

Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Non-linear Least Squares¹

Matthew C. Harding²

Jerry Hausman³

March 14, 2007

¹This paper is forthcoming in the *International Economic Review*. We thank Ketan Patel for excellent research assistance. We thank Ronald Butler, Kenneth Train, Joan Walker and participants at the MIT Econometrics Lunch Seminar and at the Harvard Applied Statistics Workshop for comments.

²Department of Economics, MIT; and Institute for Quantitative Social Science, Harvard University. Email: MHarding@MIT.edu

³Department of Economics, MIT. E-mail: JHausman@MIT.edu.

Abstract

Current methods of estimating the random coefficients logit model employ simulations of the distribution of the taste parameters through pseudo-random sequences. These methods suffer from difficulties in estimating correlations between parameters and computational limitations such as the curse of dimensionality. This paper provides a solution to these problems by approximating the integral expression of the expected choice probability using a multivariate extension of the Laplace approximation. Simulation results reveal that our method performs very well, both in terms of accuracy and computational time.

1 Introduction

Understanding discrete economic choices is an important aspect of modern economics. McFadden (1974) introduced the *multinomial logit* model as a model of choice behavior derived from a random utility framework. An individual i faces the choice between K different goods $i = 1..K$. The utility to individual i from consuming good j is given by $U_{ij} = x'_{ij}\beta + \epsilon_{ij}$, where x'_{ij} corresponds to a set of choice relevant characteristics specific to the consumer-good pair (i, j) . The error component ϵ_{ij} is assumed to be independently identically distributed with an extreme value distribution $f(\epsilon_{ij}) = \exp(-\epsilon_{ij}) \exp(-\exp(-\epsilon_{ij}))$.

If individual i is constrained to choose a single good within the available set, utility maximization implies that some good j will be chosen over all other goods $l \neq j$ such that $U_{ij} > U_{il}$, for all $l \neq j$. We are interested in deriving the probability that consumer i chooses good j , which is

$$P_{ij} = \Pr[x'_{ij}\beta + \epsilon_{ij} > x'_{il}\beta + \epsilon_{il}, \text{ for all } l \neq j]. \quad (1)$$

McFadden (1974) shows that the resulting integral can be solved in closed form implying the familiar expression:

$$P_{ij} = \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^K \exp(x'_{ik}\beta)} (= s_{ij}). \quad (2)$$

In some analyses it is also useful to think of the market shares of different firms. Without loss of generality we can also consider the choice probability described above to be the share of the total market demand which goes to good j in market i and we will denote this by s_{ij} . All the results derived in this paper will be valid for either interpretation. For convenience we shall focus on the market shares interpretation of the above equation.

The vector of coefficients β can be thought of as a representation of the individual tastes

and determines the choice, conditional on the observable consumer-good characteristics. Although an extremely useful model, the multinomial model suffers from an important limitation: it is built around the the assumption of independence of irrelevant alternatives (IIA), which implies equal cross price elasticities across all choices, as demonstrated by Hausman (1975). Additionally, it does not allow for correlations between the random components of utility, thus limiting the complexity of individual choice which can be modeled (Hausman and Wise, 1978).

While a number of more flexible specifications have been proposed, few proved to be computationally tractable. The addition of a random coefficients framework to the logit model provides an attractive alternative (Cardell and Dunbar, 1980). In many applications however it is important to think of tastes as varying in the population of consumers according to a distribution $F(\beta)$. It is particularly important not to assume the taste parameters to be independent. The estimation of correlations between the components of the vector β is of major interest. The resulting correlations describe patterns of substitution between different product characteristics.

In practice, we often assume that the distribution $F(\beta)$ is Normal with mean b and covariance Σ . The purpose of random coefficients models is to estimate the unknown parameters b and Σ from the available sample. From a computational point of view, the aim is to obtain the expected share of good j in market i from the evaluation of the following expectation:

$$E_{\beta}(s_{ij}) = \int_{-\infty}^{+\infty} \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^K \exp(x'_{ik}\beta)} dF(\beta) \quad (3)$$

We denote this model to be the random coefficients logit model. The above expression corresponds to a multivariate integral over the dimension of the space of the taste parameters. Since the integral does not have a known analytic solution, the use of simulation

methods currently plays an important part in the implementation of these models (Lerman and Manski, 1981) with recent applications employing pseudo-random Halton sequences (Small, Winston and Yan, 2005; Train, 2003).

