15-50

The third and fourth models in Table 15.8 present the mixed model estimates. The first of them imposes the restriction that $\Gamma_{21} = 0$, or that the two random parameters are uncorrelated. The second mixed model allows Λ_{21} to be a free parameter. The implied estimators for σ_{u1}, σ_{u2} and $\sigma_{u,21}$ are the elements of $\Lambda\Lambda'$, or

$$\sigma_{u1}^2 = \Lambda_{11}^2,$$
$$\sigma_{u,21} = \Lambda_{11}\Lambda_{21},$$
$$\sigma_{u2}^2 = \Lambda_{21}^2 + \Lambda_{22}^2.$$

These estimates are shown separately in the table. Note that in all three random parameters models (including the random effects model which is equivalent to the mixed model with all $\alpha_{lm} = 0$ save for $\alpha_{1,1}$ and $\alpha_{2,1}$ as well as $\Lambda_{21} = \Lambda_{22} = 0.0$), the estimate of σ_ε is relatively unchanged. The three models decompose the variation across groups in the parameters differently, but the overall variation of the dependent variable is largely the same.

The interesting coefficient in the model is $\beta_{2,i}$. Reading across the row for *Educ*, one might suspect that the random parameters model has washed out the impact of education, since the "coefficient" declines from 0.04072 to 0.007607. However, in the mixed models, the "mean" parameter, $\alpha_{2,1}$, is not the coefficient of interest. The coefficient on education in the model is $\beta_{2,i} = \alpha_{2,1} + \alpha_{2,2}$ *Ability* + $\beta_{2,3}$ *Mother's education* + $\beta_{2,4}$ *Father's education* + $u_{2,i}$. A rough indication of the magnitude of this result can be seen by inserting the sample means for these variables, 0.052374, 11.4719, and 11.7092, respectively. With these values, the mean value for the education coefficient is approximately 0.0327. This is comparable, though somewhat smaller, than the estimates for the pooled and random effects model. Of course, variation in this parameter across the sample individuals was the objective of this specification. Figure 15.7 plots a kernel density estimate for the estimated conditional means for the 2,178 sample individuals. The figure shows the very wide range of variation in the sample estimates. 15.9

FIG 15.9

15.51

15.8

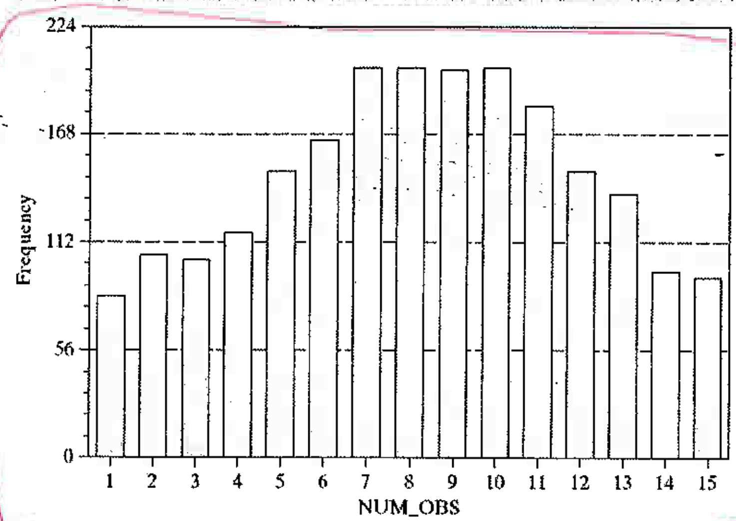
TABLE [illegible] Estimated Random Parameter Models

