Part III. Estimation

In this part, we show how to estimate and implement the models introduced in part II. Chapter 9 implements the full Bayesian version of our model via Markov Chain Monte Carlo algorithms. Chapter 10 shows how to implement a faster estimation procedure, not requiring Markov Chains.

9 Markov Chain Monte Carlo Estimation

In practical applications, researchers can build a prior using any combination of the results developed in part II. Once this step has been performed we are left with a prior of the mathematical form given in equation 7.12 (page 132), except for the fact that the exponent is likely to contain a sum of l terms of that form, with different parameters $\theta_1, \ldots, \theta_l$ and different matrices $\mathbf{C}_{ij}$. Then the only thing left to do is to assume some reasonable prior densities for σ and θ, plug them in equation 4.3, and estimate the mean of the posterior distribution of $\boldsymbol{\beta}$. In this chapter, we summarize the complete model, filling in these remaining details, and then describe a method of estimation based on the Gibbs sampler to calculate quantities of interest (Tanner, 1996). We report our calculations for the case in which there is only one prior (e.g., the one for smoothness over age groups). The full details of the more general case appears in the software accompanying this book.

9.1 Complete Model Summary

We now review the model and identify the main quantities involved in the estimation. We begin with the full posterior density (reproduced from equation 4.2, page 58):

$$\mathcal{P}(\boldsymbol{\beta}, \sigma, \theta | m) \propto \mathcal{P}(m | \boldsymbol{\beta}, \sigma) \mathcal{P}(\boldsymbol{\beta} | \theta) \mathcal{P}(\theta) \mathcal{P}(\sigma). \tag{9.1}$$

In this section, we now formally define each of the densities on the right side of equation 9.1, so that we can compute the mean posterior of the coefficients (from equation 4.3):

$$\boldsymbol{\beta}^{\text{Bayes}} \equiv \int \boldsymbol{\beta} \mathcal{P}(\boldsymbol{\beta}, \sigma, \theta | m) d\boldsymbol{\beta} d\theta d\sigma, \tag{9.2}$$

and our forecasts.

9.1.1 Likelihood

Each cross section i includes T_i observations and has its own standard deviation σ_i. As explained in section 3.1.2, we allow the observed values of the dependent variable m_{it} to be weighted. Therefore, instead of m and $\mathbf{Z}$, and the likelihood $\mathcal{P}(m|\boldsymbol{\beta}, \sigma)$, we use their weighted counterparts y and $\mathbf{X}$ (see equations 3.8 and 3.10, pages 45). The weighted version of the likelihood is then:

$$\mathcal{P}(y|\boldsymbol{\beta}, \sigma) \propto \left(\prod_i (\sigma_i^{-2})^{\frac{T}{2}}\right) \exp\left(-\frac{1}{2}\sum_i \frac{1}{\sigma_i^2}\sum_t (y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_i)^2\right). \tag{9.3}$$

9.1.2 Prior for β

We consider a prior for $\boldsymbol{\beta}$ of the form described in equation 7.12 (page 132). This prior has only one hyperparameter, θ, and can be expressed as

$$\mathcal{P}(\boldsymbol{\beta}|\theta) = K\theta^{\frac{r}{2}} \exp\left(-\frac{1}{2}\theta \sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j\right), \tag{9.4}$$

where r is the rank of the matrix defining the quadratic form in the exponent in 9.4 (see equation C.8, page 244).

9.1.3 Prior for σ_i

Because we do not desire it to have a major influence on our results, the functional form of the prior for σ_i is chosen for convenience. Therefore, we follow standard practice and choose an inverse Gamma prior for σ_i^2 (see Gelman, 2006 for an alternative):

$$\sigma_i^{-2} \sim \mathcal{G}(\mathbb{d}/2, \mathbb{e}/2). \tag{9.5}$$

Here, $\mathbb{d}$ and $\mathbb{e}$ are user-specified parameters that determine the mean and the variance of σ_i^{-2} as follows:

$$\mathbb{E}[\sigma_i^{-2}] = \frac{\mathbb{d}}{\mathbb{e}}, \quad \mathbb{V}[\sigma_i^{-2}] = \frac{\mathbb{d}}{\mathbb{e}^2}. \tag{9.6}$$

In order to specify these parameters, the user must specify the mean and variance of σ_i^{-2} and then solve equation 9.6 for $\mathbb{d}$ and $\mathbb{e}$. The user may not have prior knowledge of σ_i^{-2} and is more likely to have some knowledge of σ_i (see section 6.5.3). Therefore, one should relate the parameters $\mathbb{d}$ and $\mathbb{e}$ to moments of σ_i rather than σ_i^{-2}. This is not totally straightforward, because the resulting formulas do not allow for closed form solutions. We derive here all the necessary formulas for a numerical solution.

