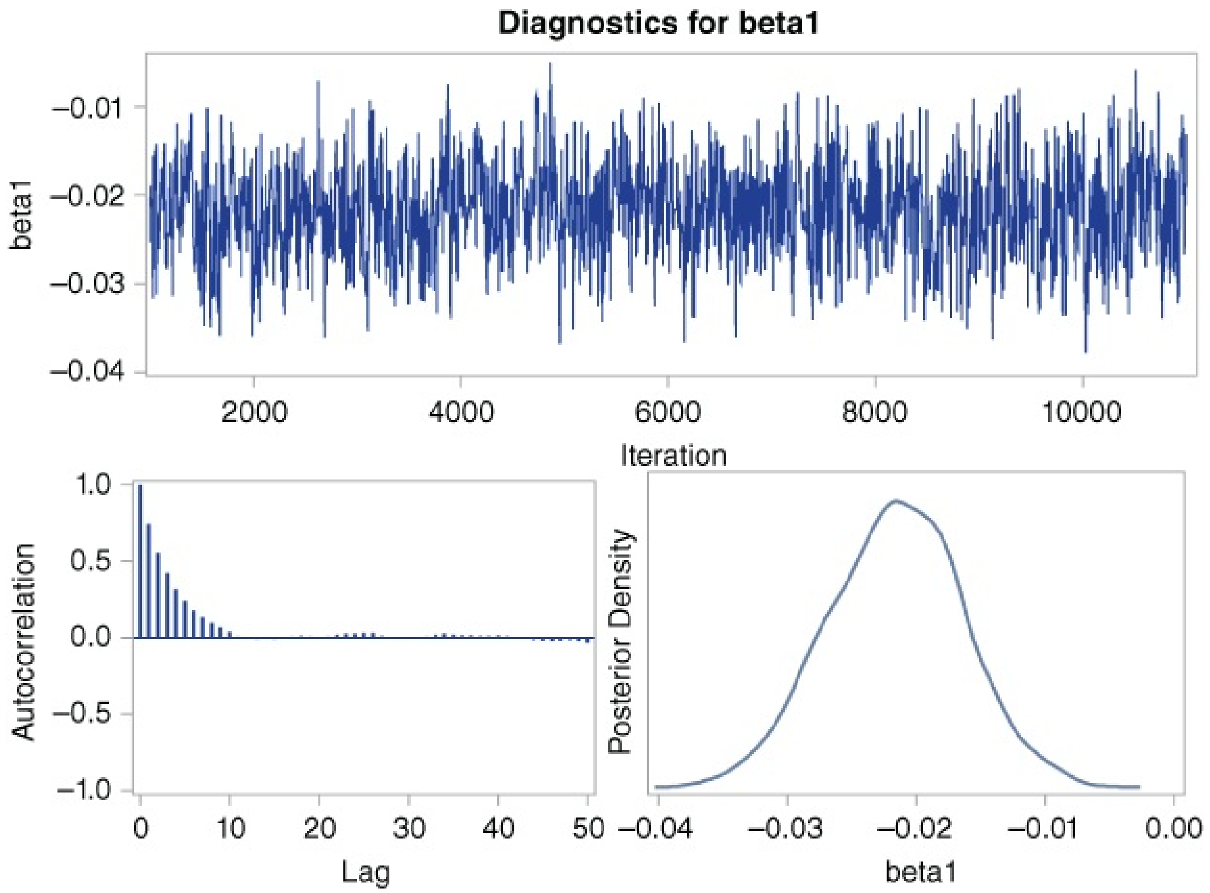

Exhibit 12.17 Diagnostic Plots

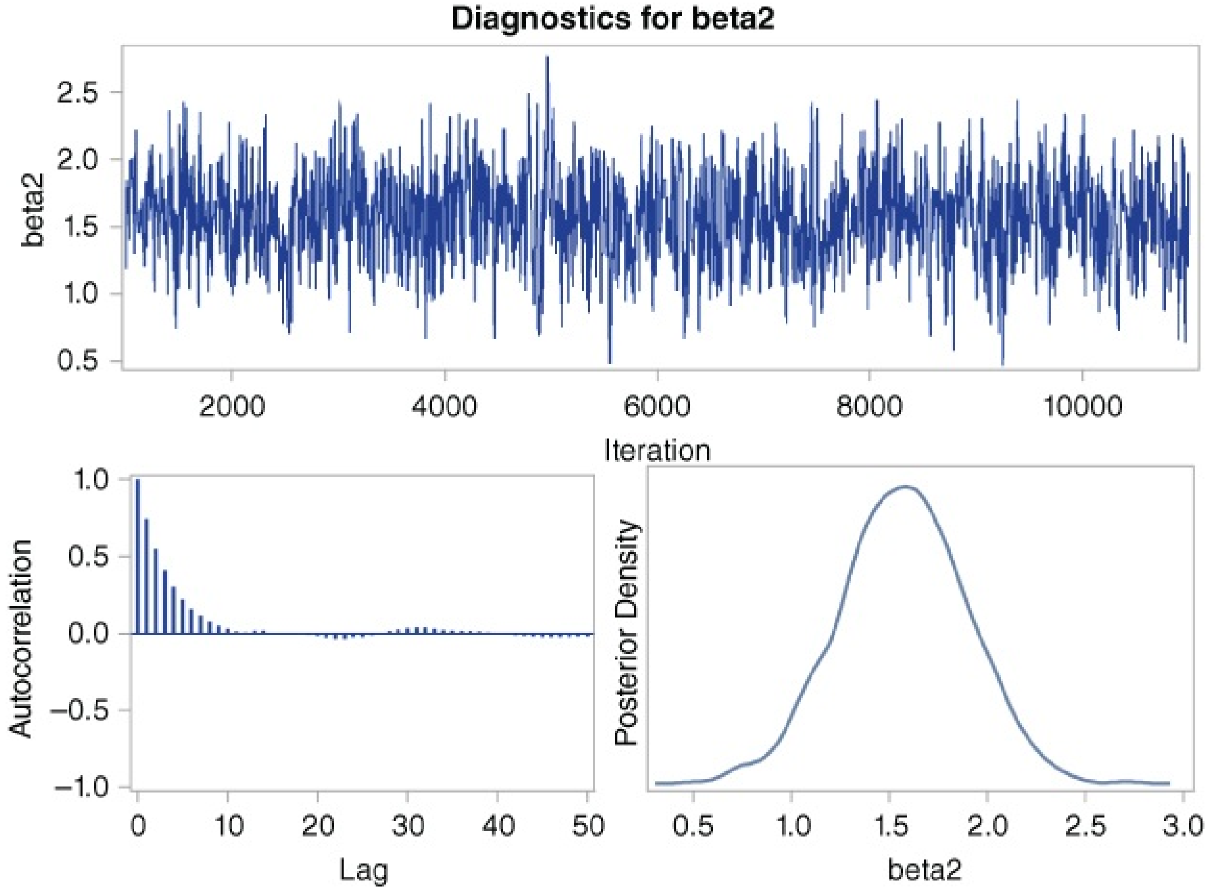

Exhibit 12.18 Diagnostic Plots

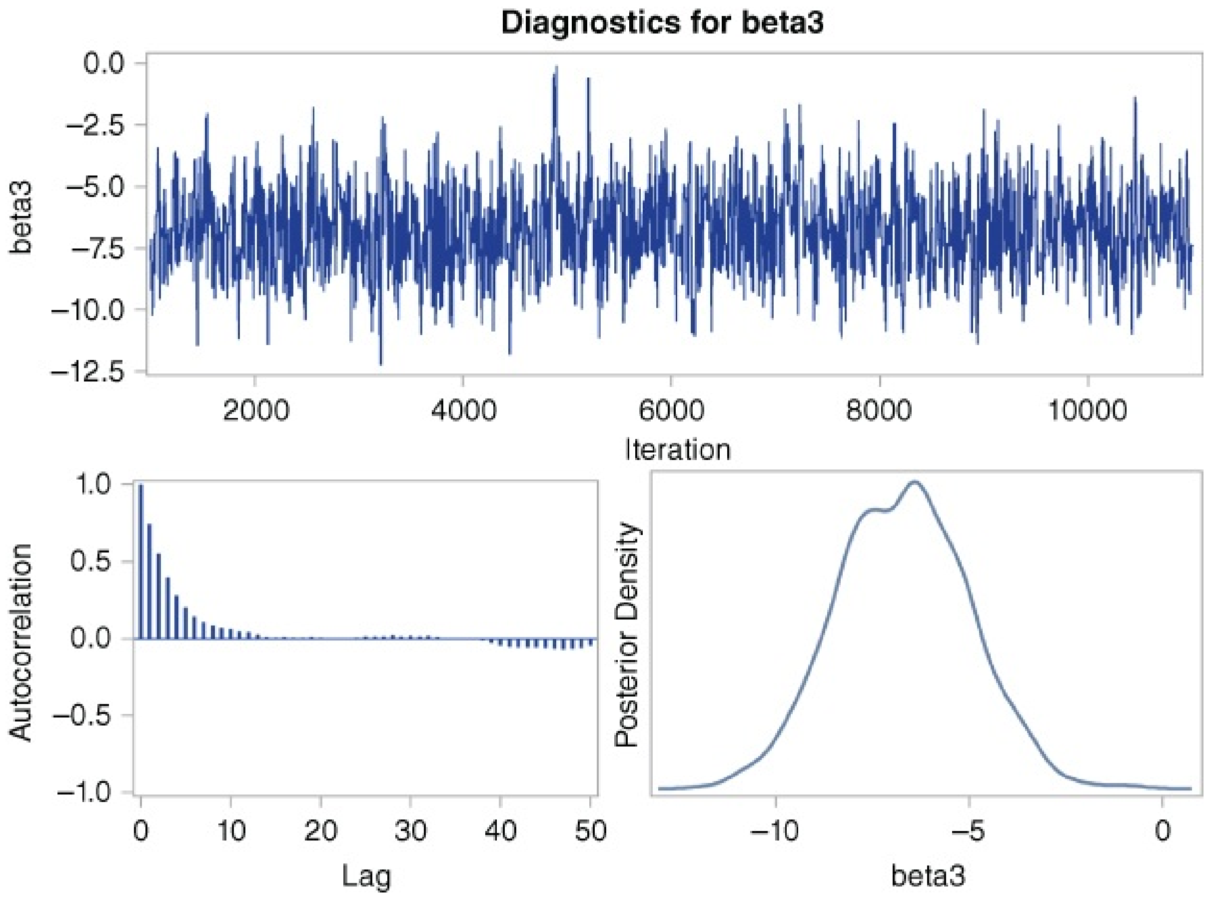

Exhibit 12.19 Diagnostic Plots

Survival Analysis

Next, we show how a simple survival model can be analyzed with Bayesian statistics. (See Exhibits 12.20 through 12.23.) As in the earlier section, we specify a Cox proportional hazards model. Under the Bayesian approach, we use a diffuse prior. SAS offers the possibility to estimate a Bayesian CPH model with PROC PHREG as a very convenient alternative to PROC MCMC; see SAS Institute Inc. (2015). The code is analogous to the earlier PHREG and MCMC code. The command line starting with BAYES requires a Bayesian model to be estimated.

The PHREG Procedure

Bayesian Analysis

Model Information	
Data Set	WORK.SAMPLE
Dependent Variable	time1
Dependent Variable	time2
Censoring Variable	default_time
Censoring Value(s)	0
Model	Cox
Ties Handling	BRESLOW
Sampling Algorithm	ARMS
Burn-In Size	2000
MC Sample Size	5000
Thinning	1

Number of Observations Read	6333	Number of Observations Used	6333

Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	95% Confidence Limits	
FICO_orig_time	1	−0.00468	0.00119	−0.00701	−0.00236
LTV_orig_time	1	0.0270	0.00838	0.0106	0.0435
gdp_time	1	−0.0899	0.0419	−0.1720	−0.00776

Independent Normal Prior for Regression Coefficients		
Parameter	Mean	Precision
FICO_orig_time	0	1E−6
LTV_orig_time	0	1E−6
gdp_time	0	1E−6

Fit Statistics	
DIC (smaller is better)	1463.113
pD (Effective Number of Parameters)	3.018

The PHREG Procedure