	Pooled OLS		Random Effects FGLS [Random Effects MLE]		Random Parameters	Random Parameters
Variable	*Estimate*	*Std.Err. (Robust)*	*Estimate [MLE]*	*Std.Err. [MLE]*	*Estimate (Std.Err.)*	*Estimate (Std.Err.)*
Exp	0.04157	0.001819 (0.002242)	0.04698 [0.04715]	0.001468 [0.001481]	0.04758 (0.001108)	0.04802 (0.001118)
Exp^2	−0.00144	0.0001002 (0.000126)	−0.00172 [−0.00172]	0.0000805 [0.000081]	−0.001750 (0.000063)	−0.001761 (0.0000631)
Broken	−0.02781	0.005296 (0.01074)	−0.03185 [−0.03224]	0.01089 [0.01172]	−0.01236 (0.003669)	−0.01980 (0.003534)
Sibs	−0.00120	0.0009143 (0.001975)	−0.002999 [−0.00310]	0.001925 [0.002071]	0.0000496 (0.000662)	−0.001953 (0.0006599)
Constant	0.09728	0.01589 (0.02783)	0.03281 [0.03306]	0.02438 [0.02566]	0.3277 (0.03803)	0.3935 (0.03778)
Ability					0.04232 (0.01064)	0.1107 (0.01077)
MEd					−0.01393 (0.0040)	−0.02887 (0.003990)
FEd					−0.007548 (0.003252)	0.002657 (0.003299)
σ_{u1}			0.172278 [0.18767]		0.004187 (0.001320)	0.5026
Educ	0.03854	0.001040 (0.002013)	0.04072 [0.04061]	0.001758 [0.001853]	0.01253 (0.003015)	0.007607 (0.002973)
Ability					−0.0002560 (0.000869)	−0.005316 (0.0008751)
MEd					0.001054 (0.000321)	0.002142 (0.0003165)
Fed					0.0007754 (0.000255)	0.00006752 (0.00001354)
σ_{u2}					0.01622 (0.000114)	0.03365
$\sigma_{u,12}$					0.0000	−0.01560
					0.0000	−0.92259
σ_ε	0.2542736		0.187017 [0.187742]		0.192741	0.1919182
Λ_{11}					0.004187 (0.001320)	0.5026 (0.008775)
Λ_{21}					0.0000 (0)	−0.03104 (0.0001114)
Λ_{22}					0.01622 (0.000113)	0.01298 (0.0006841)
ln *L*	−885.6740		[10480.18]		3550.594	3587.611

15-52

15.~~6~~8

FIGURE ~~9.7~~ Group Sizes for Wage Data Panel.



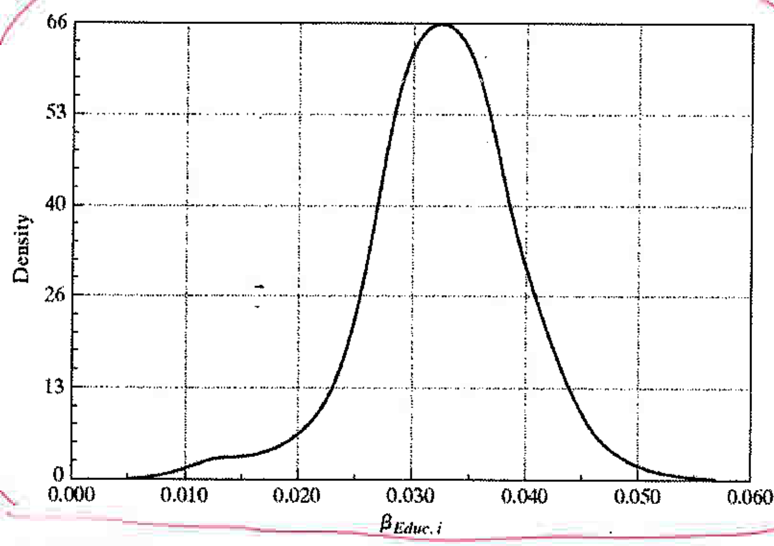


FIGURE ~~9.8~~ Kernel Density Estimate for Education Coefficient.

15.~~7~~9

15-53

15.11 Mixed Models and Latent Class Models

AU: Confirm section number

Sections 15.7-15-10 examined different approaches to modeling parameter heterogeneity. The fixed effects approach begun in Section 11.4 is extended to include the full set of regression coefficients in Section 11.11.1. where

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

and no restriction is placed on $E[\mathbf{u}_i|\mathbf{X}_i]$. Estimation produces a feasible GLS estimate of $\boldsymbol{\beta}$. Estimation of $\boldsymbol{\beta}$ begins with separate least squares estimation with each group, i – because of the correlation between $\mathbf{u}_i$ and $\mathbf{x}_{it}$, the pooled estimator is not consistent. The efficient estimator of $\boldsymbol{\beta}$ is then a mixture of the bis. We also examined an estimator of $\boldsymbol{\beta}_i$, using the optimal predictor from the conditional distributions, (15-39). The crucial assumption underlying the analysis is the possible correlation between $\mathbf{X}_i$ and $\mathbf{u}_i$. We also considered two modifications of this random coefficients model. First, a restriction of the model in which some coefficients are nonrandom provides a useful simplification. The familiar fixed effects model of Section 11.4 is such a case, in which only the constant term varies across individuals. Second, we considered a hierarchical form of the model

b's

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \mathbf{u}_i. \tag{15-42}$$

umlaut

This approach is applied to an analysis of mortgage rates in Example 11.20. [Plümper and Troeger's (2007) FEVD estimator examined in Section 11.4.5 is essentially this model as well.]