Because the object about which we have prior knowledge is σ_i, it is important to understand how the prior density in equation 9.5 looks in terms of σ_i. Defining the auxiliary variable $s_i \equiv \sigma_i^{-2}$ from the definition of Gamma density, we rewrite the density of equation 9.5 as

$$\mathcal{P}(s_i; \mathbb{d}, \mathbb{e}) = \frac{1}{\Gamma(\frac{\mathbb{d}}{2})} \left(\frac{\mathbb{e}}{2}\right)^{\frac{\mathbb{d}}{2}} s_i^{\frac{\mathbb{d}}{2}-1} e^{-\frac{\mathbb{e}}{2} s_i}.$$

The density for σ_i is derived with a simple change of variable, and it has the following form:

$$\mathcal{P}(\sigma_i; \mathbb{d}, \mathbb{e}) = \frac{2}{\Gamma(\frac{\mathbb{d}}{2})} \left(\frac{\mathbb{e}}{2}\right)^{\frac{\mathbb{d}}{2}} \sigma_i^{-\mathbb{d}-1} e^{-\frac{\mathbb{e}}{2}\frac{1}{\sigma_i^2}}.$$

A lengthy but straightforward calculation shows that the moments of the preceding density are as follows:

$$\mathbb{E}[\sigma_i^n] = \frac{\Gamma(\frac{\mathbb{d}-n}{2})}{\Gamma(\frac{\mathbb{d}}{2})} \left(\frac{\mathbb{e}}{2}\right)^{\frac{n}{2}}. \tag{9.7}$$

The mean and variance of the prior density for σ_i are therefore:

$$\mathbb{E}[\sigma_i] = \frac{\Gamma(\frac{\mathbb{d}-1}{2})}{\Gamma(\frac{\mathbb{d}}{2})} \left(\frac{\mathbb{e}}{2}\right)^{\frac{1}{2}}$$

$$\mathbb{V}[\sigma_i] = \frac{\mathbb{e}}{\mathbb{d}-2} - \mathbb{E}[\sigma_i]^2. \tag{9.8}$$

Notice that the variance of σ_i goes to infinity as $\mathbb{d}$ tends to 2, and therefore we impose the constraint $\mathbb{d} > 2$. The relationship 9.8 cannot be inverted to express $\mathbb{d}$ and $\mathbb{e}$ as a function of $\mathbb{E}[\sigma_i]$ and $\mathbb{V}[\sigma_i]$, and so a numerical procedure is necessary. To this end, it is convenient to rewrite the equations 9.8 as:

$$\frac{\mathbb{E}[\sigma_i]}{\sqrt{\mathbb{V}[\sigma_i] + \mathbb{E}[\sigma_i]^2}} = \frac{\Gamma(\frac{\mathbb{d}-1}{2})}{\Gamma(\frac{\mathbb{d}}{2})} \sqrt{\frac{\mathbb{d}}{2} - 1}$$

$$\mathbb{e} = (\mathbb{d} - 2)\left(\mathbb{V}[\sigma_i] + \mathbb{E}[\sigma_i]^2\right). \tag{9.9}$$

The top part of equation 9.9 is easily solved numerically (for $d > 2$), because it has only one solution, and once a value for $\mathbb{d}$ is obtained, the value for $\mathbb{e}$ follows via simple substitution.

9.1.4 Prior for θ

Here again we follow standard practice and choose a Gamma prior for θ:

$$\theta \sim \mathcal{G}(\mathbb{f}/2, \mathbb{g}/2), \tag{9.10}$$

where $\mathbb{f}$ and $\mathbb{g}$ are user specified parameters, which control the mean and the variance of θ through formulas like those in equation 9.6. We have shown in section 6.2 (page 97) how to make reasonable choices for the mean and the variance of θ, and therefore we can use those results to set $\mathbb{f}$ and $\mathbb{g}$.