Bayesian Analysis

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
FICO_orig_time	5000	−0.00469	0.00118	−0.00547	−0.00469	−0.00390
LTV_orig_time	5000	0.0267	0.00837	0.0212	0.0268	0.0323
gdp_time	5000	−0.0892	0.0425	−0.1188	−0.0899	−0.0594

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
FICO_orig_time	0.050	−0.00701	−0.00239	−0.00697	−0.00236
LTV_orig_time	0.050	0.0101	0.0428	0.00943	0.0420
gdp_time	0.050	−0.1718	−0.00558	−0.1747	−0.00991

Posterior Correlation Matrix			
Parameter	FICO_orig_time	LTV_orig_time	gdp_time
FICO_orig_time	1.0000	0.1745	0.0131
LTV_orig_time	0.1745	1.0000	−.0306
gdp_time	0.0131	−.0306	1.0000

The PHREG Procedure

Bayesian Analysis

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
FICO_orig_time	0.0251	−0.0007	0.0022	0.0114
LTV_orig_time	0.0349	0.0059	0.0118	−0.0038
gdp_time	−0.0039	−0.0014	−0.0087	−0.0114

Geweke Diagnostics		
Parameter	z	Pr > |z|
FICO_orig_time	−0.6374	0.5239
LTV_orig_time	−3.4405	0.0006
gdp_time	0.3633	0.7164

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/−0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
FICO_orig_time	2	3742	3746	0.9989
LTV_orig_time	2	3681	3746	0.9826
gdp_time	2	3655	3746	0.9757

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
FICO_orig_time	0.0956	0.6067	Passed	0	0.000032	−0.00469	−0.00691	Passed
LTV_orig_time	0.2301	0.2161	Passed	0	0.000320	0.0267	0.0120	Passed
gdp_time	0.1904	0.2868	Passed	0	0.00123	−0.0892	−0.0138	Passed

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
FICO_orig_time	4760.9	1.0502	0.9522
LTV_orig_time	4479.9	1.1161	0.8960
gdp_time	5000.0	1.0000	1.0000

