Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models

William Greene[*]

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

May 6, 2004

Abstract

Recent studies in econometrics and statistics include many applications of random parameter models. The underlying structural parameters in these models are often not directly 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 a suggestion on how to examine the results of estimation of a general form of random parameter model. We also extend results on computing individual level parameters in a random parameters setting and show how simulation based estimates of parameters in conditional distributions can be used to examine the influence of model covariates (marginal effects) at an individual level

Keywords: Panel data, random effects, random parameters, maximum simulated likelihood, conditional mean, conditional 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 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 often 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 parameters and interesting 'post estimation' results. An application is provided to illustrate the computations. Section 4 extends the results to estimation of marginal effects in a nonlinear model and demonstrates with a second application. A comparison to a similar set of procedures based on Allenby and Rossi's (1999) Bayesian formulation is discussed in Section 5. Some conclusions are drawn in Section 6.

2. Estimating Random Parameters Models

This section will describe a general form of the random parameters model. Many of the received applications of random parameters models have used the linear regression framework, though there is a large and growing literature on nonlinear models with random parameters, which we will consider again later in the paper. To begin, the generic model framework is set up. The mathematical framework will be laid out first, then attention is turned to the specific model parameters.

We consider a formulation of an RP model that encompasses many of the applications in the literature [e.g. Revelt and Train (1998), Train (2003), Layton and Brown (2000), Allenby and Rossi (1999), and Greene (2004)] 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 (2003) and in Greene (2003).

The density for an observation is

(1) $f(y_{it} \mid \mathbf{x}_{it}, \mathbf{z}_i, \mathbf{v}_i, \boldsymbol{\beta}_i, \boldsymbol{\theta}) = g(y_{it}, \boldsymbol{\beta}_i'\mathbf{x}_{it}, \boldsymbol{\theta})$, $i = 1,\ldots,n$, $t = 1,\ldots, T_i \geq 1$,

(2) $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Delta\mathbf{z}_i + \Gamma\mathbf{v}_i$

where $\mathbf{x}_{it}$ contains all the main covariates, including both time varying and time invariant variables, $\mathbf{z}_i$ is a set of time invariant variables that enter the means of the random

parameters and y_{it} is the response variable. The random vector $\mathbf{v}_i$ induces the random variation in the reduced form parameters of the model.[1] The structural parameters can include a set of fixed (nonrandom) parameters, $\boldsymbol{\theta}$, which would include, e.g., the standard deviation of the disturbance, σ, in a tobit or linear regression model, the shape parameters in a gamma regression model, or the dispersion parameter in a negative binomial model. (Note that $\boldsymbol{\theta}$ might be null, e.g., in a binary choice model, such as probit or logit.) The remaining structural parameters of the model are $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, and $\boldsymbol{\Gamma}$ which define the random parameters; $\boldsymbol{\beta}$ is the vector of constant terms in the means of the random parameters, $\boldsymbol{\Delta}$ is a matrix of parameters that multiply the covariates in the distribution of the random parameters, and $\boldsymbol{\Gamma}$ is an unrestricted lower triangular matrix, also to be estimated. Since $\boldsymbol{\beta}$ and the nonzero elements of $\boldsymbol{\Gamma}$ are free parameters, no generality is lost by assuming that $\mathbf{v}_i$ has mean vector zero and diagonal covariance matrix with no unknown parameters. For example, if random parameters are assumed to be normally distributed, then $\text{Var}[\mathbf{v}_i] = \mathbf{I}$. The conditional (on $\mathbf{x}_{it}$ and $\mathbf{z}_i$) variance of $\boldsymbol{\beta}_i$ is $\boldsymbol{\Gamma}\,\boldsymbol{\Gamma}'$, which is an unrestricted full matrix. 'Fixed' (nonrandom) slope parameters in the model may be specified by constraining corresponding rows in $\boldsymbol{\Delta}$ and $\boldsymbol{\Gamma}$ to equal zero. (See the application below.) A 'hierarchical' model with interactions is obtained when rows of $\boldsymbol{\Gamma}$ are constrained to equal zero while $\boldsymbol{\Delta}$ remains unrestricted. A familiar, simple random coefficients model is obtained when $\boldsymbol{\Delta}$ is a zero matrix, so that $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{w}_i$. The 'random effects' model familiar in econometric panel data models results if $\boldsymbol{\Delta}$ 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 LIMDEP (2003).] Finally, the simulation estimation method described here allows the distribution of $\mathbf{v}_i$ to be nonnormal - 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 $\mathbf{v}_i$. Conditioned on $\mathbf{v}_i$, the contribution of the observations from individual (or 'group') i is the joint density,