9.1.5 The Posterior Density

Now we can bring together all the quantities just defined and write the full posterior density as

$$\mathcal{P}(\boldsymbol{\beta}, \sigma, \theta | y) \propto \left(\prod_i (\sigma_i^{-2})^{\frac{\mathbb{d}+T_i}{2}-1} e^{-\frac{1}{2}\mathbb{e}\sigma_i^{-2}} \right) \left(\theta^{\frac{\mathbb{f}+r}{2}-1} e^{-\frac{\mathbb{g}}{2}\theta} \right)$$
$$\times \exp\left(-\frac{1}{2} \left[\sum_i \frac{1}{\sigma_i^2} \sum_t (y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_i)^2 + \theta \sum_{ij} W_{ij} \boldsymbol{\beta}_i' \mathbf{C}_{ij} \boldsymbol{\beta}_j \right] \right). \quad (9.11)$$

In the next section, we briefly describe the Gibbs algorithm used to sample from this density and to compute the posterior mean in equation 9.1.

9.2 The Gibbs Sampling Algorithm

We evaluate of the conditional mean in equation 9.2 using a Markov Chain Monte Carlo (MCMC) approach. In this section, we give the expressions needed to implement the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990), one of the most commonly used MCMC techniques. We refer the reader to standard textbooks on MCMC for a description of the Gibbs sampler (Gelman et al., 2003; Gilks, Richardson, and Spiegelhalter, 1996; Gill, 2002; Tanner, 1996). We describe this algorithm with reference to the prior in equation 7.12, with only one hyperparameter θ and only one type of cross-sectional index, generically denoted by i.

To draw random samples from the posterior density in equation 9.11, we use the Gibbs sampling algorithm. The essence of the Gibbs sampler lies in breaking a complicated joint probability density into a set of full conditional densities, and sampling one variable (or a group of variables) at a time, conditional on the values of the others.

In our case we have three sets of variables, $\boldsymbol{\beta}$, σ, and θ, so that one iteration of the algorithm consists of sampling each of these sets. To simplify our notation, we denote the density of a variable x conditional on all the others by $\mathcal{P}(x|\text{else})$. Then, we write an iteration of the Gibbs sampler containing the following steps:

1. $\sigma_i^{-2} \sim \mathcal{P}(\sigma_i^{-2}|\text{else}) \quad i = 1, \dots, n$
2. $\theta \sim \mathcal{P}(\theta|\text{else})$
3. $\boldsymbol{\beta}_i \sim \mathcal{P}(\boldsymbol{\beta}_i|\text{else}) \quad i = 1, \dots, n$

Once we know how to sample from the preceding conditional densities, we compute the posterior mean of $\boldsymbol{\beta}$ in equation 9.1 by averaging the values of $\boldsymbol{\beta}$ obtained by repeating these steps a large number of times (after having discarded a suitable number of "burn-in" iterations to ensure that the algorithm has converged, possibly also with separate chains; we do not worry about autocorrelation in the series unless we are computing standard errors). Because the conditional densities need to be known only up to a normalization factor, we only need terms in the posterior that include the variables of interest.

We now derive each of these conditional densities. We also show that each can be understood simply as a weighted average of a maximum likelihood estimate and the prior mean.

9.2.1 Sampling σ

The Conditional Density

When we choose the prior for σ_k, we implicitly assume that the relevant variable (the one with the gamma density) was σ_k^{-2}, rather than σ_k. Consistently with that choice we use σ_k^{-2} as the sampled variable and pick from equation 9.11 all the terms that contain σ_k^{-2}, grouping the others in a generic normalization constant. We thus obtain

$$\mathcal{P}(\sigma_k^{-2} \mid \text{else}) \propto (\sigma_k^{-2})^{\frac{\mathbb{d}+T_k}{2}-1} e^{-\frac{1}{2}\mathbb{e}\sigma_k^{-2}} \exp\left(-\frac{1}{2}\sigma_k^{-2}\sum_t (y_{kt} - \mathbf{X}'_{kt}\boldsymbol{\beta}_k)^2\right),$$

which is a Gamma distribution for σ_k^{-2}. Thus, by defining

$$\text{SSE}_k \equiv \sum_t (y_{kt} - \mathbf{X}'_{kt}\boldsymbol{\beta}_k)^2, \tag{9.12}$$

we conclude that sampling for σ_k^{-2} should be as follows:

$$\sigma_k^{-2}|\text{else} \sim \mathcal{G}\left(\frac{\mathbb{d}+T_k}{2}, \frac{\mathbb{e}+\text{SSE}_k}{2}\right). \tag{9.13}$$

