Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models

William Greene*

Department of Economics, Stern School of Business,
New York University,

September, 2003

Abstract

Recent studies in econometrics and statistics include many applications of random parameter models. There is some ambiguity in how estimation results in these models are interpreted. The underlying structural parameters are often not informative about the statistical relationship of interest. As a result, standard significance tests of structural parameters in random parameter models do not necessarily indicate the presence or absence of a 'significant' relationship among the model variables. This note offers some suggestions on how to interpret and use the results of estimation of a general form of random parameter model and how simulation based estimates of parameters in conditional distributions can be used to examine the influence of model covariates.

Keywords: Panel data, random effects, random parameters, maximum simulated likelihood, posterior mean, posterior variance, marginal effects, confidence interval

JEL classification: C1, C4

* 44 West 4th St., New York, NY 10012, USA, Telephone: 001-212-998-0876; fax: 01-212-995-4218; e-mail: wgreene@stern.nyu.edu, URL www.stern.nyu.edu/~wgreene.

1. Introduction

The increasing availability of large, high quality panel data sets has made models that can accommodate individual heterogeneity, such as the random parameters (RP) model, increasingly attractive. The hierarchical nature of this model's parameterization makes ambiguous the computation of certain results, such as marginal effects, that are usually of interest. Moreover, since the models are typically nonlinear and the 'parameters' are random variables, not fixed estimated quantities, assessments of statistical significance, also of common interest, cannot be assessed by the usual approach. (The model shares this aspect with Bayesian formulations of econometric models that are also becoming increasingly common.) This note will lay out a generic form of the random parameters model, then suggest computations that can be used to display empirical evidence on these two model characteristics.

Section 2 defines a broad class of random parameter models that encompasses many of the received applications. Section 3 discusses simulation based estimation of the model and interesting "post estimation" results. An application is provided to illustrate the computations. A comparison to a similar set of procedures based on Allenby and Rossi's (1999) Bayesian formulation is discussed in Section 4. Some conclusions are drawn in Section 5.

2. Random Parameters Models

We consider a formulation of the RP model which encompasses many of the applications in the literature [e.g. Revelt and Train (1998), Train (2002), Layton and Brown (2000), Allenby and Rossi (1999), and Greene (2003b)] and implementations such as Stata (2002), SAS (2003), LIMDEP (Econometric Software, 2003) and MLWin (2002)]. We formulate the RP models in terms of the likelihood function for a set of observations collected in a panel data setting (though the model can be applied in cross section data, so the notation is only used to achieve a greater level of generality). The following will sketch the procedures. More detailed treatments may be found in Train (2002) and in Greene (2003a).

The density for an observation is

$$(1) \qquad f(y_{it} \mid x_{it}, z_i, v_i, \beta_i, \theta) \;=\; g(y_{it}, \beta_i' x_{it}, \theta),\; i = 1,\ldots,n,\; t = 1,\ldots,T_i \geq 1,$$

$$(2) \qquad \beta_i \;=\; \beta \;+\; \Delta z_i \;+\; \Gamma v_i$$

where the components of the model are as follows: x_{it} contains all the main covariates, including both time varying and time invariant variables, z_i is a set of time invariant variables that enter the mean of the random parameters, y_{it} is the response variable, v_i is the random variation in the reduced form parameters of the model.[1] The structural parameters include a fixed set of ancillary parameters, θ, which would include, e.g., the standard deviation of the disturbance, σ, in a tobit or linear regression model, the shape

[1] We will assume continuous variation of the parameters. The finite mixtures (latent classes) approach to modeling heterogeneity provides an attractive alternative, but this is considered elsewhere. See,. e.g., Allenby and Rossi (1999), Hagenaars and McCutcheon (2001), Greene (2003a).