A second approach to random parameters modeling builds from the crucial assumption added to (15-42) that $\mathbf{u}_i$ and $\mathbf{X}_i$ are uncorrelated. The general model is defined in terms of the conditional density of the random variable, $f(y_{it}|\mathbf{x}_{it},\boldsymbol{\beta}_i,\boldsymbol{\theta})$ and the marginal density of the random coefficients, $f(\boldsymbol{\beta}_i|\mathbf{z}_i,\boldsymbol{\Omega})$ in which $\boldsymbol{\Omega}$ is the separate parameters of this distribution. This leads to the mixed models examined in this chapter. The random effects model that we examined in Section 11.5 and several other points is a special case in which only the constant term is random (like the fixed effects model). We also considered the specific case in which u_i is distributed normally with variance σ_u^2.

A third approach to modeling heterogeneity in parametric models is to use a discrete distribution, either as an approximation to an underlying continuous distribution, or as the model of the data generating process in its own right. (See Section 14.10.) This model adds to the preceding a noparametric specification of the variation in $\boldsymbol{\beta}_i$,

non

$$\text{Prob}(\boldsymbol{\beta}_i = \boldsymbol{\beta}_j\,|\mathbf{z}_i) = \pi_j,\ j = 1,\ldots,J.$$

A somewhat richer, semiparametric form that mimics (15-42) is

$$\text{Prob}(\boldsymbol{\beta}_i = \boldsymbol{\beta}_j\,|\mathbf{z}_i) = \pi_j(\mathbf{z}_i,\boldsymbol{\Omega}),\ j = 1,\ldots,J.$$

We continue to assume that the process generating variation in $\boldsymbol{\beta}_i$ across individuals is independent of the process that produces $\mathbf{X}_i$ – that is, in a broad sense, we retain the random effects approach. This latent class model is gaining popularity in the current literature. In the last example of this chapter, we will examine a comparison of mixed and finite mixture models for a nonlinear model.

15-54

AQ: Confirm footnote renumbering

Example 15.17 Maximum Simulated Likelihood Estimation of a Binary Choice Model

Bertschek and Lechner (1998) analyzed the product innovations of a sample of German manufacturing firms. They used a probit model (Sections 17.2–17.4) to study firm innovations. The model is for Prob[y_{it} = 1|$\mathbf{x}_{it}$,$\boldsymbol{\beta}_i$] where

y_{it} = 1 if firm i realized a product innovation in year t and 0 if not.

The independent variables in the model are

$x_{it,1}$ = constant,
$x_{it,2}$ = log of sales,
$x_{it,3}$ = relative size = ratio of employment in business unit to employment in the industry,
$x_{it,4}$ = ratio of industry imports to (industry sales + imports),
$x_{it,5}$ = ratio of industry foreign direct investment to (industry sales + imports),
$x_{it,6}$ = productivity = ratio of industry value added to industry employment,
$x_{it,7}$ = dummy variable indicating firm is in the raw materials sector,
$x_{it,8}$ = dummy variable indicating firm is in the investment goods sector.

The sample consists of 1,270 German firms observed for five years, 1984–1988. (See Appendix Table F15.1.) The density that enters the log-likelihood is

$$f(y_{it} \mid \mathbf{x}_{it},\boldsymbol{\beta}_i) = \text{Prob}[y_{it} \mid \mathbf{x}_{it}'\boldsymbol{\beta}_i] = \Phi[(2y_{it} - 1)\mathbf{x}_{it}'\boldsymbol{\beta}_i],\ y_{it} = 0, 1,$$

where

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{v}_i,\ \mathbf{v}_i \sim N[\mathbf{0}, \boldsymbol{\Sigma}].$$

To be consistent with Bertschek and Lechner (1998) we did not fit any firm specific, time invariant components in the main equation for β_i.[10] Table 15.9 presents the estimated coefficients for the basic probit model in the first column. These are the values reported in the 1998 study. The estimates of the means, $\boldsymbol{\beta}$, are shown in the second column. There appear to be large differences in the parameter estimates, although this can be misleading as there is large variation across the firms in the posterior estimates. The third column presents the square roots of the implied diagonal elements of $\boldsymbol{\Sigma}$ computed as the diagonal elements of $\mathbf{CC}'$. These estimated standard deviations are for the underlying distribution of the parameter in the model—they are not estimates of the standard deviation of the sampling distribution of the estimator. That is shown for the mean parameter in the second column. The fourth column presents the sample means and standard deviations of the 1,270 estimated conditional estimates of the coefficients.