Exhibit 12.20 Summary Statistics

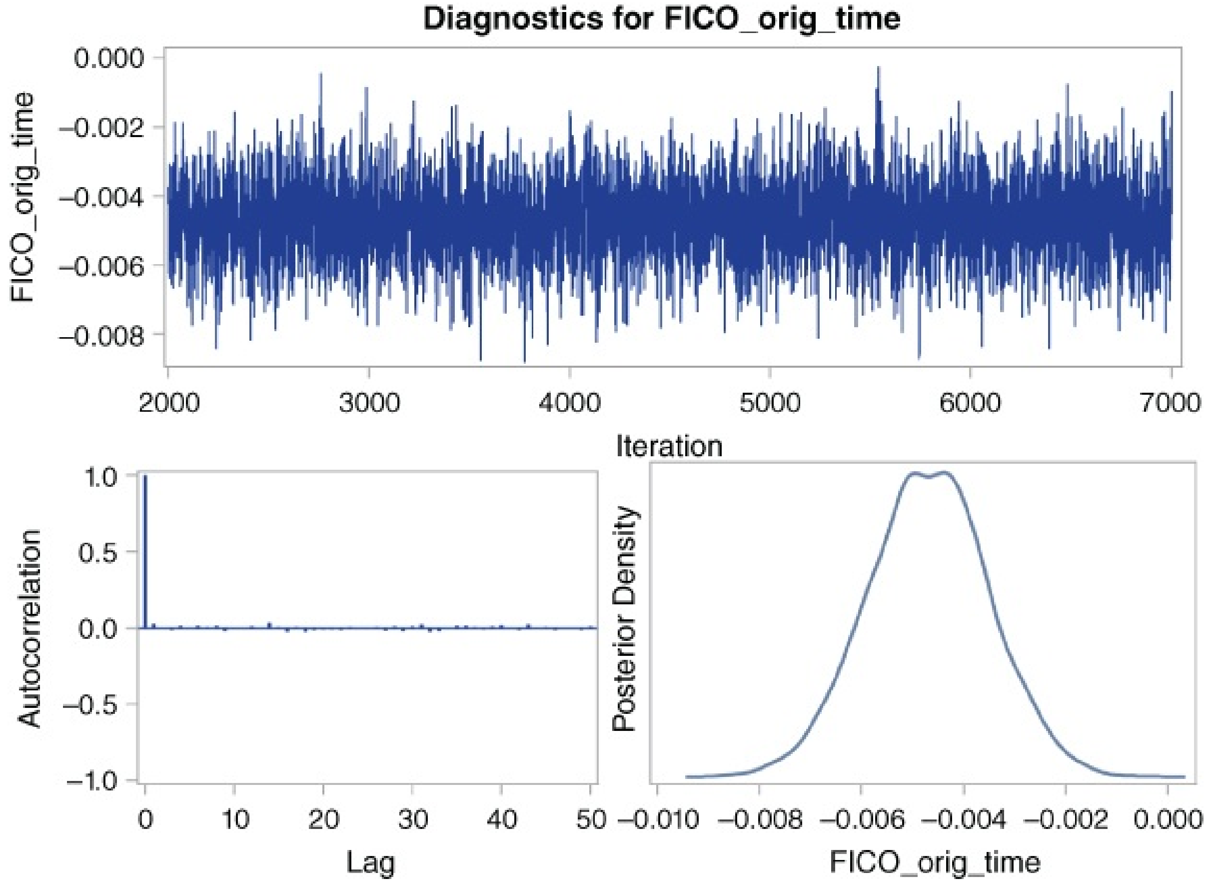

Exhibit 12.21 Diagnostic Plots

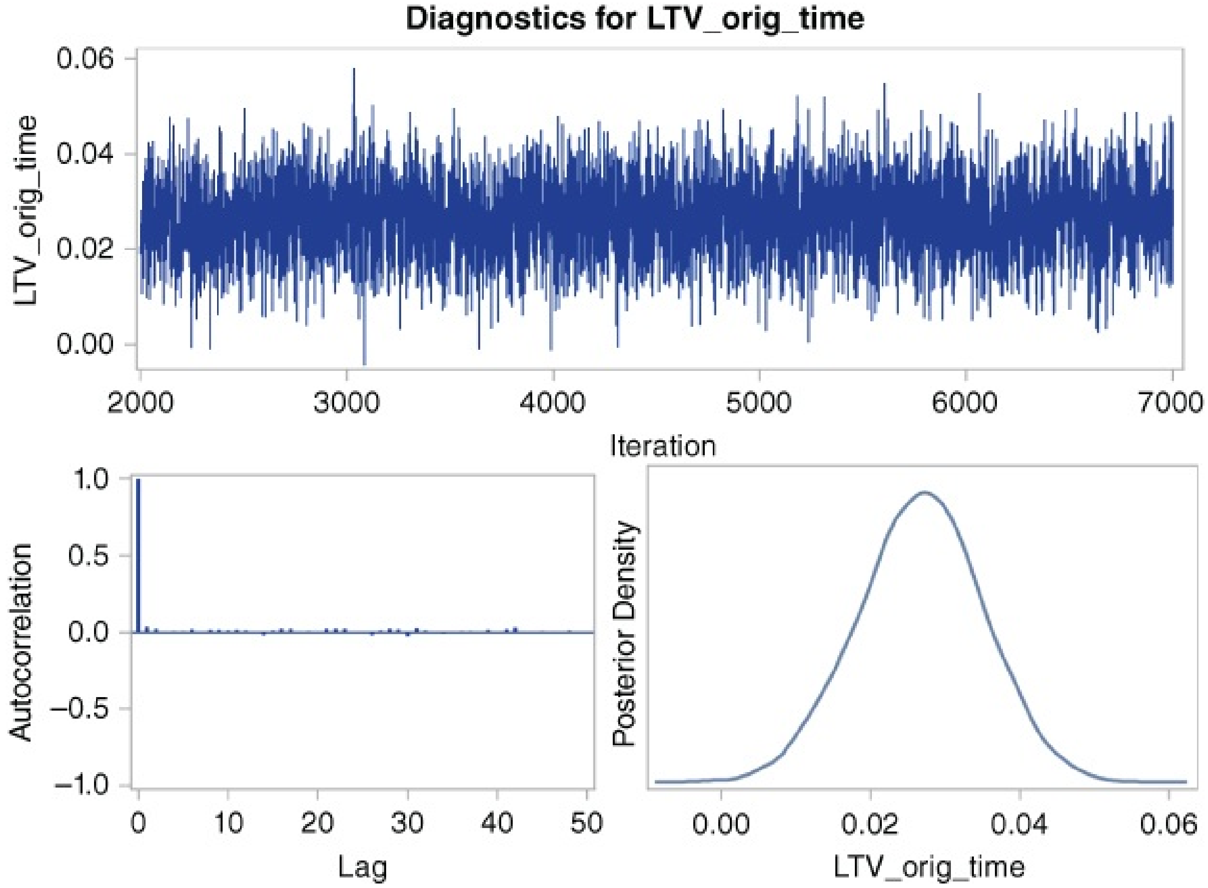

Exhibit 12.22 Diagnostic Plots

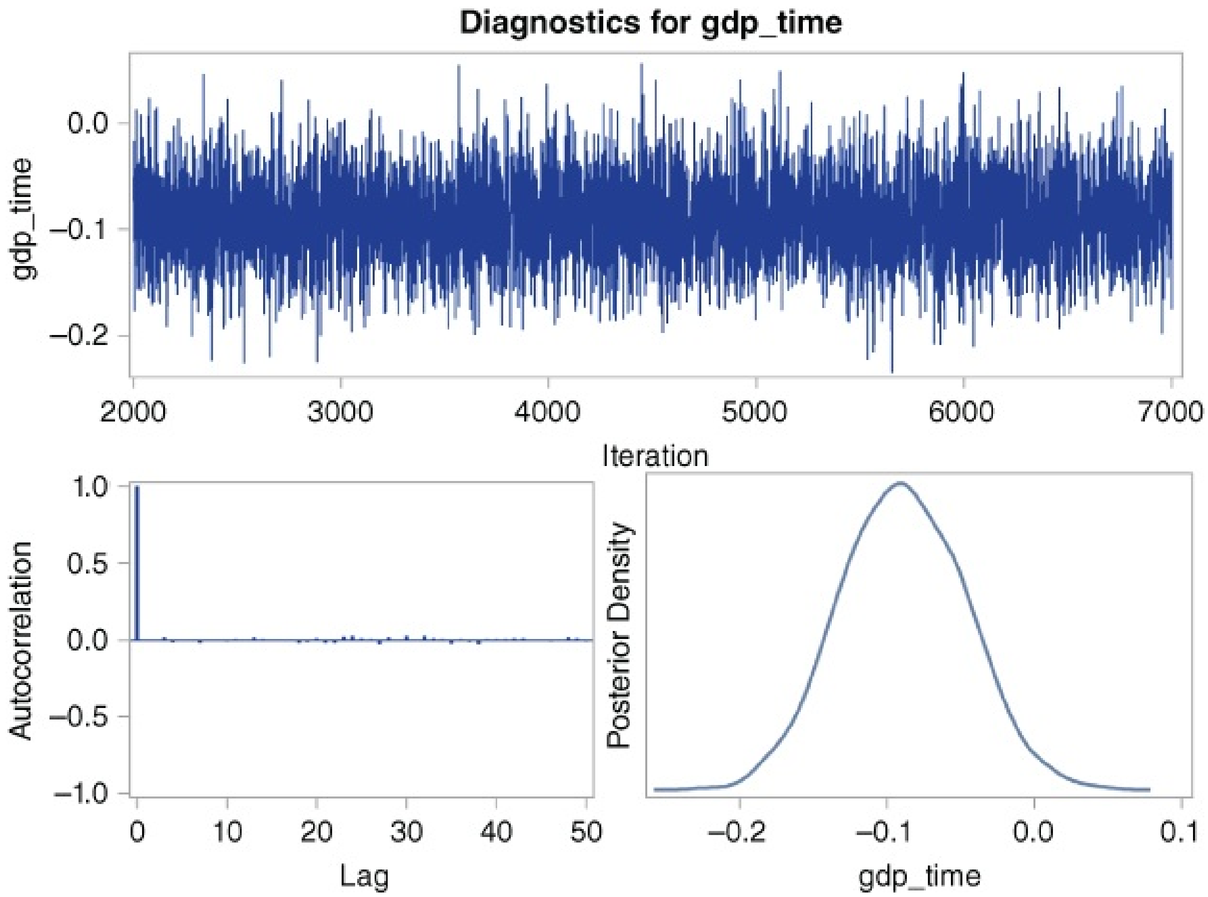

Exhibit 12.23 Diagnostic Plots

```
/*Survival Model using PROC PHREG with Bayesian approach*/
ODS GRAPHICS ON / IMAGEFMT=PDF IMAGENAME='bayessurvival';
PROC PHREG DATA=sample PLOTS=CUMHAZ PLOTS(OVERLAY)=SURVIVAL;
MODEL (Time1,Time2)*default_time(0)=FICO_orig_time
    LTV_orig_time gdp_time;
BAYES SEED=1 COEFFPRIOR=NORMAL STATISTICS=ALL NMC=5000
DIAGNOSTICS=ALL PLOTS=ALL;
ID ID Time default_time;
STORE out=score_phreg;
RUN;
ODS GRAPHICS OFF;
```

Similarly to PROC MCMC, the output starts with some general model information and returns the results of the ML estimates as comparison. There is no need to run a separate ML model. The Bayesian output is actually very similar to the MCMC output and shows the same information. Due to the diffuse prior and availability of rich data, the posterior means for the parameters are very close to those from the ML estimation and the model has converged well. As in the PD estimation, Bayesian statistics are particularly useful if only sparse sample data were available. This will be shown in a later section.

CORRELATION ESTIMATION WITH BAYESIAN STATISTICS

Asset correlations can also be estimated using the Bayesian approach, as will be shown in this section. We use the same simple aggregated model and the same data as in the chapter on correlations. So, we regress $\Phi^{-1}(dr_t)$ on a constant. The extension to covariates is left to the reader for training purposes. In other words, we have two parameters, namely a constant coded as β_0 and the residual standard variance. These can be transformed into the PD and the asset correlation as discussed earlier. For both, we use noninformative, diffuse priors. β_0 is assumed to be normally distributed with mean 0 and variance 10,000 whereas σ^2 is assumed uniformly distributed between 0 and 1. Then, $\Phi^{-1}(dr_t)$ is modeled as $N(\beta_0, \sigma^2)$.

```
/*Probit-Linear Regression without Covariates with MCMC*/
ODS LISTING CLOSE;
ODS LATEX;
ODS GRAPHICS ON / IMAGEFMT=PDF  IMAGENAME='bayescorr';
PROC MCMC DATA = data.tmp2 Statistics=ALL Diagnostics=ALL
SCALE=5 MINTUNE=5 NBI=1000 NMC=10000 THIN=2 PROPCOV=QUANEW SEED=12345;
PARMS beta0 0 sigma2 0.1;
PRIOR beta0  ~ normal(0, var = 10000);
PRIOR sigma2 ~ Uniform(0,1);
mu = beta0;
MODEL probit_dr ~ n(mu, var = sigma2);
RUN;
ODS GRAPHICS OFF;
```

The output (see Exhibits 12.24 through 12.26) shows good convergence and low autocorrelation for both coefficients. The means of the posterior distributions are very similar to those from the classical approach, which can be attributed to good sample information and the diffuse priors. Now consider that an expert strongly argues that the variance or the correlation should be a lot higher. As an example, we use a lognormal prior for σ with mean 0.1 and variance 0.1 in the next program code. It can be seen that the mean value for σ^2 is now more than twice as high. If this value is transformed into the asset correlation, one obtains approximately $\rho = \frac{\sigma^2}{1+\sigma^2} = 0.17$. In other words, the expert's prior information is linked with the data and leads to more conservative estimates (see Exhibits 12.27 through 12.29).