$$(3) \qquad L_i(\boldsymbol{\Lambda} \mid \mathbf{y}_i, \mathbf{X}_i, \mathbf{z}_i, \mathbf{v}_i) = \prod_{t=1}^{T_i} g(y_{it}, \boldsymbol{\beta}_i' \mathbf{x}_{it}, \boldsymbol{\theta})$$

where $\boldsymbol{\Lambda}$ is the full set of structural parameters, $[\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\theta}]$. In order to estimate these parameters, it is necessary to operate on the unconditional likelihood – the unobserved random term $\mathbf{v}_i$ must be integrated out. The contribution of observation i to the unconditional likelihood is

---

[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 (2003).

$$(4) \quad L_i(\boldsymbol{\Lambda} \mid \mathbf{y}_i, \mathbf{X}_i, \mathbf{z}_i) = \int_{\mathbf{v}_i} \prod_{t=1}^{T_i} g(y_{it}, \boldsymbol{\beta}_i' \mathbf{x}_{it}, \boldsymbol{\theta}) f(\mathbf{v}_i) d\mathbf{v}_i$$

$$= \int_{\boldsymbol{\beta}_i} \prod_{t=1}^{T_i} g(y_{it}, \boldsymbol{\beta}_i' \mathbf{x}_{it}, \boldsymbol{\theta}) f(\boldsymbol{\beta}_i) d\boldsymbol{\beta}_i.$$

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

$$(5) \quad \log L_S = \sum_{i=1}^{n} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} g(y_{it}, \boldsymbol{\beta}_{i,r}' \mathbf{x}_{it}, \boldsymbol{\theta}).$$

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

The preceding encompasses many of the received random parameters models. The linear model with random coefficients results if $g(\ldots)$ is the normal density. Other familiar cases, some of which have already appeared in the literature, such as the random coefficients probit model [Sepanski (2000), Guilkey and Murphy (1993), Greene (2004)] are special cases.

Other applications which formulate the parameter heterogeneity in terms of discrete variation [i.e., latent class models - see, for example, deSarbo et al. (1995)] are similar. Estimation of individual level parameters in latent class models is treated essentially the same as suggested here for continuous variation. [See Greene (2004) and Kamakura and Wedel (2004) for discussion and applications].

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

To illustrate the models and calculations, we will consider two applications. 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 is used to suggest a useful way to examine the coefficients in a random parameters model and to give evidence on statistical significance. An application based on the probit model for binary choice is used to demonstrate the computation of marginal effects in the next section.

The structural equations of the linear RP model examined by Craig et al. are

(6)
(a) $LogSales_{i,c} = \alpha_{i,c} + \beta_{i,c} LogSales_{i,US} + \tau_1 CD_c + \tau_2 McD_c + \tau_3 E_c$
$\qquad + \Sigma_g \lambda_g Genre_{i,g} + \Sigma_t \phi_t D_t + \varepsilon_{i,c}$

(b) $\quad \alpha_{i,c} = \alpha + \delta_\alpha logIncome_c + \gamma_\alpha v_{i,\alpha}$

(c) $\quad \beta_{i,c} = \beta + \delta_\beta logIncome_c + \gamma_\beta v_{i,\beta}$

where '*c*' indicates the country (UK, Australia, Germany, Austria, Argentina, Chile, Spain 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.) Note that the panel nature of the data set arises from the multiple country release of each film, not through observation of the same film at several points in time. Since there is no 'time series' aspect to the model, this grouping of observations produces the same characteristics as the more familiar panel data arrangement.

We note before proceeding that the reduced form of the model obtained by inserting (6.b.) and (6.c.) in (6.a.) becomes a simple linear model with $logIncome_c$ and an interaction variable, $LogSales_{i,US} \times logIncome_c$ added to the equation. In this form, although the compound disturbance, $\varepsilon_{i,c} + \gamma_\alpha v_{i,\alpha} + \gamma_\beta v_{i,\beta} LogSales_{i,US}$, is heteroscedastic, the structural parameters can be estimated consistently, albeit not efficiently, by ordinary least squares. However, in what follows, the structural parameters are of secondary interest; we are interested in estimation of the individual level parameters. Moreover, the consistency of the conventional estimation method applied to the reduced form will not carry over to nonlinear models such as the probit, logit or Poisson regression models.