The random coefficients logit model is an extremely versatile tool for the analysis of discrete choices since it can be thought of as an arbitrarily close approximate representation of any random utility model consistent with choice probabilities (McFadden and Train, 2000). This has prompted researchers to think of this model as “one of the most promising state of the art discrete choice models” (Hensher and Green, 2003). Applications of the random coefficients logit model abound, not only within economics, but also in related disciplines such as marketing or transportation research (Hess and Polak, 2005). The random coefficients model is also an important building block for more complex models. Thus, Berry, Levinsohn and Pakes (1995) employ the random coefficients logit model to analyse demand based on market-level price and quantity data. Bajari, Hong and Ryan (2005) incorporate it into an econometric model of discrete games with perfect information, where it selects the probability of different equilibria.

The implementation of the random coefficients model remains a challenging application of the method of simulated moments. In particular the estimation of a full covariance matrix of the taste parameters, which fully incorporates all the possible correlations between parameters, seems to elude most researchers and appears to be a serious limitation of the simulation approach. In Section 2 of this paper we will derive an analytic approximation of the integral expression in Equation 3 which can be incorporated into an extremely convenient non-linear least squares framework for the estimation of all mean and variance-covariance parameters of the taste distribution. Section 3 shows the superior performance of the new method based on the Laplace method compared to the simulation alternative in cases where the model is specified with non-zero correlations.

2 A Laplace Approximation of the Expected Share

Consider the expected share of product j in market i under the random coefficients logit model introduced above.

$$E_{\beta}(s_{ij}) = E_{\beta} \left\{ \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^K \exp(x'_{ik}\beta)} \right\} = E_{\beta} \left\{ \left(\sum_{k=1}^K \exp(x'_{ijk}\beta) \right)^{-1} \right\}, \quad (4)$$

where $x_{ijk} = x_{ik} - x_{ij}$ for all k . Assume that the taste parameters β are drawn from a normal multivariate distribution with mean b and covariance matrix Σ ,

$$f(\beta) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - b)' \Sigma^{-1} (\beta - b) \right\}. \quad (5)$$

For simplicity we focus in our derivations on the case where all coefficients are random. More generally, we may wish to allow for mixture of fixed and random coefficients. The results in this paper will continue to hold in this case too and we restate the main result of this paper in terms of both random and fixed coefficients in Appendix B.

Then the expected share is given by the following multivariate integral:

$$E_{\beta}(s_{ij}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \int_{-\infty}^{+\infty} \exp[-g(\beta)] d\beta, \text{ where} \quad (6)$$

$$g(\beta) = \frac{1}{2} (\beta - b)' \Sigma^{-1} (\beta - b) + \log \left(\sum_{k=1}^K \exp(x'_{ijk}\beta) \right) \quad (7)$$

In this section we provide an approximation to the integral expression above using the asymptotic method of Laplace. While univariate applications of this method are common to mathematics and physics, where they are routinely applied to the complex functions

in order to derive “saddle-point approximations”, few applications to econometrics or statistics have been attempted. The extension of the method to multivariate settings was developed by Hsu (1948) and Glynn (1980). A statement of the main theorem is given in Appendix A together with the technical conditions required for the approximation to exist. Statistical applications of the Laplace approximation were developed by Daniels (1954) and Barndorff-Nielsen and Cox (1979) who employ the Laplace approximation to derive the *indirect Edgeworth expansion*, a generalization of the Edgeworth expansion method for distributions to exponential families. The Laplace method was also applied in Bayesian statistics to derive approximations to posterior moments and distributions (Tierney and Kadane, 1986; Efstathiou, Guthierrez-Pena and Smith, 1998). More recently, Butler (2002) noticed that the Laplace approximation often produces accurate results in sub-asymptotic situations which are not covered by the traditional setting. It is this insight which we will use below.