parameters in a gamma regression model, or the dispersion parameter in a negative binomial model. (Note that θ might be null, e.g., in a binary choice model, such as probit or logit.) The remaining structural parameters of the model are β, Δ, and Γ which define the random parameters; β is the vector of constant terms in the means of the random parameters, Δ is a matrix of parameters that multiply the covariates in the distribution of the random parameters, and Γ is an unrestricted lower triangular matrix, also to be estimated. Since the nonzero elements of Γ are free parameters, no generality is lost (indeed identification requires it) by assuming that v_i has mean vector zero and diagonal covariance matrix with no unknown parameters. For example, if random parameters are assumed to be normally distributed, then Var[v_i] = I. The conditional variance of β_i is $\Gamma\Gamma'$, which is unrestricted. 'Fixed' (nonrandom) parameters in the model are specified simply by constraining corresponding rows in Δ and Γ to equal zero. (See the application below.) A 'hierarchical' model with interactions is obtained when rows of Γ are constrained to equal zero while Δ remains unrestricted. The more familiar, simple random coefficients model is obtained when Δ is a zero matrix. The 'random effects' model results if Δ is zero and only the overall constant term in the model is random. Many other permutations of the model can be cast in this framework through suitable modifications of the parameters and/or stochastic specifications. The density can be specified to accommodate many cases of interest to practitioners, such as static probit and logit models, models for counts, linear regression models, duration models, and many others. [See ESI (2003)]. Note, finally, that the simulation estimation method described here allows the distribution of v_i to be nonnormal (indeed, it may be any type of random variable for which random draws can be simulated).

Observations are conditionally (on the random effects) independent. Dependence of the T_i observations for a particular individual results from the common, invariant v_i. Conditioned on v_i, the contribution of the observations from individual (or 'group') i is the joint density,

$$(3) \qquad L_i(\Lambda \mid \mathbf{y}_i, \mathbf{X}_i, z_i, v_i) = \prod_{t=1}^{T_i} g(y_{it}, \beta_i' x_{it}, \theta)$$

where Λ is the full set of structural parameters, [β, Δ, Γ, θ]. In order to estimate the parameters, it is necessary to operate on the unconditional likelihood – the unobserved random term v_i must be integrated out. The contribution of observation i to the unconditional likelihood is

$$(4) \qquad \begin{aligned} L_i(\Lambda \mid \mathbf{y}_i, \mathbf{X}_i, z_i) &= {}_{v_i} \prod_{t=1}^{T_i} g(y_{it}, \beta_i' x_{it}, \theta) f(v_i) dv_i \\ &= {}_{\beta_i} \prod_{t=1}^{T_i} g(y_{it}, \beta_i' x_{it}, \theta) f(\beta_i) d\beta_i . \end{aligned}$$

The integrals will not exist in closed form, but since they are of the form of expectations, they can be estimated by simulation, instead. The simulated log likelihood is

$$(5) \qquad \log L_S = {}_{i=1}^{n} \log \frac{1}{R} {}_{r=1}^{R} \prod_{t=1}^{T_i} g(y_{it}, \beta_{i,r}' x_{it}, \theta).$$

Note that the simulation is over R draws on $v_{i,r}$ through $\beta_{i,r}$ as defined in (2). The maximum simulated likelihood estimator is obtained by maximizing (5) over the full set of structural parameters, Λ. (The relevant theory for this class of estimators can be found elsewhere, including Train (2002), Greene (2003a) and Gourieroux and Monfort (1996). Estimates of the structural parameters and estimates of their asymptotic standard errors are based on maximization of (5).