The latent class formulation developed in Section 14.10 provides an alternative approach for modeling latent parameter heterogeneity.[11] To illustrate the specification, we will reestimate the random parameters innovation model using a three-class latent class model. Estimates of the model parameters are presented in Table 15.10. The estimated conditional mean shown, which is comparable to the empirical means in the rightmost column in Table 17.4 for the random parameters model, are the sample average and standard deviation of the 1,270 firm-specific posterior mean parameter vectors. They are computed using $\hat{\boldsymbol{\beta}}_i = \Sigma_{j=1}^{3}\hat{\pi}_{ij}\hat{\boldsymbol{\beta}}_j$ where $\hat{\pi}_{ij}$ is the conditional estimator of the class probabilities in (14-102). These estimates differ considerably from the probit model, but they are quite similar to the empirical means in Table 15.9. In each case, a confidence interval around the posterior mean contains the one-class, pooled probit estimator. Finally, the (identical) prior and average of the sample posterior class probabilities are shown at the bottom of the table. The much larger empirical standard deviations reflect that the posterior estimates are based on aggregating the sample data and involve, as well, complicated functions of all the model parameters. The estimated numbers of class members are computed by assigning to each firm the predicted class associated with the highest posterior class probability.

[10] Apparently they did not use the second derivatives to compute the standard errors—we could not replicate these. Those shown in the Table 15.9 are our results.

[11] See Greene (2001) for a survey. For two examples, Nagin and Land (1993) employed the model to study age transitions through stages of criminal careers and Wang et al. (1998) and Wedel et al. (1993) and used the Poisson regression model to study counts of patents.

15-55

TABLE 15.9 Estimated Random Parameters Model

	Probit	*RP Mean*	*RP Std. Dev.*	*Empirical Distn.*
Constant	−1.96	−3.91	2.70	−3.27
	(0.23)	(0.20)		(0.57)
ln Sales	0.18	0.36	0.28	0.32
	(0.022)	(0.019)		(0.15)
Relative Size	1.07	6.01	5.99	3.33
	(0.14)	(0.22)		(2.25)
Import	1.13	1.51	0.84	2.01
	(0.15)	(0.13)		(0.58)
FDI	2.85	3.81	6.51	3.76
	(0.40)	(0.33)		(1.69)
Productivity	−2.34	−5.10	13.03	−8.15
	(0.72)	(0.73)		(8.29)
Raw Materials	−0.28	−0.31	1.65	−0.18
	(0.081)	(0.075)		(0.57)
Investment	0.19	0.27	1.42	0.27
	(0.039)	(0.032)		(0.38)
ln L	−4114.05		−3498.654	

TABLE 15.10 Estimated Latent Class Model

	Class 1	*Class 2*	*Class 3*	*Posterior*
Constant	−2.32	−2.71	−8.97	−3.38
	(0.59)	(0.69)	(2.20)	(2.14)
ln Sales	0.32	0.23	0.57	0.34
	(0.061)	(0.072)	(0.18)	(0.09)
Relative Size	4.38	0.72	1.42	2.58
	(0.89)	(0.37)	(0.76)	(1.30)
Import	0.94	2.26	3.12	1.81
	(0.37)	(0.53)	(1.38)	(0.74)
FDI	2.20	2.81	8.37	3.63
	(1.16)	(1.11)	(1.93)	(1.98)
Productivity	−5.86	−7.70	−0.91	−5.48
	(2.70)	(4.69)	(6.76)	(1.78)
Raw Materials	−0.11	−0.60	0.86	−0.08
	(0.24)	(0.42)	(0.70)	(0.37)
Investment	0.13	0.41	0.47	0.29
	(0.11)	(0.12)	(0.26)	(0.13)
ln L		−3503.55		
Class Prob (Prior)	0.469	0.331	0.200	
	(0.0352)	(0.0333)	(0.0246)	
Class Prob (Posterior)	0.469	0.331	0.200	
	(0.394)	(0.289)	(0.325)	
Pred. Count	649	366	255	

15-56

15.12

~~15.7~~ **SUMMARY AND CONCLUSIONS**

This chapter has outlined several applications of simulation assisted estimation and inference. The essential ingredient in any of these applications is a random number generator. We examined the most common method of generating what appear to be samples of random draws from a population—in fact, they are deterministic Markov chains that only appear to be random. Random number generators are used directly to obtain draws from the standard uniform distribution. The inverse probability transformation is then used to transform these to draws from other distributions. We examined several major applications involving random sampling:

• Random sampling, in the form of bootstrapping, allows us to infer the characteristics of the sampling distribution of an estimator, in particular its asymptotic variance. We used this result to examine the sampling variance of the median in random sampling from a nonnormal population. Bootstrapping is also a useful, robust method of constructing confidence intervals for parameters.