Now perform a Taylor expansion of the function $g(\beta)$ around the point $\tilde{\beta}_{ij}$, such that $g(\tilde{\beta}_{ij}) < g(\beta)$ for all $\beta \neq \tilde{\beta}$. This expansion is given by:

$$g(\beta) \cong g(\tilde{\beta}_{ij}) + (\beta - \tilde{\beta}_{ij})' \left[\frac{\partial g}{\partial \beta} \Big|_{\beta=\tilde{\beta}_{ij}} \right] + \frac{1}{2} (\beta - \tilde{\beta}_{ij})' \left[\left(\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'} \right)_{\beta=\tilde{\beta}_{ij}} \right] (\beta - \tilde{\beta}_{ij}) + O((\beta - \tilde{\beta}_{ij})^3). \quad (8)$$

Substituting in the integral expression above we obtain:

$$E_{\beta}(s_{i,j}) \cong |\Sigma|^{-1/2} \exp(-g(\tilde{\beta}_{i,j})) \times \int_{-\infty}^{+\infty} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} (\beta - \tilde{\beta}_{ij})' \left[\left(\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'} \right)_{\beta=\tilde{\beta}_{ij}} \right] (\beta - \tilde{\beta}_{ij}) + O((\beta - \tilde{\beta}_{ij})^3) \right\} d\beta \quad (9)$$

The intuition for this approach is given by the fact that if $g(\beta)$ has a minimum at the point $\tilde{\beta}_{ij}$, then the contribution of the function $g(\beta)$ to the exponential integral will be

dominated by a small region around the point $\tilde{\beta}_{ij}$. Furthermore by using a second order Taylor expansion around $\tilde{\beta}_{ij}$, we make the further assumption that the higher order terms of the expansion may be safely ignored. Let $\tilde{\Sigma}_{ij}$ be the inverse of the Hessian of $g(\beta)$ evaluated at $\tilde{\beta}_{ij}$, i.e. $\tilde{\Sigma}_{ij}^{-1} = \left(\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'} \right)_{\beta = \tilde{\beta}_{ij}}$. Note that both $\tilde{\beta}_{ij}$ and $\tilde{\Sigma}_{ij}$ are indexed by i and j to remind us that these values depend on the covariates of product j in market i explicitly and in general will not be constant across products or markets.

Then, we can re-write the integral above as:

$$E_{\beta}(s_{ij}) \cong |\Sigma|^{-1/2} \exp(-g(\tilde{\beta}_{ij})) \left| \tilde{\Sigma}_{ij} \right|^{1/2} (2\pi)^{-p/2} \left| \tilde{\Sigma}_{ij} \right|^{-1/2} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} (\beta - \tilde{\beta}_{ij})' \tilde{\Sigma}_{ij}^{-1} (\beta - \tilde{\beta}_{ij}) \right\} d\beta. \quad (10)$$

We recognize the right hand side of this expression to be the Gaussian integral, that is the integral over the probability density of a Normal variable β with mean $\tilde{\beta}_{ij}$ and covariance $\tilde{\Sigma}_{ij}$. Since this area integrates to 1 we have,

$$(2\pi)^{-p/2} \left| \tilde{\Sigma}_{ij} \right|^{-1/2} \left[\int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2} (\beta - \tilde{\beta}_{ij})' \tilde{\Sigma}_{ij}^{-1} (\beta - \tilde{\beta}_{ij}) \right\} d\beta \right] = 1 \quad (11)$$

and we can write the expected share of product i in market j as

$$E_{\beta}(s_{ij}) \cong \sqrt{\frac{\left| \tilde{\Sigma}_{ij} \right|}{|\Sigma|}} \exp(-g(\tilde{\beta}_{ij})) \quad (12)$$

The expansion point $\tilde{\beta}_{ij}$ has to be chosen optimally for each share, that is $\tilde{\beta}_{ij}$ solves the

equation $g'(\beta)|_{\beta=\tilde{\beta}_{ij}} = 0$, *i.e.*

$$\left(\tilde{\beta}_{ij} - b\right)' \Sigma^{-1} + \sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk} \tilde{\beta}_{ij})}{\sum_{k=1}^K \exp(x'_{ijk} \tilde{\beta}_{ij})} \right\} = 0 \quad (13)$$