3. Simulation Based Estimation of Parameters with an Application to American Movies

There is a possible ambiguity in how one should interpret the estimated parameters in a random parameters model. To illustrate the point, we consider a recent application in marketing, Craig, Douglas and Greene (2003) which modeled the foreign box office receipts of 299 movies released in 8 countries over a 6 year period. The structural equations of the model are

$$
\begin{aligned}
LogSales_{i,c} &= \alpha_i + \beta_i\, LogSales_{i,US} + \tau_1 CD_c + \tau_2 McD_c + \tau_3 E_c \\
&\quad + \Sigma_g\, \lambda_g Genre_{i,g} + \Sigma_t\, \phi_t\, D_t + \varepsilon_{i,c} \\
\alpha_i &= \alpha + \delta_\alpha\, logIncome_c + \gamma_\alpha v_{i,\alpha} \\
\beta_\mathrm{i} &= \beta + \delta_\beta\, logIncome_c + \gamma_\beta v_{i,\beta}
\end{aligned} \tag{6}
$$

where 'c' indicates the country (UK, Australia, Germany, Austria, Argentina, Chile, and Mexico), 'i' indicates the film, $i = 1,...,299$, the box office sales and income figures are in per capita terms, $Genre_{i,g}$ are twelve (after one is dropped) dummy variables indicating which of 13 genres characterizes the film, D_t, $t = 1997,...,2002$, are year dummy variables (the first is dropped), E_c is a dummy variable for the English speaking countries, and two variables, cultural distance, CD_c, and McDonald's restaurants per capita, McD_c are used to measure cultural similarity to the U.S. (Details on the data may be found in the cited paper.)

In the hierarchical model, we have

$$\beta_i \,|\, Income_c = \beta + \delta logIncome_c + \sigma_\beta v_{i,\beta}, \text{ where } v_{i,\beta} \sim \mathrm{N}[0,1]. \tag{7}$$

We seek to ascertain whether U.S. box office sales are a 'significant' determinant of foreign box office sales for a film and, ultimately, to obtain film specific estimates of the parameters, β_i. A finding that estimates of β and δ are 'significant' does not imply that the random coefficient on the left is, in total, correspondingly so. Large variation due to the normally distributed component, $\sigma_\beta v_{i,\beta}$, might dominate the random parameter. (Nor, it turns out, does lack of significance of these coefficients necessarily imply the absence of a 'significant' relationship.) Computing $\tilde{\beta}_i = \hat{\beta} + \hat{\delta}_\beta logIncome_c$ will likewise be uninformative since there is no film specific information in the estimate – this is merely an estimate of the common prior mean. More generally, in the model in (1)-(2), this computation would correspond to some kind of average individual with these specific characteristics, rather than this specific individual. [See Train (2002, Chapter 10, for discussion of this distinction.]

For each film, we can estimate the posterior mean, $E[\beta_i \mid logSales_{i,c}, \mathbf{x}_{i,c}, z_c,\ c = 1,\ldots,8]$, using Bayes theorem, where z_c is log per capita income and $\mathbf{x}_{f,c}$ is all other variables in the model including the log per capita U.S. box office. The empirical distribution of the estimated film specific estimates will then suggest whether the results document a systematic relationship between U.S. and local box office receipts.

For convenience, let $\mathbf{y}_i$ denote the observations on the local box office for this film for all countries for which it is observed, and let $\mathbf{X}_i$ denote the observations on all other variables for this film, including US box office, genre, McD, log of per capita income, and so on, again for all countries. Then, the posterior mean for the specific film is

$$(8) \qquad E[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i] = \int_{\beta_i} \beta_f\, f(\beta_i \mid \mathbf{y}_i, \mathbf{X}_i)\, d\beta_i$$

where $f(\beta_i|\mathbf{y}_i,\mathbf{X}_i)$ is the conditional (posterior) density of β_i given all the information available in the sample on this film. This conditional distribution is constructed using Bayes theorem as follows:

$$f(\beta_i|\mathbf{y}_i,\mathbf{X}_i) = \frac{f(\beta_i, \mathbf{y}_i \mid \mathbf{X}_i)}{f(\mathbf{y}_i \mid \mathbf{X}_i)}$$

$$(9) \qquad = \frac{f(\mathbf{y}_i \mid \beta_i, \mathbf{X}_i)\, f(\beta_i \mid \mathbf{X}_i)}{\int_{\beta_i} f(\mathbf{y}_i \mid \beta_i, \mathbf{X}_i)\, f(\beta_i \mid \mathbf{X}_i)\, d\beta_i}.$$