• Monte Carlo studies are used to examine the behavior of statistics when the precise sampling distribution of the statistic cannot be derived. We examined the behavior of a certain test statistic and of the maximum likelihood estimator in a fixed effects model.

• Many integrals that do not have closed forms can be transformed into expectations of random variables that can be sampled with a random number generator. This produces the technique of Monte Carlo integration. The technique of maximum simulated likelihood estimation allows the researcher to formulate likelihood functions (and other criteria such as moment equations) that involve expectations that can be integrated out of the function using Monte Carlo techniques. We used the method to fit random parameters models.

The techniques suggested here open up a vast range of applications of Bayesian statistics and econometrics in which the characteristics of a posterior distribution are deduced from random samples from the distribution, rather than brute force derivation of the analytic form. Bayesian methods based on this principle are discussed in the next chapter.

15-57

End 15

Insert msp 15-57A here

Key Terms and Concepts

- Bootstrapping
- Cholesky factorization
- Fundamental probability transformation
- GHK simulator
- Halton sequence
- Importance function
- Incidental parameters problem
- Markov chain
- Maximum simulated likelihood
- Monte Carlo integration
- Monte Carlo study
- Period
- Power of a test
- Pseudo–random number generator
- Random parameters
- Seed
- Simulation
- Size of a test
- Specificity
- Shuffling
- Discrete uniform distribution
- Gauss–Hermite quadrature
- Pseudo maximum likelihood estimator
- Paired bootstrap
- Parametric bootstrap
- Nonparametric bootstrap
- Block bootstrap
- Antithetic draws
- Hadamard product
- Schur product
- Direct product
- Kronecker product
- Mixed model
- Hierarchical linear model
- Latent class model

Exercises

1. The exponential distribution has density $f(x) = \theta\exp(-\theta x)$. How would you obtain a random sample of observations from an exponential population?

2. TheWeibull population has survival function $S(x) = \lambda p\ \exp(-(\lambda x)p)$. How would you obtain a random sample of observations from a Weibull population? (The survival function equals one minus the cdf.)

3. Suppose x and y are bivariate normally distributed with zero means, variances equal to one and correlation equal to ρ. Show how to use a Gibbs sampler to estimate $E[x2\exp(y) + y2\exp(x)]$.

4. Derive the first order conditions for nonlinear least squares estimation of the parameters in (15-2). How would you estimate the asymptotic covariance matrix for your estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta},\sigma)$?

Application

1. Does the Wald statistic reject the null too often? Construct a Monte Carlo study of the behavior of the Wald statistic for testing the hypothesis that γ equals zero in the model of Section 17.4.1. Recall, theWald statistic is the square of the t ratio on the parameter in question. The procedure of the test is to reject the null hypothesis if theWald statistic is greater than 3.84, the critical value from the chi squared distribution with one degree of freedom. Replicate the study in Section 17.4.1, that is for all three assumptions about the underlying data.

2. A regression model that describes income as a function of experience is

$$\ln \mathit{Income}_i = \beta_1 + \beta_2\, \mathit{Experience}_i + \beta_3\, \mathit{Experience}_i^2 + \varepsilon_i.$$

The model implies that ln *Income* is largest when $\partial \ln \mathit{Income}/\partial \mathit{Experience}$ equals zero. The value of *Experience* at which this occurs is where $\beta_4 + 2\beta_5 \mathit{Experience} = 0$, or $\mathit{Experience}^* = -\beta_2/\beta_3$. Describe how to use the delta method to obtain a confidence interval for *Experience**. Now, describe how to use bootstrapping for this computation. A model of this sort using the Cornwell and Rupert data appears in Example 15.6. Using your proposals here, carry out the computations for that model, using the Cornwell and Rupert data.

minus

Insert on ms p 15-57 where indicated

15-57A

AU: Terms with blue check were not bold KTs in chapter

Antithetic draws
Block bootstrap
Bootstrapping
Cholesky factorization
Direct product
Discrete uniform distribution
Fundamental probability transformation
Gauss–Hermite quadrature
GHK smooth recursive stimulator
Hadamard product
✓Halton sequence
Hierarchical linear model
✓Importance function
Incidental parameters problem
Kronecker product
✓Latent class model
Markov chain
✓Maximum stimulated likelihood
Mixed model
Monte Carlo integration
✓Monte Carlo study
Nonparametric bootstrap
Paired bootstrap
Parametric bootstrap
Period
Power of a test
Pseudo-maximum likelihood estimator
Pseudo-random number generator
✓Random parameters
Schur product
Seed
✓Simulation
Size of a test
Specificity
Shuffling

end of insert