In Appendix B we show that $-g(\beta)$ is the sum of two strictly concave functions and thus it is also concave. Hence, the function $g(\beta)$ attains a unique minimum at the point $\tilde{\beta}_{ij}$. We can also think of the optimal expansion point $\tilde{\beta}_{ij}$ as solving a fixed-point equation, $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$, where

$$B(\tilde{\beta}_{ij}) = b' - \left[\sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk} \tilde{\beta}_{ij})}{\sum_{k=1}^K \exp(x'_{ijk} \tilde{\beta}_{ij})} \right\} \right] \Sigma. \quad (14)$$

Additionally, the Hessian of $g(\beta)$ is given by

$$\frac{\partial^2 g(\beta)}{\partial \beta \partial \beta'} = \Sigma^{-1} + \frac{\sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk} \tilde{\beta}_{ij})}{\sum_{k=1}^K \exp(x'_{ijk} \tilde{\beta}_{ij})} - \frac{\left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk} \tilde{\beta}_{ij}) \right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk} \tilde{\beta}_{ij}) \right]}{\left[\sum_{k=1}^K \exp(x'_{ijk} \tilde{\beta}_{ij}) \right]^2} \quad (15)$$

The following proposition summarizes the main result of this paper by approximating the Gaussian integral corresponding to the expected share of product i in market j using a Laplace approximation.

Proposition 1: *If β has a Normal distribution with mean b and covariance Σ , we can*

approximate $E_{\beta}(s_{ij}) = E_{\beta}\left\{\left(\sum_{k=1}^K \exp(x'_{ijk}\beta)\right)^{-1}\right\}$ by

$$E_{\beta}(s_{ij}) \cong \sqrt{\frac{|\tilde{\Sigma}_{ij}|}{|\Sigma|}} \exp\left\{-\frac{1}{2}(\tilde{\beta}_{ij} - b)' \tilde{\Sigma}_{ij}^{-1}(\tilde{\beta}_{ij} - b)\right\} \left(\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})\right)^{-1}, \quad (16)$$

where

$$\tilde{\Sigma}_{ij} = \left\{ \Sigma^{-1} + \frac{\sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})} - \frac{\left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij})\right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk}\tilde{\beta}_{ij})\right]}{\left[\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})\right]^2} \right\}^{-1}, \quad (17)$$

and $\tilde{\beta}_{ij}$ solves the fixed-point equation $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$ for

$$B(\tilde{\beta}_{ij}) = b' - \left[\sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk}\tilde{\beta}_{ij})}{\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})} \right\} \right] \Sigma. \quad (18)$$

In the next section we present detailed simulation results which show the performance of the approximation in estimating the unknown parameters b and Σ of the model. The figure below shows the remarkably good fit between of the Laplace approximation of the true market share at fixed values of b and Σ for two covariates.

The exact expected share obtained by numerical integration coincides with the expected share obtained by the Laplace approximation almost everywhere. The only noticeable deviation occurs for values of the expected share close to 1. Fortunately, this case is relatively infrequent in economic applications, where in multi-brand competition models we may expect to have many small shares in any given market but it is unlikely to have more than a few very large shares in the entire sample. The Laplace approximation introduced

in this section has the peculiar property of being an asymmetrical approximation to a symmetrical function. This feature however proves to be extremely useful for economic applications since it provides an very close approximation to small shares which are much more likely to occur in economic data than shares close to 1, where the approximation tends to underestimate the true expected share.

The optimal expansion point $\tilde{\beta}_{ij}$ used in Proposition 1 can be computed by standard iterative methods which solve the fixed-point equation $\tilde{\beta}_{ij} = B(\tilde{\beta}_{ij})$. While such methods are widely available in commercial software packages and tend to be extremely fast, the optimal expansion point $\tilde{\beta}_{ij}$ needs to be computed for each firm in each market separately, which may potentially slow down numerical optimization routines if large data sets are used. To improve computational efficiency we can further derive an approximate solution to the fixed point equation, which as we will show in the next section, performs very well.