The joint density in the numerator is the product of the marginal distribution of β_i, which is the normal distribution defined by (6), and the conditional distribution of the dependent variable given the parameter β_i, which is the term in the likelihood function before the integration in (4). The denominator is the marginal distribution of $\mathbf{y}_i$ obtained by integrating β_i out of the joint distribution. The conditional mean of this distribution is then obtained by the definition,

$$(10) \qquad E[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i] = \frac{\int_{\beta_i} \beta_i\, f(\mathbf{y}_i \mid \beta_i, \mathbf{X}_i)\, f(\beta_i \mid \mathbf{X}_i)\, d\beta_i}{\int_{\beta_i} f(\mathbf{y}_i \mid \beta_i, \mathbf{X}_i)\, f(\beta_i \mid \mathbf{X}_i)\, d\beta_i}.$$

In order to estimate the conditional mean in (10), we would insert the estimated parameters for the remainder of the model in the likelihood function and the marginal density of β_i, then compute the integrals. However, the integrals will not exist in closed form. They can be computed by simulation, by the same method used to compute the simulated likelihood earlier. The simulation estimator of the posterior mean is, then

$$(11) \qquad \hat{E}[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i] = \frac{(1/R)\sum_{r=1}^{R} \hat{\beta}_{i,r}\, \hat{L}(\mathbf{y}_i, \mathbf{X}_i, v_{i,\beta,r})}{(1/R)\sum_{r=1}^{R} \hat{L}(\mathbf{y}_i, \mathbf{X}_i, v_{i,\beta,r})}$$

where $\hat{L}(\mathbf{y}_i, \mathbf{X}_i, v_{i,\beta,r})$ is the contribution to the likelihood function (not its log) of film i evaluated at all the estimated parameters and the rth simulated value, $\hat{\beta}_{i,r} = \hat{\beta} + \hat{\delta}\log Income_c + \hat{\sigma}_\beta v_{i,\beta,r}$. Note that the simulation is over the draws of $v_{i,\beta,r}$. (Also, we note that the random constant term in the model is also simulated.) In the results below, we will also make use of the estimated posterior variance of β_i, $Var[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i] = E[\beta_i^2|\mathbf{y}_i,\mathbf{X}_i] - (E[\beta_i|\mathbf{y}_i,\mathbf{X}_i])^2$. This is estimated in the same fashion by first estimating the posterior expected square, with

$$(12) \qquad \hat{E}[\beta_i^2 \mid \mathbf{y}_i, \mathbf{X}_i] = \frac{(1/R)\sum_{r=1}^{R}\hat{\beta}_{i,r}^2 \hat{L}(\mathbf{y}_i, \mathbf{X}_i, v_{i,\beta,r})}{(1/R)\sum_{r=1}^{R}\hat{L}(\mathbf{y}_i, \mathbf{X}_i, v_{i,\beta,r})}.$$

The standard deviation of the posterior distribution is estimated with

$$(13) \qquad S.D.[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i] = \sqrt{\hat{E}[\beta_i^2 \mid \mathbf{y}_i, \mathbf{X}_i] - \left(\hat{E}[\beta_i \mid \mathbf{y}_i, \mathbf{X}_i]\right)^2}.$$

Table 1 (reproduced from Craig, Douglas and Greene (2003) presents the maximum simulated likelihood estimates for the model. The estimates in the first column are the nonrandom parameters counterparts to the RPM in (6). This model is a linear regression model, where (2) implies that logIncome and an interaction with logUSBox will also appear in the model. The second column gives the least squares estimates of (6) ignoring the random parameters specification, including *logIncome*. (Note the zero 'income effect.') The third column presents the maximum simulated likelihood estimates of the full RP model. Based on the likelihood ratio test, the hypothesis of homogeneity of the model coefficients is soundly rejected; χ^2 = 2(3154.532 - 3007.729) = 293.606 with 2 degrees of freedom. As to whether US Box office results significantly effect foreign sales, the least squares results strongly suggest so – the raw coefficient (1.131 in the presence of the income effect, 1.204 without it) appears to be strongly 'significant' with t ratios well in excess of 4.0. But, in the random parameters model, the simple t ratio on the estimate of β is 1.82, suggesting a much weaker conclusion.