The MCMC Procedure

Number of Observations Read	60	Number of Observations Used	60

Parameters				
Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	beta0	*N*−Metropolis	0	normal(0, var = 10000)
	sigma2		0.1000	uniform(0,1)

The MCMC Procedure

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25	50	75
beta0	5000	−2.2772	0.0358	−2.3013	−2.2761	−2.2528
sigma2	5000	0.0793	0.0154	0.0686	0.0773	0.0880

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
beta0	0.050	−2.3484	−2.2084	−2.3533	−2.2143
sigma2	0.050	0.0545	0.1149	0.0523	0.1105

Posterior Correlation Matrix		
Parameter	beta0	sigma2
beta0	1.0000	−0.0099
sigma2	−0.0099	1.0000

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.5959	0.0601	−0.0180	0.0055
sigma2	0.6112	0.0809	−0.0215	−0.0144

Geweke Diagnostics		
Parameter	z	Pr > |z|
beta0	0.5783	0.5631
sigma2	0.2690	0.7880

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/−0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
beta0	10	10534	3746	2.8121
sigma2	11	12172	3746	3.2493

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
beta0	0.1014	0.5782	Passed	0	0.00179	−2.2772	−0.00079	Passed
sigma2	0.1666	0.3426	Passed	0	0.000798	0.0793	0.0101	Passed

Effective Sample Sizes			
Parameter	ESS	Autocorrelation Time	Efficiency
beta0	1321.6	3.7834	0.2643
sigma2	1249.2	4.0026	0.2498

Exhibit 12.24 Summary Statistics

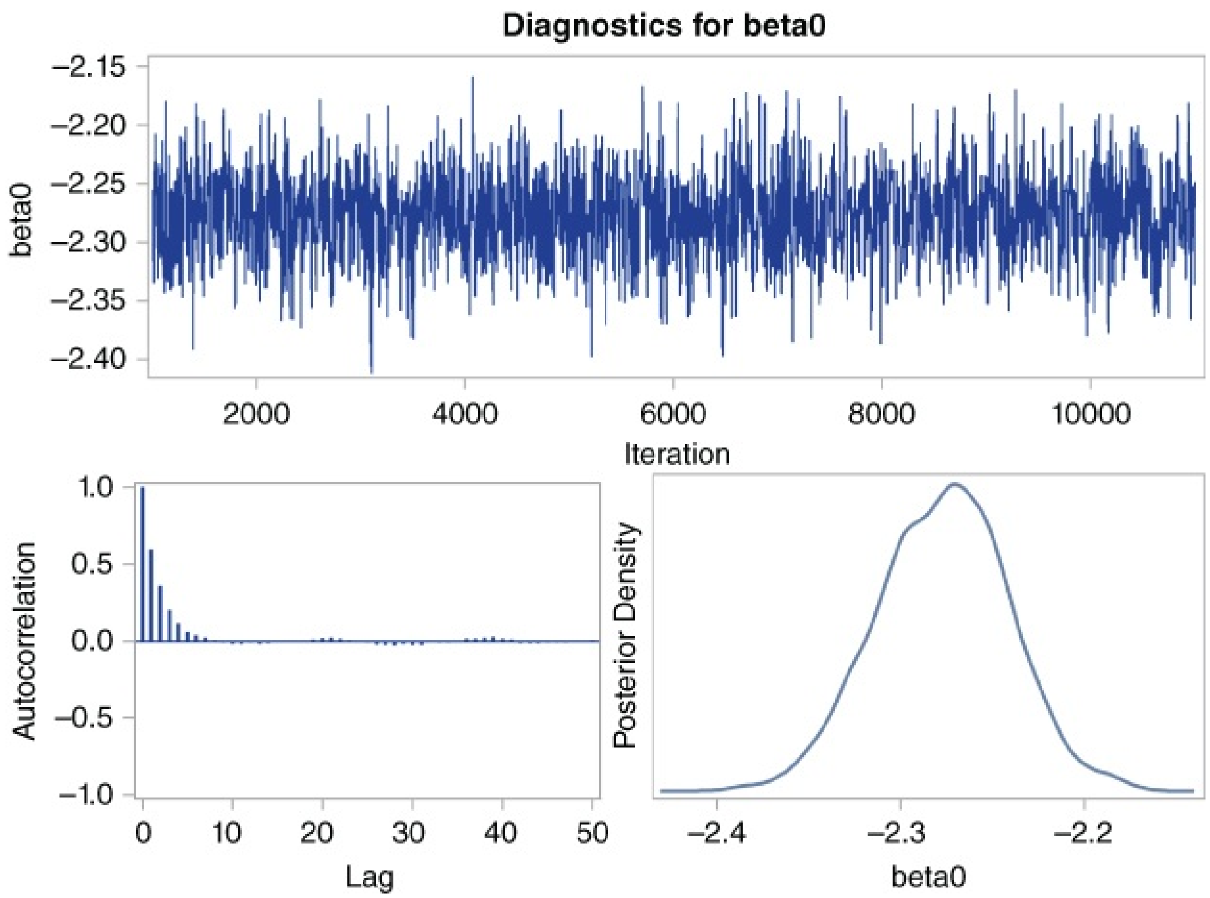

Exhibit 12.25 Diagnostic Plots

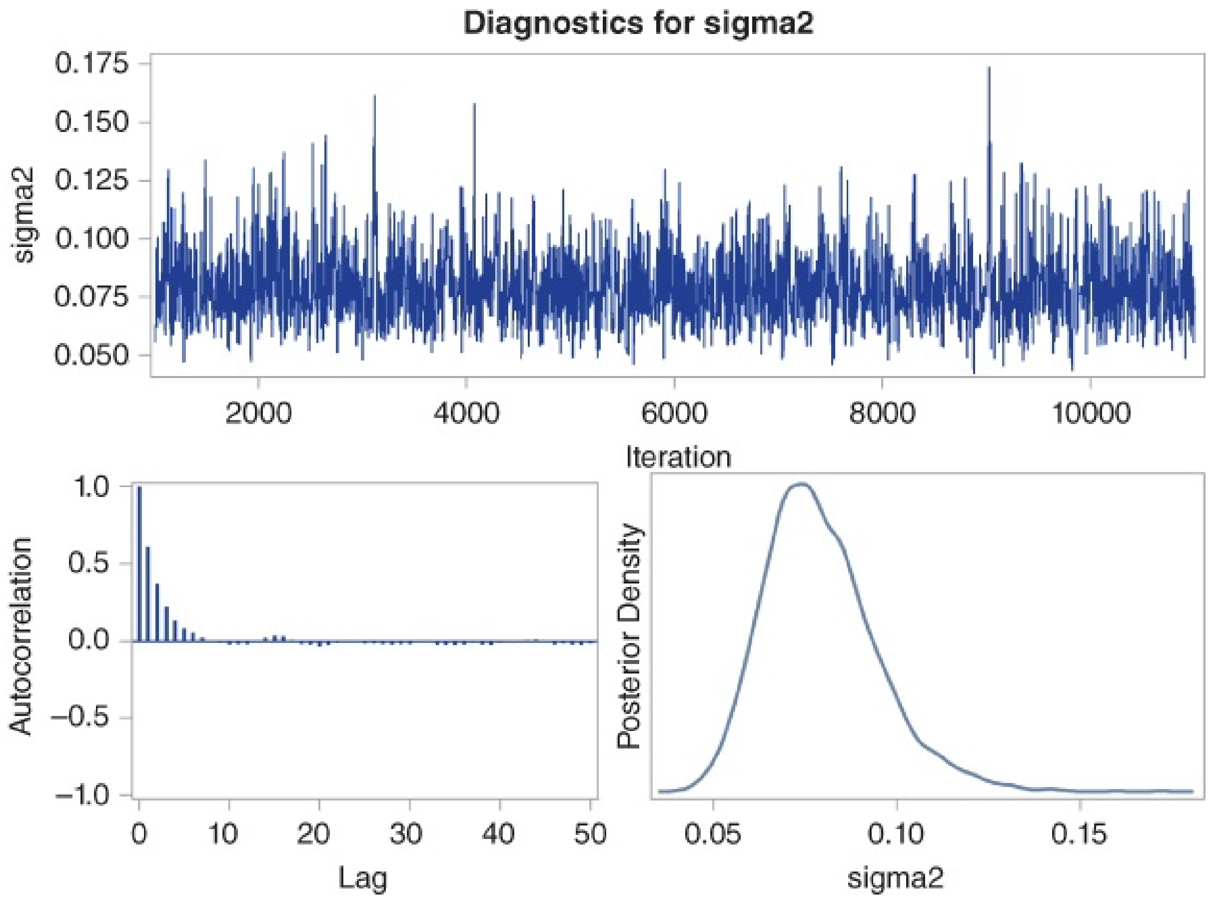

Exhibit 12.26 Diagnostic Plots

The MCMC Procedure

Number of Observations Read	60	Number of Observations Used	60

Parameters				
Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	beta0	*N*–Metropolis	0	normal(0, var = 10000)
	sigma2		0.1000	lognormal(0.1, var = 0.1)