The 'effect' of interest, the effect of (log) US sales on (log) foreign sales for a particular film (in a particular country) in this model is

$$(7) \qquad \beta_i \,|\, Income_c \quad = \partial E[LogSales_{i,c}|\, logIncome_c \,, v_{i,\beta}]/\partial LogSales_{i,US}$$

$$= \beta + \delta logIncome_c + \sigma_\beta v_{i,\beta},$$

where $v_{i,\beta} \sim N[0,1]$. 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, $\beta_{i,c}$.[2] Obviously, a finding that individual estimates of β and δ are individually 'significant' does not imply that the random coefficient on the left is, in total, correspondingly so. Of course, even neglecting the random parameters aspect of the model, this interpretation of marginal effects in the presence of interactions is a common problem even in conventional linear regression models. [See Wooldridge (2000, pp. 224-231) for commentary.] Moreover, in the present setting, large variation due to the normally distributed random component, $\sigma_\beta v_{i,\beta}$, might dominate the random parameter. Since the 'coefficient' differs across individuals by construction, the meaning of the significance results for individual structural coefficients is somewhat ambiguous. See, for example, Layton and Brown (2000), who fit a random parameters discrete choice model and base their assessment of 'significance' on estimates of the structural

---

[2] In the model formulation in (1) and (2), the heterogeneity in the means of the coefficients, z_i, varies across individuals but is time invariant. In this application, the income variable varies across country, but not across films, so, in fact, the heterogeneity in the parameters is provided largely by the random component, $v_{\beta,i}$, not by z_i.

parameters only. The presence of $\Delta \mathbf{z}_i$ in β_i brings a level of generality that is not typical in the received applications. Most are of the simple form $\beta_i = \beta + \mathbf{w}_i$. It is the additional source of heterogeneity, $\Delta \mathbf{z}_i$, that creates the interaction terms, and this term which makes ambiguous the interpretation of the coefficients alone as marginal effects.

In this application, the computation of (7) is complicated by the presence of the random effect and by our interest in individual level effects. Computing $\tilde{\beta}_{i,c} = \hat{\beta} + \hat{\delta}_\beta logIncome_c$ will be uninformative since there is no film specific information in the estimate – this is merely an estimate of the common unconditional mean. More generally, in the model in (1)-(2), this computation would correspond to some kind of average across films rather than this specific individual. [See Train (2003, Chapter 10, for discussion of this distinction.] For each film, we can estimate the conditional mean, $E[\beta_{i,c} \mid logSales_{i,c}, logSales_{i,US}, \mathbf{x}_{i,c}, \mathbf{z}_c, c = 1,\dots,8]$,[3] 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 conditional mean for the specific film in a particular country is

$$(8) \quad E[\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i] = \int_{\beta_{i,c}} \beta_{i,c} f(\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i) d\beta_{i,c}$$

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

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

The joint density in the numerator is the product of the marginal distribution of $\beta_{i,c}$, which is the normal (or other) distribution defined by (6), and the conditional distribution of the dependent variable given the parameter $\beta_{i,c}$, 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 $\beta_{i,c}$ out of the joint distribution. The conditional mean of this distribution is then obtained by the definition,

---

[3] Note that although (7) is in terms of a specific film in a specific country, what follows will present a film specific effect conditioned on the incomes of all countries in which it was observed.

6

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

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 conditional mean is, then

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

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, $\hat{\beta}_{i,c,r} = \hat{\beta} + \hat{\delta} \log Income_c + \hat{\sigma}_\beta v_{i,\beta,r}$. 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 conditional variance of $\beta_{i,c}$, $Var[\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i] = E[\beta_{i,c}^2 \mid \mathbf{y}_i, \mathbf{X}_i] - (E[\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i])^2$. This is estimated in the same fashion by first estimating the conditional expected square, with

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

The standard deviation of the conditional distribution is estimated with

$$(13) \quad \widehat{S.D}[\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i] = \sqrt{\hat{E}[\beta_{i,c}^2 \mid \mathbf{y}_i, \mathbf{X}_i] - \left(\hat{E}[\beta_{i,c} \mid \mathbf{y}_i, \mathbf{X}_i]\right)^2}.$$