Figure 1 below shows for each of the 299 films the range given by the estimated posterior mean plus and minus 2.5 posterior standard deviations. With conditional normality, this range would encompass over 99% of the mass of each posterior distribution. Since the posterior distributions are not necessarily normal or symmetric, the actual mass may be slightly less than this, but will be more than 95%. The horizontal lines in the figure are drawn at the sample mean of the 299 estimated conditional means (1.21) and at zero. The dots in the centers of the bars show the film specific point estimates, the posterior means in (11). The vertical bars divide the data into the six years of observations. Only two of these 299 intervals include zero, and those only slightly. We conclude that the relationship between US and foreign box office is indeed, positive and significant.

Table 1. Estimated Regressions for Log Per Capita Box Office, All Countries (Estimated Standard Errors in Parentheses)[a]

	Fixed Parameters Models (OLS)		Random Parameters Model (MSL)
Variable	Fixed Parameters		
Cultural Dist.	-0.192 (0.168)**	-0.155 (0.017)**	-0.156 (0.010)**
Macs Per Capita	0.057 (0.004)**	0.040 (0.004)**	0.040 (0.003)**
English Lang.	-0.260 (0.083)**	0.120 (0.079)	0.138 (0.052)**
Drama	-0.138 (0.137)	-0.130 (0.142)	-0.138 (0.086)
Romance	0.008 (0.201)	0.022 (0.207)	0.042 (0.127)
Comedy	-0.165 (0.134)	-0.151 (0.138)	-0.140 (0.085)
Action	0.110 (0.133)	0.118 (0.137)	0.092 (0.084)
Fantasy	0.509 (0.185)**	0.526 (0.191)**	0.559 (0.123)**
Adventure	0.108 (0.156)	0.110 (0.161)	0.141 (0.099)
Family	-0.476 (0.160)**	-0.498 (0.165)**	-0.591 (0.092)**
Animated	0.148 (0.160)	0.152 (0.165)	0.149 (0.103)
Thriller	-0.049 (0.170)	-0.033 (0.176)	-0.074 (0.106)
Mystery	0.403 (0.295)	0.383 (0.304)	0.227 (0.198)
Science Fiction	0.042 (0.173)	0.039 (0.179)	0.009 (0.112)
Horror	0.172 (0.160)	0.157 (0.165)	0.058 (0.112)
Year 1998	-0.237 (0.081)**	-0.314 (0.084)**	-0.566 (0.050)**
Year 1999	-0.160 (0.084)*	-0.245 (0.087)**	-0.445 (0.050)**
Year 2000	-0.332 (0.083)**	-0.410 (0.085)**	-0.587 (0.051)**
Year 2001	-0.339 (0.083)**	-0.419 (0.085)**	-0.640 (0.052)**
Year 2002	-0.619 (0.084)**	-0.701 (0.087)**	-0.929 (0.050)**
	Random Parameters		
	Constant		
Intercept	0.556 (0.358)	-0.975 (0.183)**	-3.649 (1.313)**
Income Effect	-0.250 (0.048)**	0.000 (0.000)	0.359 (0.164)**
Std. Deviation	0.000	0.000	0.177
	PerCapita US Box		
Intercept	1.131 (0.245)**	1.204 (0.047)**	1.669 (0.917)*
Income Effect	0.011 (0.037)	0.000[b] (0.000)	-0.057 (0.114)
Std. Deviation	0.000[b]	0.000[b]	0.369
			Γ_{21} = 0.000[b]
Disturbance S.D.	1.022	1.053	0.934
Log Likelihood	-3154.532	-3220.691	-3007.729
R^2	0.505	0.475	

[a]*(**) Indicates significant at 95% (99%) significance level.
[b]Fixed at this value.