The MCMC Procedure

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25	50	75
beta0	5000	−2.2784	0.0574	−2.3164	−2.2792	−2.2409
sigma2	5000	0.1919	0.0415	0.1620	0.1873	0.2164

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
beta0	0.050	−2.3941	−2.1647	−2.3994	−2.1720
sigma2	0.050	0.1253	0.2844	0.1223	0.2803

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.6043	0.0785	0.0281	−0.0017
sigma2	0.5857	0.1188	−0.0234	−0.0319

Geweke Diagnostics		
Parameter	z	Pr > |z|
beta0	0.3971	0.6913
sigma2	1.4173	0.1564

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/–0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
beta0	11	11744	3746	3.1351
sigma2	10	10668	3746	2.8478

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
beta0	0.1595	0.3618	Passed	0	0.00341	−2.2784	−0.00150	Passed
sigma2	0.2558	0.1811	Passed	0	0.00193	0.1919	0.0101	Passed

Exhibit 12.27 Summary Statistics

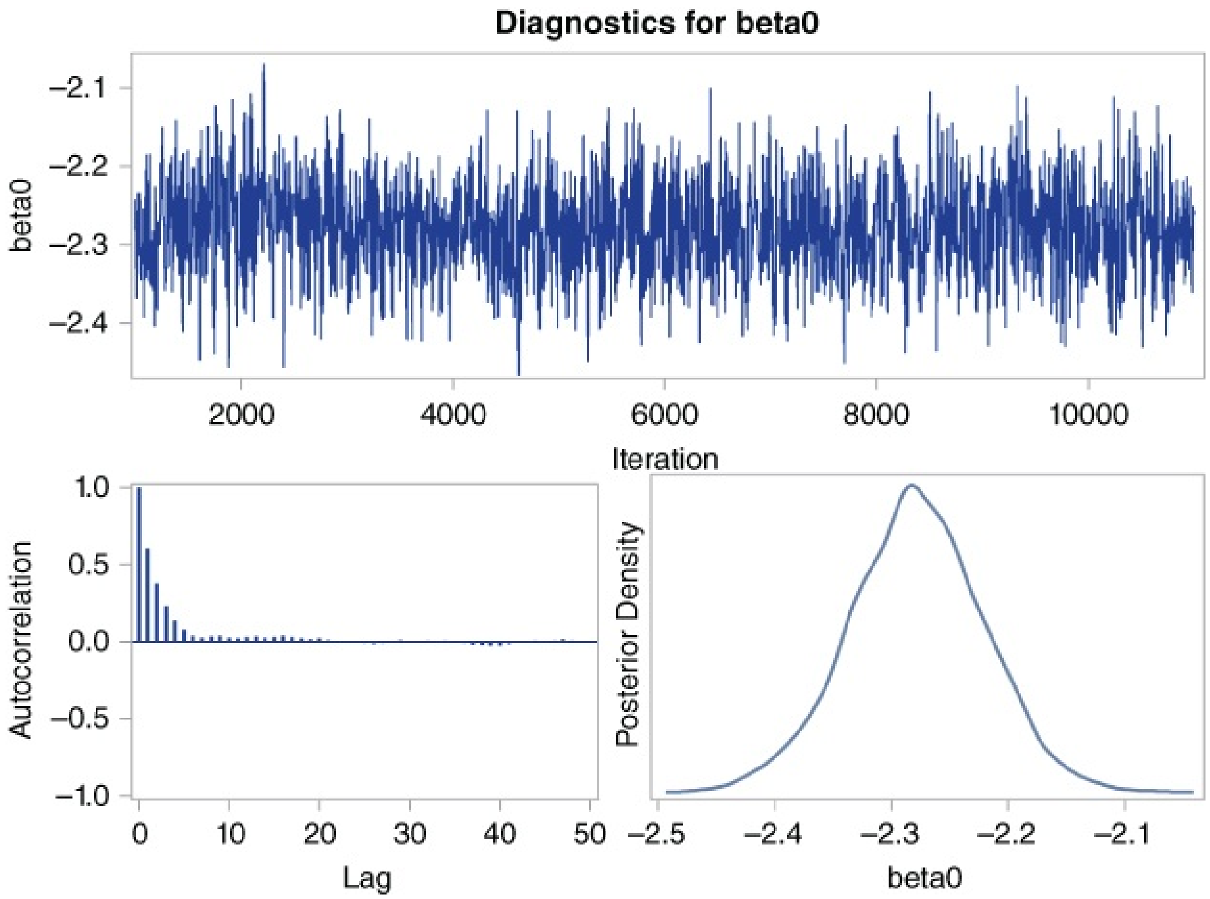

Exhibit 12.28 Diagnostic Plots

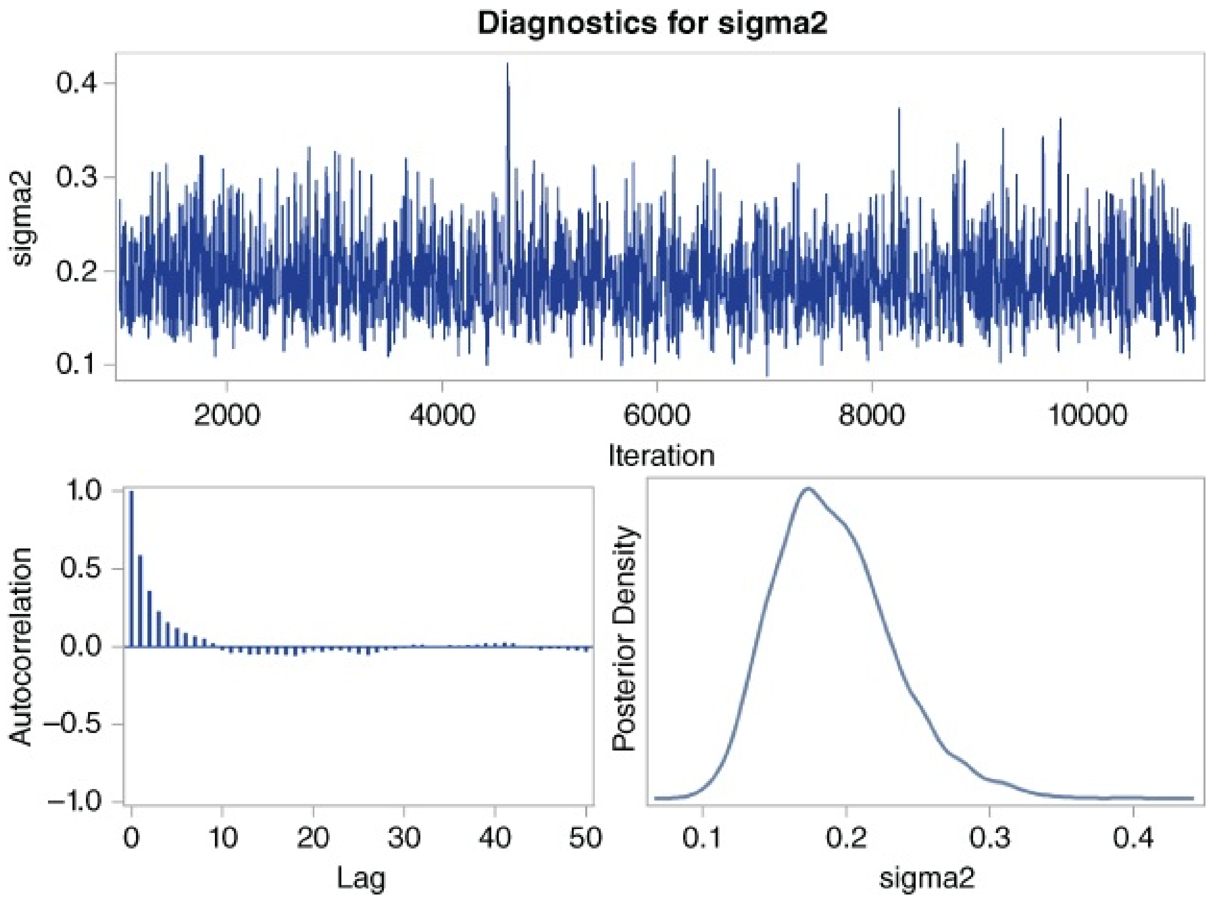

Exhibit 12.29 Diagnostic Plots

```
/*Probit-Linear Regression without Covariates with MCMC*/
ODS LISTING CLOSE;
ODS GRAPHICS ON / IMAGEFMT=PDF  IMAGENAME='bayes2corr';
PROC MCMC DATA = data.tmp2 Statistics=ALL Diagnostics=ALL Scale=5
MINTUNE=5 NBI=1000 NMC=10000 THIN=2 PROPCOV=QUANEW SEED=12345;
PARMS beta0 0 sigma2 0.1;
PRIOR beta0  ~ normal(0, var = 10000);
PRIOR sigma2 ~ lognormal(0.1, var = 0.1);
mu = beta0;
MODEL probit_dr ~ n(mu, var = sigma2);
RUN;
ODS GRAPHICS OFF;
```

PD ESTIMATION FOR LOW DEFAULT PORTFOLIOS

An interesting and helpful application of Bayesian statistics is in the case where only sparse data are available, such as low default portfolios (LDPs). The approaches already presented in the chapter on LDPs can be augmented by Bayesian approaches. Early work on this has been done by Kiefer (2009), from whom we use an illustration of the approach before we apply it to real data. Consider a portfolio of $n = 100$ loans and $d = 0$ defaults. Assume further that all loans have the same PD π and are uncorrelated. Hence, the number of defaults is binomially distributed. The likelihood function for the parameter π is then given by the binomial distribution and is plotted in Exhibit 12.30. The maximum is at value $\pi = 0$, which yields the maximum-likelihood estimator. Similarly, if there were $d \in \{1, 2, 5\}$ defaults, the ML estimate would be at the peak of the likelihood with 1 percent, 2 percent, and 5 percent, respectively. Using 0 percent as an estimate for the PD of a segment of a loan portfolio obviously does not make much sense.