Interpretation

In order to clarify this expression further, we write the conditional expected value of σ_k^{-2}:

$$\mathbb{E}[\sigma_k^{-2}|\text{else}] = \frac{\mathbb{d}+T_k}{\mathbb{e}+\text{SSE}_k},$$

and define, respectively, $\sigma^2_{\text{P},k}$, which is related to the expected value of σ_k^2 under the prior in equation 9.5, and $\sigma^2_{k,\text{ML}}$, the usual maximum likelihood estimator of σ_k^2:

$$\sigma^2_{\text{P},k} \equiv \frac{\mathbb{e}}{\mathbb{d}} = \frac{1}{\mathbb{E}[\sigma_k^{-2}]}, \quad \sigma^2_{k,\text{ML}} = \frac{\text{SSE}_k}{T_k}.$$

Now rewrite the equation for the conditional mean of σ_k^{-2}:

$$\mathbb{E}[\sigma_k^{-2}|\text{else}] = \left(\frac{\mathbb{d}\,\sigma_{\text{P},k}^2 + T_k\sigma_{k,\text{ML}}^2}{T_k + \mathbb{d}}\right)^{-1}.$$

This expression helps clarify that when $\mathbb{d}$ is large—when the prior density is highly concentrated around its mean—the conditional mean of σ_k^{-2} is very close to the prior mean. On the other hand, when the number of observations T_k is large, then the likelihood dominates, and the conditional mean becomes determined by the likelihood. Although this conclusion is to be expected, it is useful, because it makes clear the "dual" role of the quantities $\mathbb{d}$ and T_k, which control the trade-off between the prior and the likelihood as measures of concentration around the mean. It is also possible to show a similar kind of duality between $\mathbb{e}$ and SSE_k.

9.2.2 Sampling θ

The Conditional Density

Proceeding in the same way, we write the conditional distribution for θ as

$$\mathcal{P}(\theta \mid \text{else}) \propto \theta^{\frac{\mathbb{f}+r}{2}-1} e^{-\frac{1}{2}\theta\mathbb{g}} \exp\left(-\frac{1}{2}\theta \sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j\right),$$

which is again a Gamma distribution. Thus, sampling for θ may be done according to the following, expressed in standard form:

$$\theta \sim \mathcal{G}\left(\frac{\mathbb{f}+r}{2}, \frac{\mathbb{g}}{2} + \frac{1}{2}\sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j\right). \tag{9.14}$$

Interpretation

Again, we examine the conditional mean of θ, which is

$$\mathbb{E}[\theta|\text{else}] = \frac{\mathbb{f}+r}{\mathbb{g} + \sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j}.$$

In order to interpret this expression, we define, respectively, θ_{P}, the expected value of θ under its prior (see equation 9.10), and θ_{ML}, the maximum likelihood (ML) estimator of θ, that is, the value of θ that maximizes $\mathcal{P}(\boldsymbol{\beta}|\theta)$ (see equation 9.4):

$$\theta_{\text{P}} \equiv \mathbb{E}[\theta] = \frac{\mathbb{f}}{\mathbb{g}}, \quad \theta_{\text{ML}} \equiv \frac{r}{\sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j}.$$

Rewriting the preceding equation for the conditional mean as an equation for its reciprocal is easier. Although the mean of the reciprocal is not the reciprocal of the mean, these quantities are related, which is enough for the purpose of explanation:

$$\frac{1}{\mathbb{E}[\theta|\text{else}]} = \frac{\mathbb{f}\frac{1}{\theta_{\text{P}}} + r\frac{1}{\theta_{\text{ML}}}}{\mathbb{f} + r}.$$

As is the case for σ, this expression depends on the trade-off between two terms: one that relates to the prior of θ and another that relates to its likelihood. The parameters which control this trade-off are $\mathbb{f}$ and r. The parameter $\mathbb{f}$ controls how concentrated is the prior distribution of θ around its mean. By the same token, we would expect that r, the rank of the matrix in the exponent of $\mathcal{P}(\boldsymbol{\beta}|\theta)$, describes the concentration of $\mathcal{P}(\boldsymbol{\beta}|\theta)$ around its mean. This can be seen by writing the density for the random variable $H = \sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j$, which is the relevant expression from the point of view of θ. Using the techniques described in appendix C, we can show that

$$H \sim \mathcal{G}(r, \theta).$$

From here we can see immediately that r plays for the likelihood the same role played by $\mathbb{f}$ in the prior and is indeed a measure of concentration.