Table 1 (reproduced from Craig, Douglas and Greene (2003) presents the maximum simulated likelihood estimates for the model coefficients. The estimates in the first column are the nonrandom parameters counterparts to the RPM in (6). These are estimated by ordinary least squares.[4] 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* and the interaction term. (Note the zero 'income effects.') 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; $\chi^2 = 2(3154.532 - 3007.729) = 293.606$ with 2 degrees of freedom. As to whether US Box office results significantly

---

[4] Since the parameters are assumed to be nonrandom, no adjustment for heteroscedasticity is needed.

effect foreign sales, the least squares results strongly suggest so – the 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, as noted above, given the interaction term in the model, this assessment is incomplete. In the center column, note that the coefficient on per capita US box is 1.204 with a t ratio of 25.6. In the first column, the counterpart would be 1.131 + 0.011(logIncome) which requires a specific value of logIncome. Inserting the full sample mean of 6.4927 gives a result of 1.204 again. Using the estimated asymptotic covariance matrix for the two OLS coefficients [.05925/ -.008909,0.0013729] produces an estimated asymptotic standard error of 0.04597, so the results are almost identical to the preceding case. Based only on the OLS results, it seems clear that the data do suggest a significant relationishup. But, in the random parameters model, the simple t ratio on the estimate of $\hat{\beta}$ is 1.82 and that on $\hat{\delta}$ is only -0.5. Neither is statistically significant, so one might be tempted to conclude that the relationship is insignificant. Once again, however, it is necessary to consider the full effect, not just the structural coefficients. However, in the random parameters case, the simple mean estimator, itself, becomes an ambiguous estimator of the population quantities of interest.

Figure 1 below shows for each of the 299 films the range given by the estimated conditional mean plus and minus 2.5 conditional standard deviations, using, once again, the full sample mean of income.[5] With conditional normality, this range would encompass over 99% of the mass of each conditional distribution. Since the conditional 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, and that for all but two films, there is almost certainly a positive relationship between US box office and foreign box office.

In principle, one could estimate the 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,

---

[5] Note that the results suggest that separate means could be computed by country. In the more general case in which z_i varies only acros individuals, only a single value would be implied.

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

$$(14) \quad \hat{E}[WTP_{program} \mid data_i] = \frac{(1/R)\sum_{r=1}^{R}\left(\hat{\beta}_{program,i} / \hat{\beta}_{price,i}\right)\hat{L}(data_{i,r})}{(1/R)\sum_{r=1}^{R}\hat{L}(data_{i,r})}$$

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

4. Estimating Marginal Effects in Nonlinear Random Parameters Models

The computations described in the preceding section 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 and the data. One could estimate these for each individual in a sample.   We consider a straightforward application, the binary probit model. The probit model with random heterogeneously distributed parameters would be

$$y_{it}^* = \boldsymbol{\beta}_i'\mathbf{x}_{it} + \varepsilon_{it}, \varepsilon_{it} \sim N[0,1]$$

$$(15) \quad y_{it} = 1(y_{it}^* > 0)$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Delta\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i, \mathbf{v}_i \sim N[\mathbf{0},\mathbf{I}].$$

Assuming that both ε_{it} and $\mathbf{v}_i$ are normally distributed, by direct substitution, we find the unconditional result,

$$(16) \quad \text{Pr}\,\text{ob}[y_{it} = 1 \mid \mathbf{x}_{it}, \mathbf{z}_i] = \Phi\left[\frac{\boldsymbol{\beta}_i'\mathbf{x}_{it}' + \mathbf{z}_i'\Delta\mathbf{x}_{it}}{\sqrt{1 + \mathbf{x}_{it}'\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\mathbf{x}_{it}}}\right]$$

Thus, as noted earlier, unlike the linear regression model, this model cannot be estimated consistently just by adding the interaction terms to the original model and using the conventional maximum likelihood estimator.[6]

The random parameters probit model can be estimated as suggested in Section 2. The conditional joint density for the ith individual is

---

[6] It can be estimated conventionally by basing a likelihood function on (16), as suggested by Sepanski (2000) and Guilkey and Murphy (1993). However, maximum simulated likelihood estimation provides a much richer set of results in that it facilitates estimation of the individual level parameters. It is also likely to be easier given the complexity of the function in (16).