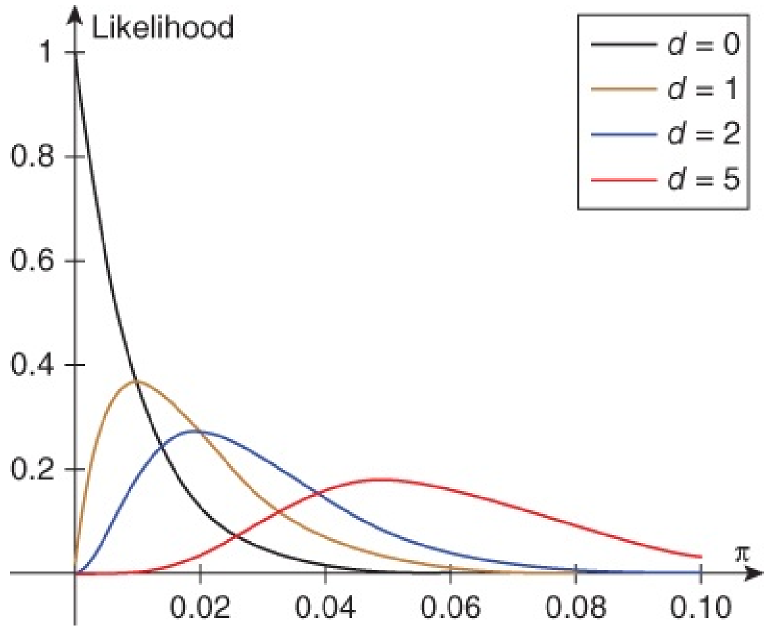

Exhibit 12.30 Various Likelihoods

We can now mix the sample information with prior information. Suppose there is an expert who specifies a prior of the PD as a beta distribution, for example with parameters $\alpha = 1$ and $\beta = 200$ yielding the black solid distribution in Exhibit 12.31.[1] We can then compute the posterior distributions for various numbers of defaults.[2] For example, for $d = 0$ the maximum of the posterior distribution would be zero. However, in Bayesian statistics we would rather use the mean of the distribution as an estimate for the parameter, which is now about 0.5 percent. Similarly, if $d = 5$ the estimate for the PD would no longer be 5 percent but rather somewhere around 2 percent. Hence, the Bayesian estimate for LDP pulls the likelihood toward the prior distribution.

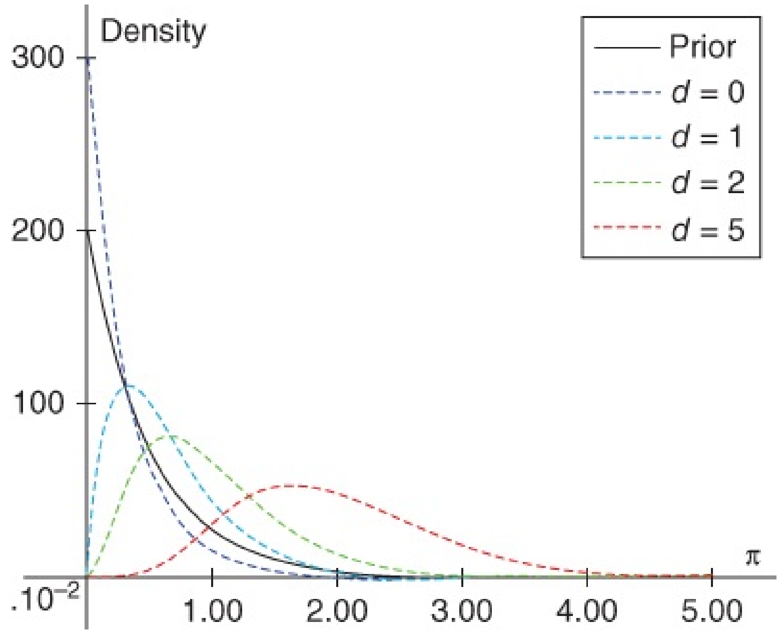

Exhibit 12.31 Prior and Posterior Distributions

Let us now apply this to our data set. Before we do so, we construct a sample that actually exhibits LDP properties. We simply pick out those mortgages that have a FICO score higher than 810, resulting in only 0 or very small fractions of defaults per time. The average default rate is about 0.27 percent. We then use PROC MCMC to specify a binomial distribution for the number of defaults per time, given the number of observations per time. For the PD parameter of the binomial distribution, we specify a $beta(1,1)$ prior distribution, which is actually a uniform distribution and reflects uninformativeness. PROC MCMC shows good convergence of the simulations, and the mean of the posterior distribution is about 0.4 percent with a 95 percent equal tail credibility interval ranging from 0.16 percent to 0.73 percent. Hence, the uninformative prior pulls the estimate away from the low average default rate. (See Exhibits 12.32 and 12.33.)

The MCMC Procedure

Number of Observations Read	47	Number of Observations Used	47

Parameters				
Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	pd	Conjugate	0.5000	beta(1,1)

The MCMC Procedure

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25	50	75
pd	10000	0.00393	0.00148	0.00284	0.00374	0.00482

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
pd	0.050	0.00160	0.00730	0.00140	0.00691

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
pd	0.0080	−0.0049	−0.0079	−0.0062

Geweke Diagnostics		
Parameter	z	Pr > |z|
pd	−2.0005	0.0454

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/−0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
pd	2	3741	3746	0.9987

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
pd	0.1329	0.4460	Passed	0	0.000031	0.00393	0.00786	Passed