9.2.3 Sampling β

The Conditional Density

In order to find the distribution of $\boldsymbol{\beta}_k$ with all the other variables held constant, we need to isolate from the posterior all terms that depend on $\boldsymbol{\beta}_k$. As a first pass, we eliminate unnecessary multiplicative terms in equation 9.4 and write

$$\mathcal{P}(\boldsymbol{\beta}_k|\text{else}) \propto \exp\left(-\frac{1}{2}\left[\frac{1}{\sigma_k^2}\sum_t (y_{kt} - \mathbf{X}_{kt}'\boldsymbol{\beta}_k)^2 + \theta\sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j\right]\right). \tag{9.15}$$

We collect in a generic term K all the terms that do not depend on $\boldsymbol{\beta}_k$ (we reuse the symbol K to refer to possibly different constants for each subsequent equation). For the first term in equation 9.15, we have

$$\sum_t (y_{kt} - \mathbf{X}_{kt}'\boldsymbol{\beta}_k)^2 = \boldsymbol{\beta}_k'\mathbf{X}_k'\mathbf{X}_k\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k'\mathbf{X}_k'y_k + K.$$

For the second term in equation 9.15, we have

$$\sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j = W_{kk}\boldsymbol{\beta}_k'\mathbf{C}_{kk}\boldsymbol{\beta}_k + 2\sum_{j\neq k} W_{jk}\boldsymbol{\beta}_k'\mathbf{C}_{kj}\boldsymbol{\beta}_j + K.$$

If we use the quadratic form identity in appendix B.2.6 (page 237), $W = s^+ - s$, and so we rewrite the preceding expression as

$$\sum_{ij} W_{ij}\boldsymbol{\beta}_i'\mathbf{C}_{ij}\boldsymbol{\beta}_j = s_k^+\boldsymbol{\beta}_k'\mathbf{C}_{kk}\boldsymbol{\beta}_k - 2\sum_j s_{jk}\boldsymbol{\beta}_k'\mathbf{C}_{kj}\boldsymbol{\beta}_j + K.$$

Putting everything together,

$$\mathcal{P}(\boldsymbol{\beta}_k|\text{else}) \propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}_k'\left(\frac{\mathbf{X}_k'\mathbf{X}_k}{\sigma_k^2} + \theta s_k^+ \mathbf{C}_{kk}\right)\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k'\left(\frac{\mathbf{X}_k' y_k}{\sigma_k^2} + \theta \sum_j s_{jk}\mathbf{C}_{kj}\boldsymbol{\beta}_j\right)\right]\right),$$

we now define

$$\Lambda_k^{-1} \equiv \frac{\mathbf{X}_k'\mathbf{X}_k}{\sigma_k^2} + \theta s_k^+ \mathbf{C}_{kk}$$

and

$$\overline{\boldsymbol{\beta}_k} \equiv \frac{\mathbf{X}_k' y_k}{\sigma_k^2} + \theta \sum_j s_{jk}\mathbf{C}_{kj}\boldsymbol{\beta}_j.$$

With these definitions we have

$$\begin{aligned}\mathcal{P}(\boldsymbol{\beta}_k|\text{else}) &\propto \exp\left(-\frac{1}{2}\left[\boldsymbol{\beta}_k'\Lambda_k^{-1}\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k'\overline{\boldsymbol{\beta}_k}\right]\right)\\ &= \exp\left(-\frac{1}{2}\left[(\boldsymbol{\beta}_k - \Lambda_k\overline{\boldsymbol{\beta}_k})'\Lambda_k^{-1}(\boldsymbol{\beta}_k - \Lambda_k\overline{\boldsymbol{\beta}_k})\right]\right).\end{aligned}$$

Therefore, we need to sample $\boldsymbol{\beta}_k$ as follows:

$$\boldsymbol{\beta}_k \sim \mathcal{N}(\Lambda_k\overline{\boldsymbol{\beta}_k}, \Lambda_k), \tag{9.16}$$

which is easily done by setting

$$\boldsymbol{\beta}_k = \Lambda_k\overline{\boldsymbol{\beta}_k} + \sqrt{\Lambda_k}\mathbf{b}, \quad \mathbf{b} \sim \mathcal{N}(0, I).$$