(17) $L_i(\mathbf{\Lambda} \mid \mathbf{y}_i, \mathbf{X}_i, \mathbf{z}_i, \mathbf{v}_i) = \prod_{t=1}^{T_i} \Phi[(2y_{it} - 1)\boldsymbol{\beta}_i' \mathbf{x}_{it}]$

where, as before, $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i$. Estimates of the structural parameters are obtained by maximizing (5) with this function. The probability is the conditional (on $\mathbf{x}_{it}$) mean function, so the marginal effects are

(18) $\boldsymbol{\delta}(\mathbf{x}_{it}, \mathbf{z}_i \mid \mathbf{v}_i) = (\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)\phi\left[(\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)' \mathbf{x}_{it}\right].$[7]

Note that since all parameters appear in the scale factor, the marginal effects for all variables are random, even if the corresponding coefficients are not. As before, it is necessary to integrate out the random unobserved heterogeneity. The unconditional marginal effects for individual i at time t are, therefore

(18) $\overline{\boldsymbol{\delta}}(\mathbf{x}_{it}, \mathbf{z}_i) \quad = \quad \mathbf{E}_{\mathbf{v}_i}\left[\boldsymbol{\delta}(\mathbf{x}_{it}, \mathbf{z}_i \mid \mathbf{v}_i)\right]$

$= \int_{\mathbf{v}_i} (\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)\phi\left[(\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i + \boldsymbol{\Gamma}\mathbf{v}_i)' \mathbf{x}_{it}\right] g(\mathbf{v}_i) d\mathbf{v}_i.$

One might compute this using the individual group means rather than at a particular point in time. As before, the integration must be done by simulation, as in (11).

(19) $\hat{E}[\hat{\boldsymbol{\delta}}(\mathbf{X}_i, \mathbf{z}_i)] = \dfrac{(1/R)\sum_{r=1}^{R} \hat{\boldsymbol{\delta}}_{i,r}\hat{L}(\mathbf{y}_i, \mathbf{X}_i, \mathbf{v}_{i,\beta,r})}{(1/R)\sum_{r=1}^{R} \hat{L}(\mathbf{y}_i, \mathbf{X}_i, \mathbf{v}_{i,\beta,r})}$

Bertschek and Lechner (1998) analyzed the product and innovation activity of 1,270 German manufacturing firms observed in five years, 1984 - 1988, in response to imports and foreign direct investment. [See Bertschek and Lechner (1998).] The model and different approaches to estimation of a panel probit model are described in Greene (2003, 2004). To illustrate these computations, we continue that analysis here.

The basic model to be estimated is a probit model based on the latent regression

$y_{it}^* = \beta_1 + \sum_{k=2}^{8} x_{k,it}\beta_k + \varepsilon_{it}$, $y_{it} = \mathbf{1}\left(y_{it}^* > 0\right)$, $i = 1,...,1270$, $t = 1984,...,1988$.

where

> y_{it} = 1 if a product innovation was realized by firm i in year t, 0 otherwise,
> $x_{2,it}$ = Log of industry sales in DM,
> $x_{3,it}$ = Relative firm size = ratio of employment in business unit to employment in the industry (times 30),
> $x_{4,it}$ = Import share = ratio of industry imports to (industry sales plus imports),
> $x_{5,it}$ = FDI share = Ratio of industry foreign direct investment to (industry sales,

---

[7] The effects would be computed as the difference in probabilities rather than as the derivative if the associated regressor were a dummy variable. The approach would be the same. In the interest of brevity, we treat only the simpler case of a continuous independent variable.

plus imports),

$x_{6,it}$ = Productivity = Ratio of industry value added to industry employment,
$x_{7,it}$ = Raw materials sector = 1 if the firm is in this sector,
$x_{8,it}$ = Investment goods sector = 1 if the firm is in this sector,
$x_{9,it}$ = Consumer goods sector = 1 if the firm is in this sector (not used)
$x_{10,it}$ = Food sector = 1 if the firm is in this sector (not used).[8]

Descriptive statistics and further discussion appear in the earlier papers. Primary interest in the study was in the effect of imports ($x_{4,it}$) and inward foreign direct investment ($x_{5,it}$) on innovation. Both are hypothesized to have a significant positive effect. Parameter estimates for the random parameters model are estimated under the specification

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{v}_i,$$

that is, with no interaction terms. Parameter estimates for the random parameters model are presented in Table 2 with the estimates of a pooled probit model (assuming that there are no common unobserved effects). As might be expected, the estimates are somewhat similar, but do differ noticeably. Figures 2 and 3 are kernel density estimates for the distribution of the individual parameters, $\beta_{4,i}$ and $\beta_{5,i}$ based on (11). Consistent with the findings in the article and in Greene (2004), we conclude that both imports and FDI do strongly influence innovation.