In principle, one could estimate the posterior, conditional means and variances for other functions of the model parameters. Discrete choice analysis of consumer preferences provides an important example. It is common in the discrete choice literature to use the model parameters to estimate 'willingness to pay' values. This is computed as the ratio of a quantity coefficient to a price coefficient. For example, Layton and Brown (2000) examine a stated preference survey over programs for mitigating forest loss due to global climate change in the context of a random parameters (mixed) multinomial logit model. Estimates of the RP model include a price coefficient and coefficients on three levels of forest loss under the programs. Discussion of results considers significance levels of structural parameters – their model corresponds to (1)-(2) with $\Delta = 0$ and some zeros placed in Γ. The basic reported results corresponding to our Table 1 include for one of the models, a price coefficient of -.1185 and coefficients on three programs corresponding to 2,500 ft. loss, 1,200 ft. and 600 ft. of –11.1871, -5.1586 and –1.9483, respectively. Discussion of willingness to pay estimates report only the three ratios of –0.1185 to these values as the point estimates, and overall confidence intervals. In precisely this application, one could, instead, compute

$$(12)\qquad \hat{E}[WTP_{program} \mid data_i] = \frac{(1/R)\sum_{r=1}^{R}\left(\hat{\beta}_{program,i}/\hat{\beta}_{price,i}\right)\hat{L}(data_{i,r})}{(1/R)\sum_{r=1}^{R}\hat{L}(data_{i,r})}$$

The conditional standard deviation could be computed in the same fashion See Hensher, Greene and Rose (2003) for another application.

This computations described here could also be applied to other more involved functions in a model. For example, the marginal effects in binary choice models or in models for counts such as the Poisson regression model are complicated nonlinear functions of all the model parameters. One could estimate these for each individual in a sample. For individual i in a panel probit model, the marginal effect of an x_{ik}, evaluated at the individual's mean values would be $\delta_i = \beta_{k,i}\ \phi(\beta_i'\overline{x}_i)$, a quantity which can be computed by simulation in the same fashion as β_i in the WTP above.

4. Bayesian Analysis of Individual Heterogeneity

Allenby and Rossi (1999) have studied the issue discussed here at length in the context of a discrete choice, brand choice model. Their model framework is different from our application, but is encompassed in the generic form given earlier. They focused on a Bayesian approach while suggesting that the classical form of these computations was likely to be extremely cumbersome. The preceding suggests that in fact, that is not actually the case. (We note that the platform for their analysis was a multinomial probit (MNP) model, which they fit with Markov Chain Monte Carlo Methods. They are certainly correct that simulation based estimation of the MNP model is cumbersome in the extreme, and has not advanced beyond quite moderate sized applications. As Train (2003) has analyzed extensively, this limitation is relaxed in several attractive directions by the mixed logit model, which is in fact, more flexible. Thus, we will eschew conclusions about the model in particular, and consider the generic issues that they raised.)

The main focus of Allenby and Rossi (1999) is the estimation of household (individual) level parameters. In the Bayesian context, this is achieved by estimation of the posterior means of β_i, as suggested earlier. The posterior distributions follow from the assumed priors (normal in their cases) and are easily drawn or analyzed (as in their figure 1.) Regarding a classical approach, they suggest (in their equation (20)) precisely the calculation suggested in our (11), but observe that the calculation is "substantially more computationally demanding than the full Bayesian approach and offers only approximate answers." The preceding suggests that the first of these is overstated. The computations described here are straightforward to apply (we used the built in routines in LIMDEP 8.0). It is true that the Bayesian 'estimates' are exact, but only as implied by the assumed priors. While they, themselves may appear to be diffuse, the influence of the assumed distributions remains a consideration. Also, it should be remembered that the