Let $h(\beta) = \log \left(\sum_{k=1}^K \exp(x'_{ijk}\beta) \right)$ and perform a quadratic Taylor approximation of $g(\beta)$ around the constant parameter vector b . Then,

$$h(\beta) \cong h_{ij}(b) + (\beta - b)' J_{ij}(b) + \frac{1}{2}(\beta - b)' H_{ij}(b) (\beta - b) + O((\beta - b)^3), \quad (19)$$

where the Jacobian and Hessian terms are given by

$$J_{ij}(b) = \sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk}b)}{\sum_{k=1}^K \exp(x'_{ijk}b)} \right\} \quad \text{and} \quad (20)$$

$$H_{ij}(b) = \frac{\sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk}b)}{\sum_{k=1}^K \exp(x'_{ijk}b)} - \frac{\left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk}b) \right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk}b) \right]}{\left[\sum_{k=1}^K \exp(x'_{ijk}b) \right]^2}. \quad (21)$$

Thus, we can re-write the expression for $g(\beta)$ as

$$g(\beta) = \frac{1}{2} (\beta - b)' \Sigma^{-1} (\beta - b) + h_{ij}(b) + (\beta - b)' J_{ij}(b) + \frac{1}{2} (\beta - b)' H_{ij}(b) (\beta - b) \quad (22)$$

The optimal expansion point $\tilde{\beta}_{ij}$ solves the equation $\partial g(\beta)/\partial \beta = 0$. Hence,

$$\frac{\partial g(\beta)}{\partial \beta} = (\beta - b)' \Sigma^{-1} + J_{ij}(b) + (\beta - b)' H_{ij}(b) = 0. \quad (23)$$

Since this expression is now linear we can easily solve for the optimal expansion point $\tilde{\beta}_{ij}$,

$$\tilde{\beta}_{ij} = b + [\Sigma^{-1} + H_{ij}(b)]^{-1} J'_{ij}(b). \quad (24)$$

We can now re-write Proposition 1 to obtain an easily implementable version of the Laplace approximation of the expected share.

Proposition 2: *If β has a Normal distribution with mean β and covariance Σ , we can approximate $E_\beta(s_{ij}) = E_\beta\left\{\left(\sum_{k=1}^K \exp(x'_{ijk}\beta)\right)^{-1}\right\}$ by*

$$E_\beta(s_{ij}) \cong \sqrt{\frac{|\tilde{\Sigma}_{ij}|}{|\Sigma|}} \exp\left\{-\frac{1}{2}(\tilde{\beta}_{ij} - b)' \tilde{\Sigma}_{ij}^{-1}(\tilde{\beta}_{ij} - b)\right\} \left(\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})\right)^{-1}, \quad (25)$$

where

$$\tilde{\beta}_{ij} = b + [\Sigma^{-1} + H_{ij}(b^*)_{b^*=b}]^{-1} J'_{ij}(b) \quad (26)$$

$$\tilde{\Sigma}_{ij}^{-1} = \Sigma^{-1} + H_{ij}(b^*)_{b^*=\tilde{\beta}_{ij}}, \quad (27)$$

and

$$J_{ij}(b) = \sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk}b)}{\sum_{k=1}^K \exp(x'_{ijk}b)} \right\} \quad (28)$$

$$H_{ij}(b^*) = \frac{\sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk}b^*)}{\sum_{k=1}^K \exp(x'_{ijk}b^*)} - \frac{\left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk}b^*) \right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk}b^*) \right]}{\left[\sum_{k=1}^K \exp(x'_{ijk}b^*) \right]^2}. \quad (29)$$

Notice that the Hessian expression $H_{ij}(b^*)$ is evaluated at different points b^* in the computation of the values of $\tilde{\beta}_{ij}$ and $\tilde{\Sigma}_{ij}$. Proposition 2 is also insightful in that it explains why a simple Taylor expansion of the Gaussian integral around the mean b will fail. Consider the expression for $\tilde{\beta}_{ij}$, which is the optimal expansion point in the Laplace approximation. Notice that $\tilde{\beta}_{ij}=b$ only if $J_{ij}(b) = 0$. But this expression can only be zero if the vectors of covariates x_{ijk} are zero for all k . Hence a Taylor approximation of the same problem will fail since it expands each expected share around a constant value when in fact it ought to perform the expansion around an optimal value which will differ from share to share depending on the covariates. The Laplace approximation developed above performs this optimal expansion.