Marginal effects for the pooled probit model are shown in Table 3. The second row in Table shows the averages of the 1,270 individual estimates of the marginal effects. The differences across the two models are large, but perhaps less than the differences in the coefficients might suggest. Table 4 shows the averages of the estimated marginal effects by the four industry groups. There is considerable difference across the industries. Finally, Figure 4 shows the estimates of the marginal effect of imports on innovation for a sample of 200 of the consumer goods firms. The horizontal lines in the figure are drawn at zero and 0.436 which is the marginal effect computed at the sample means using the pooled probit model.

---

[8]Firms are classified into four industry groups, raw materials, investment goods, consumer goods and food. Only two of the industry dummy variables are included in the model. The original study does not elaborate on why the third was not included. We have retained the original specification.

5. Bayesian Analysis of Individual Heterogeneity

Allenby and Rossi (1999) have studied the issue discussed here at length in the context of a discrete 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 more flexible. Thus, we will eschew conclusions about the model in particular, and consider the general 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. The posterior distributions follow from the assumed priors (normal in their cases) and are easily drawn from 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 shows that the first of these is clearly 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 for the assumed priors and the data used. (A different posterior is asssociated with every different data set.) Also, it should be remembered that the Gibbs sampler is also a simulation based estimator subject, at least, to simulation variance; it is not a clear window into the true population. The Bayesian posterior means are not 'exact' measures of quantities that are only measured approximately by the classical methods. The classical methodology is explicitly directed toward characterizing the sampling distribution. There is no counterpart to the posterior. The two methods are estimating different quantities. We do note, though, whether this point/counterpoint is actually substantive may itself be moot. Regardless of the setting, with uninformative priors, the posterior is dominated by the likelihood. Huber and Train (2001) have compiled some empirical 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 unconditional normal distribution implied by (2) does not imply normality of the conditional – as they note, the shape of the conditional density will be influenced by the data. How one could sample directly from the conditional distribution, 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.

6. 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.

Bertschek, I. and M. Lechner, "Convenient Estimators for the Panel Probit Model," *Journal of Econometrics*, 87, 1998, pp. 329-372.

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*, 29, 1, 2004, pp. 21-48.

Guilkey, D., and J. Murphy, "Estimation and Testing in the Random Effects Probit Model," *Journal of Econometrics,* 59, 1993, pp. 301-317.

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

Kamakura W. and Wedel M. "An Empirical Bayes Procedure for Improving Individual-Level Estimates and Predictions From Finite Mixtures of Multinomial Logit Models *Journal of Business & Economic Statistics*, 22, 2004, pp. 121-125.

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

Sepanski, J., "On a Random Coefficient Probit Model," *Communications in Statistics - Theory and Methods*, 29, 11, 2000, pp. 2493-2505.

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.

Wooldridge, J., *Introductory Econometrics: A Modern Approach*, New York, Southwestern Publishers, 2000.

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	
	$\mathbf{\Gamma}_{21} = \mathbf{r}$[b]			
Disturbance std.deviation.	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.

Table 2. Estimated Components of the Fixed and Random Parameters Models for Innovation.
(Estimated Standard Errors in Parentheses)

Variable	Pooled Probit β	Random Parameters Probit β	Constant	log Sales	RelSize	Imports	FDI	Prod.	Raw Mtl.	Invst.	Std. Devs.
Constant	-1.96 (.230)	-3.620 (0.575)	3.56 (0.703)	0	0	0	0	0	0	0	3.562
log Sales	.137 (.022)	0.339 (0.056)	0.269 (0.067)	0.0849 (0.0228)	0	0	0	0	0	0	.2827
Relative size	1.073 (.142)	5.009 (0.596)	2.632 (0.603)	-4.56 (1.066)	1.90 (0.511)	0	0	0	0	0	5.597
Imports	1.134 (.151)	1.727 (0.293)	0.937 (0.543)	-0.113 (0.621)	1.227 (0.287)	1.161 (0..286)	0	0	0	0	1.935
FDI	2.853 (.402)	3.655 (0.739)	5.665 (1.400)	1.183 (1.048)	0.631 (0.611)	-1.431 (0.742)	2.479 (0.751)	0	0	0	6.487
Prod.	-2.34 (.715)	-6.005 (1.924)	-4.62 (1.771)	6.673 (2.393)	-1.339 (1.332)	-3.987 (1.098)	7.964 (1.113)	5.999 (1.057)	0	0	13.526
Raw Material	-.278 (.081)	-0.230 (0.213)	0.775 (0.382)	-0.654 (0.453)	0.215 (0.751)	-0.130 (0.666)	-0.665 (0.346)	1.215 (0.425)	0.354 (0.378)	0	1.770
Inv Goods	.187 (.039)	0.312 (0.077)	0.191 (0.190)	0.597 (0.207)	0.404 (0.0947)	-0.0372 (0.0894)	-0.375 (0.098)	0.117 (0.105)	-0.622 (0.044)	0.288 (0.048)	1.087
Log L	-4114.046	-3504.078									