Gibbs sampler is a simulation based estimator, not a window into the true population. Whether the Bayesian posterior means are 'exact' measures of quantities that are only measured approximately by the classical methods is an issue that is at least open to question. The counterparts in the classical framework are the underlying stochastic assumptions, which are parametric and also occasionally controversial, and the use of simulation based estimation to obtain the posterior moments. We do note, though, whether this point/counterpoint is actually substantive may itself be moot. Train (2001) has compiled evidence that the numerical answers that one obtains with hierarchical Bayes and classical 'mixed models' are likely to be essentially the same.

One comparison does remain. In Allenby and Rossi's analysis, the posterior analysis provides a full statement of the posterior distribution, not just its first few moments. (Again, this is derived from the assumed prior). It is not likely that the same information could be obtained from the classical estimates. The prior normal distribution implied by (2) does not imply normality of the posterior – as they note, the shape of the posterior will be influenced by the data. How one could sample directly from the posterior for example, to construct a kernel density estimate, remains to be established. However, again, this may be of limited practical import – the quantities of interest are likely to be the HPD intervals in the Bayesian or the 'confidence' intervals in the classical case, and in either framework, there is no great difficulty in obtaining these.

5. Conclusion

Random parameters (RP) models, also known as hierarchical models, mixed models, and random coefficients models, are enjoying a flowering in the applied as well as theoretical literature. [For a sample in just one area, discrete choice modeling, see Train (2003).] This note has proposed an extension of the existing classical, simulation based techniques that suggests a useful approach to the question of 'significance' as well as to the more general issue of how one can make effective use of the results after estimating an RP model for estimating individual level quantities of interest. The proposed device combines use of the numerical statistical results with a useful graphical summary of the estimates to provide data on individual level heterogeneity as well as more general conclusions about the model's implications for relationships among the component variables.

References

Allenby, G. and P. Rossi "Marketing Models of Consumer Heterogeneity" *Journal of Econometrics*, 89, 1999, pp. 57-78.

Craig, S., S. Douglas and W. Greene, "Culture Matters: A Hierarchical Linear Random Parameters Model for Predicting Success of US Films in Foreign Markets," Manuscript, Department of Marketing, Stern School of Business, NYU, 2003.

Gourieroux, C. and A. Monfort, *Simulation Based Methods: Econometric Methods*, Oxford University Press, Oxford, 1996.

Greene, W., *Econometric Analysis*, Prentice Hall, Englewood Cliffs, 2003a.

Greene, W., "Convenient Estimators for the Panel Probit Model: Further Results," *Empirical Economics*, 2003b, forthcoming.

Hagenaars, J. and A. McCutcheon, *Applied Latent Class Analysis*, Cambridge University Press, Cambridge, 2002.

Hensher, D., W. Greene and Jr. Rose, "Taking Advantage of Priors in Estimation and Posteriors in Application to Reveal Individual-Specific Parameter Estimates and Avoid the Potential Complexities of WTP Derived from Population Moments," Manuscript, Institute of Transport Studies, University of Sydney, 2003.

Layton, D. and G. Brown, "Heterogeneous Preferences Regarding Global Climate Change," *Review of Economics and Statistics*, 82, 4, 1998, pp. 616-624.

McLachlan, G. and D. Peel, *Finite Mixture Models*, John Wiley and Sons, New York, 2000.

LIMDEP Version 8.0, Econometric Software, Inc., Plainview, NY, 2003.

MLWin, Centre for Multilevel Modeling, Bedford, UK, 2003

Revelt D. and K. Train, "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, 80, 1998, pp. 1-11.

SAS, Cary, NC, 2003

Stata, Version 8, College Station, Tx, 2002.

Train, K., "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," Manuscript, Department of Economics, University of California, Berkeley, 2001.