Interpretation

Despite the apparent complexity, equation 9.16 has a clear interpretation, similar to the interpretation of the formulas for the conditional means of σ_k^{-2} and θ. In those cases the conditional mean was a weighted average of two terms: one was interpreted as a maximum likelihood estimate, and the other was the prior mean. Because we expect to see the same phenomenon, we define the two quantities:

$$\boldsymbol{\beta}_k^{\text{ML}} \equiv (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k' y_k\ , \quad \boldsymbol{\beta}_k^{\text{p}} \equiv \sum_j \frac{s_{kj}}{s_k^+}\mathbf{C}_{kk}^{-1}\mathbf{C}_{kj}\boldsymbol{\beta}_j.$$

The quantity $\boldsymbol{\beta}_k^{\text{ML}}$ is simply the maximum likelihood estimator of $\boldsymbol{\beta}_k$. The quantity $\boldsymbol{\beta}_k^{\text{p}}$ is the conditional mean of the prior, in equation 8.5 (page 147).

In order to see the meaning of equation 9.16, we consider a special, but informative, case. Remember that by definition we have $\mathbf{C}_{kk} = T^{-1}\mathbf{Z}_k'\mathbf{Z}_k$. Here, $\mathbf{Z}_k$ is a vector of covariates extending over T time periods—the time over which we think prior knowledge is relevant. In general, the data matrix $\mathbf{Z}_k$ differs from the data matrix $\mathbf{X}_k$: $\mathbf{X}_k$ might include population weights, and it reflects the same pattern of missing values of m_{it}. Therefore, even without population weighting, the rows of $\mathbf{X}_{it}$ are a subset of the rows of $\mathbf{Z}_k$. Here, for the purpose of explanation, we assume that $\mathbf{X}_k$ and $\mathbf{Z}_k$ are identical, so that $\mathbf{C}_{kk} = T_k^{-1}\mathbf{X}_k'\mathbf{X}_k$. Using this assumption and a bit of algebra, we rewrite the conditional mean for $\boldsymbol{\beta}_k$ as follows:

$$\mathbb{E}[\boldsymbol{\beta}_k|\text{else}] = \frac{\frac{T_k}{\sigma_k^2}\boldsymbol{\beta}_k^{\text{ML}} + \theta s_k^+ \boldsymbol{\beta}_k^{\text{P}}}{\frac{T_k}{\sigma_k^2} + \theta s_k^+}.$$

As expected, the conditional mean of $\boldsymbol{\beta}_k$ is a weighted average of $\boldsymbol{\beta}_k^{\text{ML}}$ and $\boldsymbol{\beta}_k^{\text{P}}$. The weight of $\boldsymbol{\beta}_k^{\text{ML}}$ is large when the number of observations T_k is large, or when the noise affecting the observation (σ_k), which also measures the variance of $\boldsymbol{\beta}_k$ (in the likelihood), is small. In order to interpret the weight on $\boldsymbol{\beta}_k^{\text{P}}$, we need to inspect equation 8.5 (page 147). From this equation we see that the term θs_k^+ is inversely proportional to the (conditional) variance of $\boldsymbol{\beta}_k$ under the prior. Therefore, the weight on $\boldsymbol{\beta}_k^{\text{P}}$ is large when $\boldsymbol{\beta}_k$ has large prior variance; it is the counterpart of $\frac{1}{\sigma_k^2}$ in the weight on $\boldsymbol{\beta}_k^{\text{ML}}$.

9.2.4 Uncertainty Estimates

Once the Gibbs sampler has been implemented, no additional effort is needed to estimate model-based standard errors or confidence intervals for the forecast. This is done by producing, at every iteration of the Gibbs sampler (after the "burn-in" period), a forecast for each cross section based on the current sample from $\boldsymbol{\beta}$ and adding to it a random disturbance, sampled from a normal distribution with standard deviation given by the current sample from σ. The standard deviation for this random set of forecasts will give us an estimate of the standard errors. Of course, model-based uncertainty estimates do not take into account the most important source of error, which is the specification error, for which other techniques must be used.

9.3 Summary

This chapter offers a method of computing forecasts from our model in chapter 4 given the choice of any of the priors in chapters 5 or 7. The following chapter offers speedier versions that do not rely on the Gibbs algorithm.