Table 3. Average Estimates of Marginal Effects

	logSales	RelSize	Imports	FDI	Prod	Raw Mtl.	Inv. Good
Pooled	0.068	0.413	0.437	1.099	-0.902	-0.110	0.072
Random Parame	0.081	1.211	0.395	0.877	-1.532	-0.031	0.058

Table 4. Sample Average Estimated Marginal Effects by Industry from Random Parameters Model

	Investment	Consumer	Raw Material	Food
Imports	0.391	0.389	0.403	0.471
FDI	0.928	0.791	0.818	1.108
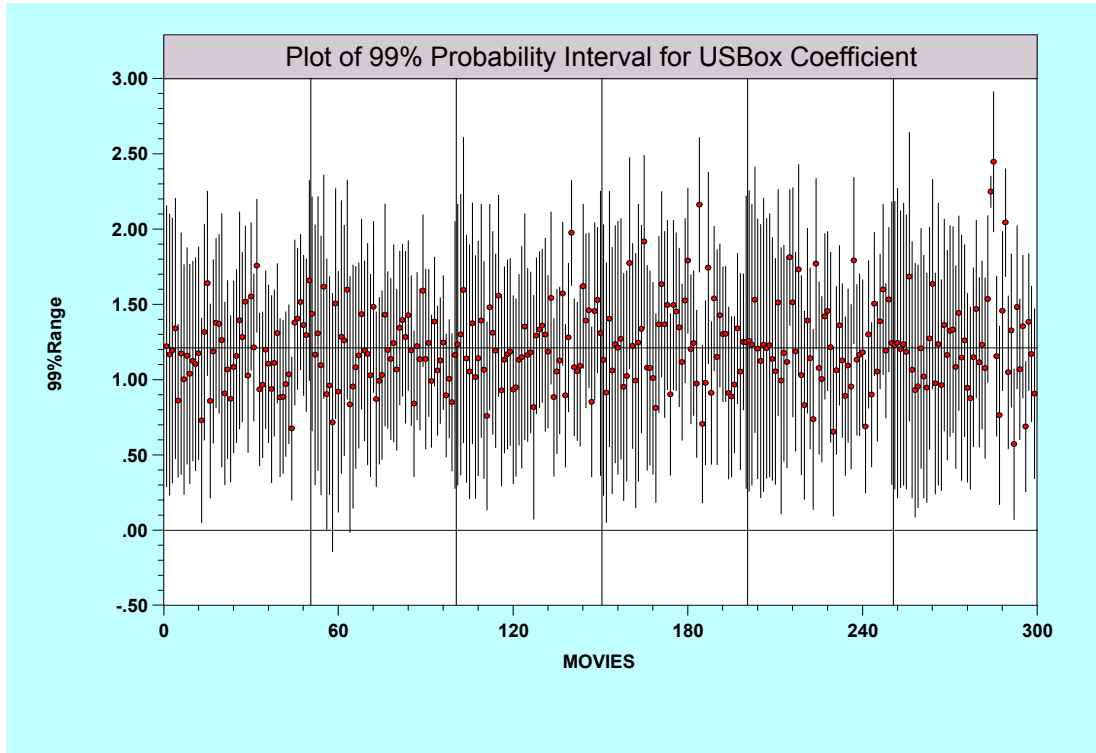
Figure 1. 99% Confidence Intervals for Film Specific Coefficients on logUSBox