Train, K., *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.

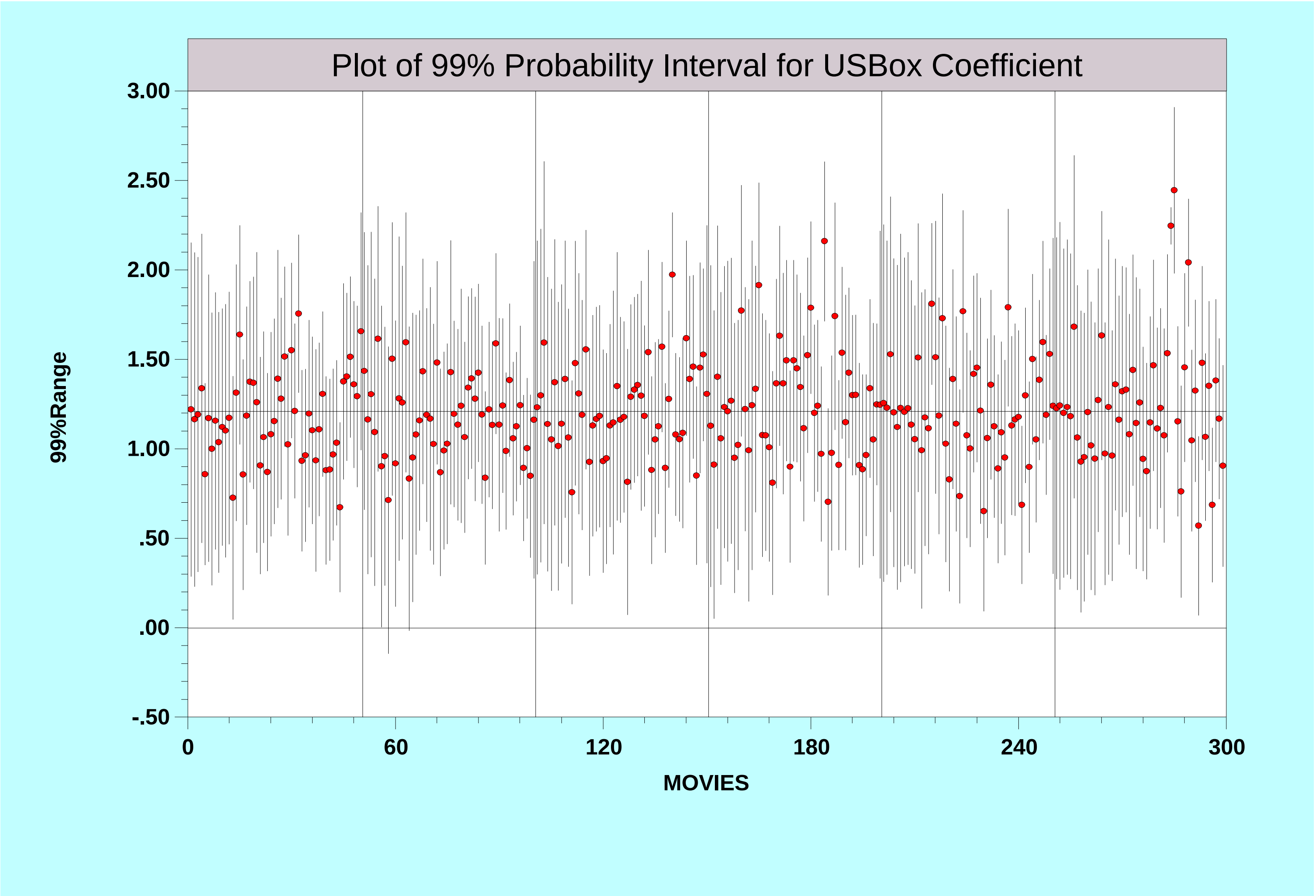

Figure 1. 99% Confidence Intervals for Film Specific Coefficients on logUSBox