Exhibit 12.32 Summary Statistics

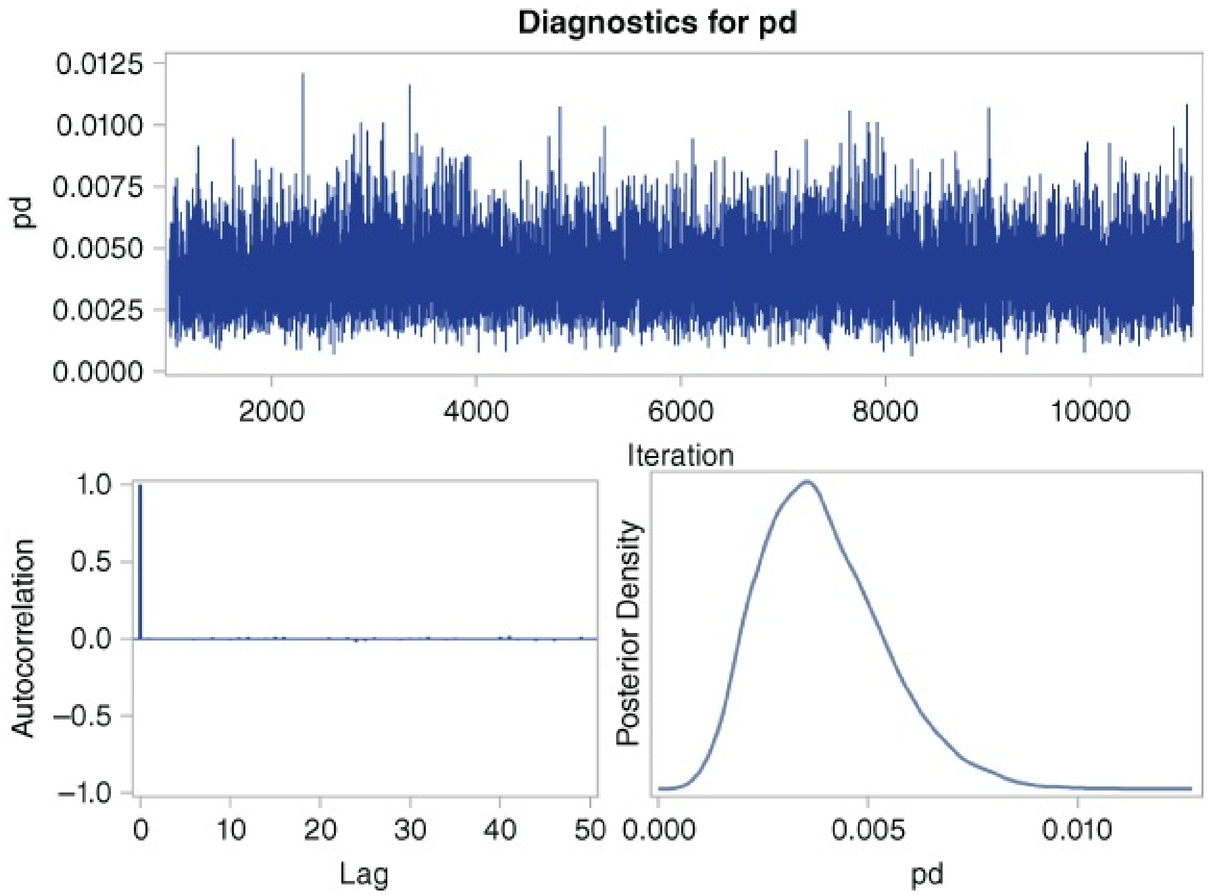

Exhibit 12.33 Diagnostic Plots

```
DATA tmp;
SET data.mortgage;
IF FICO_orig_time> 810;
RUN;
PROC SORT DATA = tmp;
BY time;
RUN;
PROC MEANS data = tmp;
VAR default_time;
BY time;
OUTPUT OUT = means;
RUN;
DATA tmp_ldp;
SET means;
IF _STAT_ = "MEAN";
n_default = default_time * _FREQ_;
default_time_1 = LAG(default_time);
RUN;

/*Independence model*/
ODS GRAPHICS ON;
PROC MCMC DATA = tmp_ldp STATISTICS=ALL DIAGNOSTICS=ALL
    NMC=10000 SEED=12345;
PARMS pd;
PRIOR pd ~ beta(1,1);
MODEL n_default ~ BINOMIAL(_FREQ_,pd);
RUN;
```

```
ODS GRAPHICS OFF;
```

One can also extend the Bayesian LDP approach to a more advanced parametric model, such as a probit model for individual data with a sample of 1,940 low default observations. (See Exhibits 12.34 through 12.38.) First, we estimate a classical probit model with PROC LOGISTIC with LTV and GDP as covariates. Neither of them is statistically significant at the 5 percent level. For the Bayes model, we use uninformative priors with normal distributions with a very high variance of 10,000. The output of the classical analysis returns estimates for LTV of 0.0815 and for GDP of 0.1062, whereas the means in the Bayesian model are lower. Hence, the means of the posterior distributions are quite different from the classical estimates. Also check whether the value of 0 lies within an equal—tail 95 percent credibility interval for GDP and LTV, which would signal low significance. We could now continue the analysis by specifying a more informative prior if there is valuable information about the parameters.

The LOGISTIC Procedure

Model Information	
Data Set	WORK.TMP
Response Variable	default_time
Number of Response Levels	2
Model	Binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1780
Number of Observations Used	1780

Response Profile		
Ordered Value	default_time	Total Frequency
1	1	6
2	0	1774

Note	Probability modeled is default_time=1.

Model Convergence Status
Convergence criterion (GCONV=1E−8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	82.291	82.441
SC	87.775	98.894
−2 Log L	80.291	76.441

R-Squared	0.0022	Max-Rescaled *R*-Squared	0.0490

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	−12.0679	3.6221	11.1006	0.0009
LTV_orig_time	1	0.0815	0.0454	3.2266	0.0725
gdp_time	1	0.1062	0.2397	0.1963	0.6578

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	67.4	Somers' D	0.353
Percent Discordant	32.1	Gamma	0.355
Percent Tied	0.6	Tau-a	0.002
Pairs	10644	c	0.676

Exhibit 12.34 Probit Model

The MCMC Procedure

Number of Observations Read	1780	Number of Observations Used	1780

Parameters				
Block	Parameter	Sampling Method	Initial Value	Prior Distribution
1	beta0	*N*−Metropolis	0	normal(0, var = 10000)
	beta1		0	normal(0, var = 10000)
	beta2		0	normal(0, var = 10000)

The MCMC Procedure

Posterior Summaries						
Parameter	*N*	Mean	Standard Deviation	Percentiles		
				25	50	75
beta0	10000	−5.1413	1.2773	−5.9951	−5.0784	−4.2407
beta1	10000	0.0300	0.0160	0.0188	0.0294	0.0408
beta2	10000	0.0556	0.0837	−0.00266	0.0486	0.1059

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
beta0	0.050	−7.7185	−2.8015	−7.6983	−2.7956
beta1	0.050	0.000467	0.0618	0.000811	0.0621
beta2	0.050	−0.0882	0.2382	−0.1007	0.2142

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
beta0	0.8732	0.5026	0.2620	0.0038
beta1	0.8690	0.4912	0.2518	−0.0027
beta2	0.8415	0.4597	0.2413	0.0009

Geweke Diagnostics		
Parameter	z	Pr > |z|
beta0	−1.9757	0.0482
beta1	1.8469	0.0648
beta2	1.0318	0.3022

Raftery-Lewis Diagnostics				
Quantile = 0.025 Accuracy = +/−0.005 Probability = 0.95 Epsilon = 0.001				
Parameter	Number of Samples			Dependence Factor
	Burn-In	Total	Minimum	
beta0	23	24192	3746	6.4581
beta1	20	21310	3746	5.6887
beta2	18	19686	3746	5.2552

Heidelberger-Welch Diagnostics								
Parameter	Stationarity Test				Half-Width Test			
	Cramer–von Mises Stat	*p*-Value	Test Outcome	Iterations Discarded	Half-Width	Mean	Relative Half-Width	Test Outcome
beta0	0.2758	0.1584	Passed	0	0.0969	−5.1413	−0.0188	Passed
beta1	0.1905	0.2867	Passed	0	0.00119	0.0300	0.0395	Passed
beta2	.	.	Failed	.	.	.	.	