Figure 2. Kernel Density Estimate for β_4

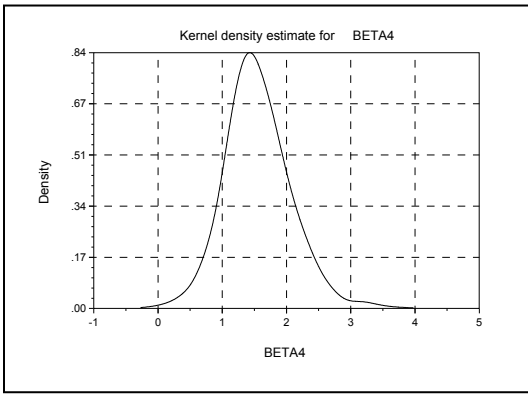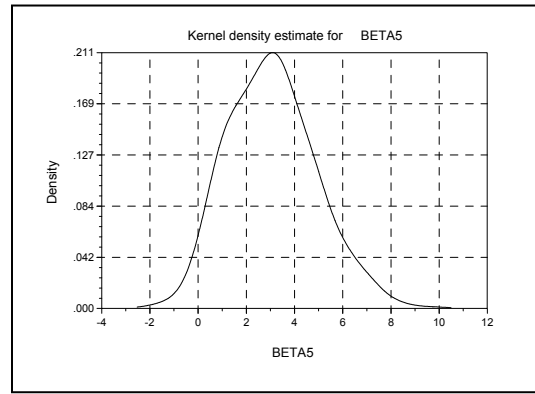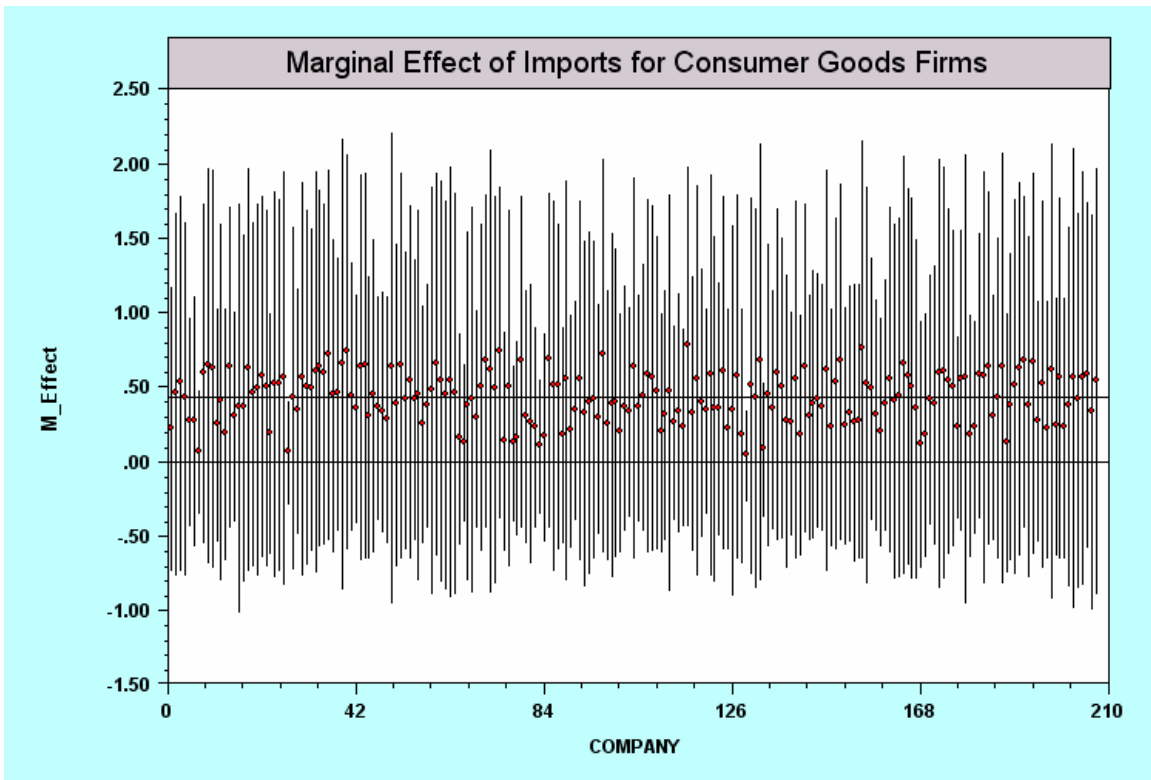**Figure 3. Kernel Density Estimate for β_5**

Figure 4. Estimated Marginal Effects