Exhibit 12.35 Summary Statistics

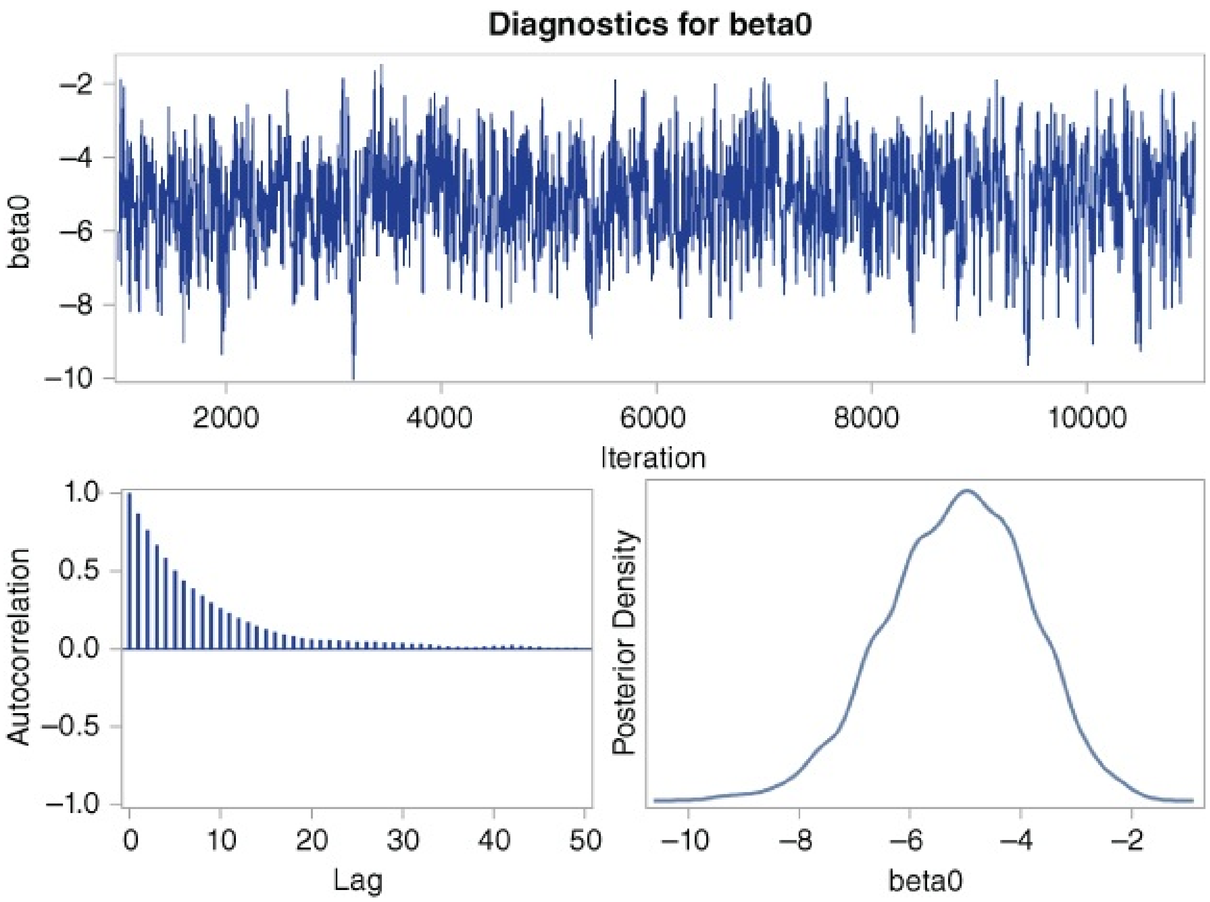

Exhibit 12.36 Diagnostic Plots

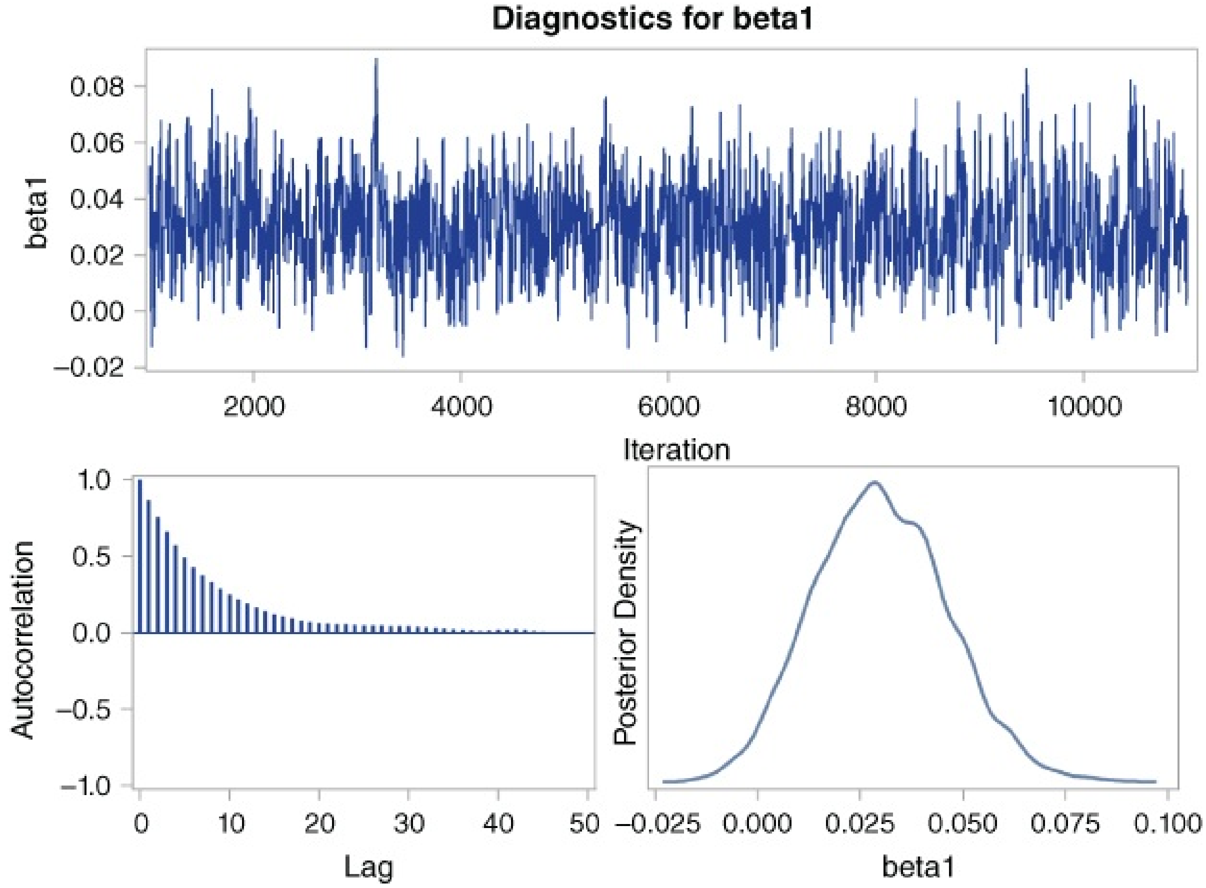

Exhibit 12.37 Diagnostic Plots

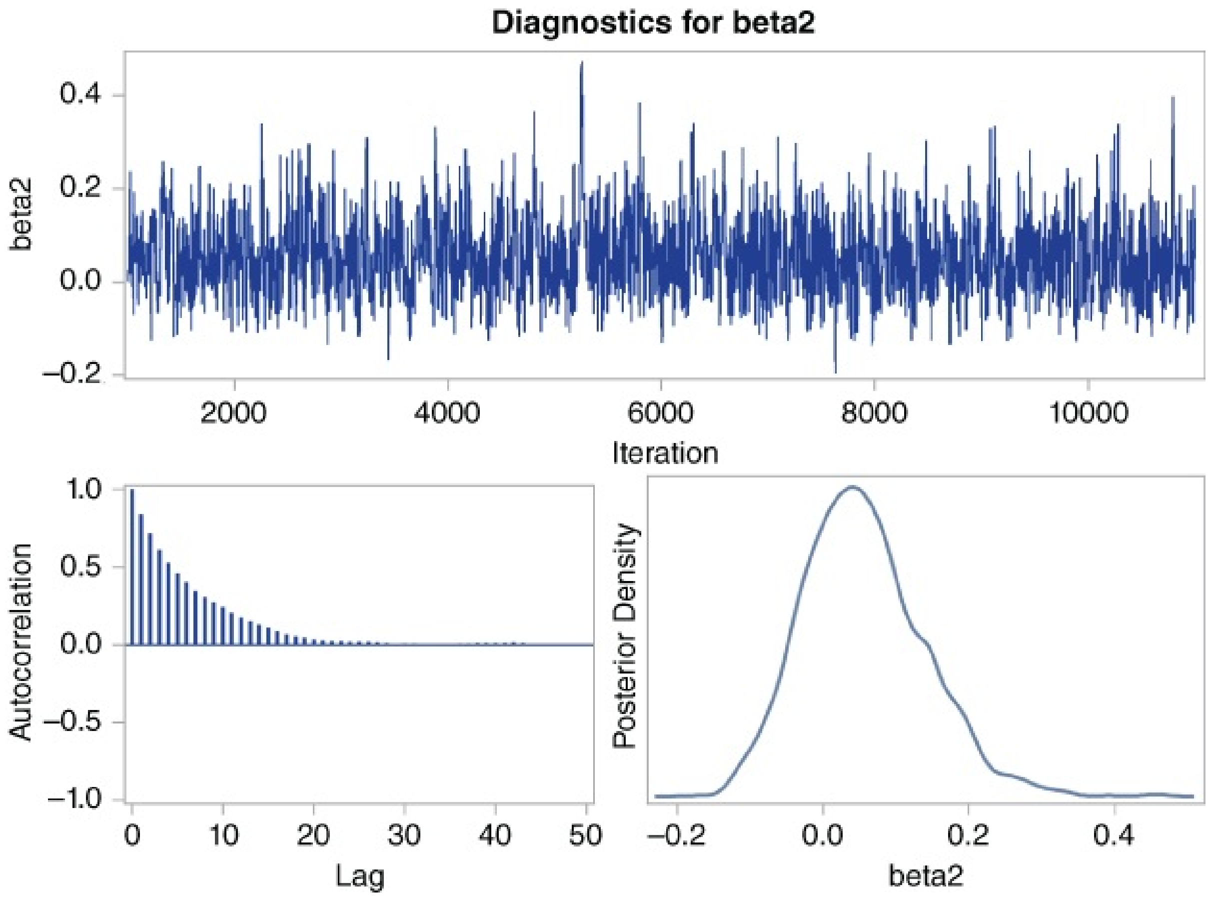

Exhibit 12.38 Diagnostic Plots

```
ODS GRAPHICS ON;
PROC LOGISTIC DATA=tmp DESCENDING;
MODEL default_time =
LTV_orig_time gdp_time
/ OUTROC=roc_logistic RSQUARE;
STORE OUT=model_logistic;
RUN;
ODS GRAPHICS OFF;

ODS GRAPHICS ON;
ODS LISTING CLOSE;
ODS LATEX;
ODS GRAPHICS ON / IMAGEFMT=PDF  IMAGENAME='bayesldp';
PROC MCMC DATA = tmp STATISTICS=ALL DIAGNOSTICS=ALL
NMC=10000 SEED=12345;
PARMS beta0 beta1 beta2;
PRIOR beta0 beta1 beta2 ~ normal(0, var = 10000);
pd = PROBNORM(beta0 +
beta1 * LTV_orig_time + beta2* gdp_time);
MODEL default_time ~ binomial(1,pd);
RUN;
ODS GRAPHICS OFF;
```

PRACTICE QUESTIONS

1. Describe the differences between the Bayesian approach and the classical frequentist approach.
2. Consider the first example in this chapter. Evaluate the Bayesian model with more informative priors for FICO and GDP in addition to LTV.
3. Run a Bayesian regression model for estimating asset correlation with inclusion of the lagged default rate, and interpret the results.
4. Run a Bayesian probit model for the low default portfolio using more informative priors where the mean of the GDP coefficient is 0.3 and the mean of the FICO coefficient is 0.02. Interpret and discuss the results.

NOTES

[1] Kiefer (2009) actually proposes a generalized 4-parameter beta distribution and obtains the prior from a real-world expert.

[2] It is well known in Bayesian statistics that a binomial likelihood and a beta prior result in a posterior that is also a beta distribution; see Greenberg (2014). Thus, the beta distribution is the *conjugate* prior for the binomial distribution.

References

Greenberg, E. 2014. *Introduction to Bayesian Econometrics*. 2nd ed. Cambridge: Cambridge University Press.

Kiefer, N. 2009. “Default Estimation for Low-Default Portfolios.” *Journal of Empirical Finance* 16: 164–173.

SAS Institute Inc. 2015. *SAS/STAT 14.1 User's Guide: Technical Report*. Cary, NC: SAS Institute.