3 Monte-Carlo Simulations

In this section we discuss the estimation of the random coefficients model by non-linear least squares after applying the Laplace approximation derived in the previous section to each expected market share. We will also compare its performance in Monte-Carlo simulations to that of alternative methods used for the estimation of these models in the econometric literature.

Since the model was introduced over thirty years ago, several estimation methods have been proposed which try to circumvent the problem that the integral expression for the expected shares does not have a closed form solution for most distributions of the taste parameters. While numerical integration by quadrature is implemented in numerous software packages it is also extremely time consuming. In practice it is not possible to use numerical integration to solve such problems if the number of regressors is greater than two or three. We have found that even for the case of a single regressor this method is extremely slow and not always reliable.

The main attempt to estimate random coefficients models is based on the method of simulated moments (McFadden, 1989; Pakes and Pollard, 1989), where the expectation is replaced by an average over repeated draws from the distribution of taste parameters:

$$E_{\beta}(s_{ij}) = \int \frac{\exp(x'_{ij}\beta)}{\sum_{k=1}^K \exp(x'_{ik}\beta)} dF(\beta) \cong \frac{1}{R} \sum_{r=1}^R \frac{\exp(x'_{ij}\bar{\beta}_r)}{\sum_{k=1}^K \exp(x'_{ik}\bar{\beta}_r)}, \quad (30)$$

where $\bar{\beta}_r$ is drawn from the distribution $F(\beta)$. Random sampling from a distribution may nevertheless provide poor coverage of the domain of integration. There is no guarantee that in a particular set of draws the obtained sequence will uniformly cover the domain of integration and may in fact exhibit random clusters which will distort the approximation. To achieve a good approximation the number of draws R will have to be very large.

More recently the use of variance reduction techniques has been advocated in an attempt to improve the properties of simulated estimation (Train, 2003). Negatively correlated pseudo-random sequences may lead to a lower variance of the resulting estimator than traditional independent sampling methods. The method currently employed in econometrics uses *Halton sequences* (Small, Winston and Yan, 2005).

Halton sequences can be constructed as follows. For each dimension r of the vector β

and some prime number k construct the sequence

$$s_{t+1} = \left\{ s_t, s_t + \frac{1}{k}, \dots, s_t + \frac{(k-1)}{k^t} \right\}, \text{ for } s_0 = 0. \quad (31)$$

This sequence is then randomized by drawing μ uniform (0,1) and for each element s , letting $s^* = \text{mod}(s + \mu)$.

This method provides coverage of the unit hypercube by associating each dimension with a different prime number k . In order to transform these points into draws from the relevant distribution, an inversion is then applied, e.g. if the desired distribution is Normal one would turn these points on the unit hypercube into values of β , by letting $\bar{\beta}_r = \Phi^{-1}(s_r^*)$, which corresponds to the inverse of the normal distribution.

The use of Halton sequences improves performance over the use of independent draws and yet nevertheless it suffers from the curse of dimensionality. Many thousand draws are required for each observation and the application of this method is extremely problematic for the estimation of even a small number of parameters since it is so time consuming.

The mathematical properties of Halton sequences are not sufficiently well understood and may represent a liability in some applications. Train (2003) reports that in estimating a random coefficients logit model for households' choice of electricity supplier repeatedly, most runs provided similar estimates of the coefficients, yet some runs provided significantly different coefficients even though the algorithm was unchanged and applied to the same data set. Similarly, Chiou and Walker (2005) report that simulation based methods may falsely identify models if the number of draws is not sufficiently large. The algorithm may produce spurious results which "look" reasonable yet are not supported by the underlying data.

Additionally, to our knowledge, it was not possible so far to reliably estimate the full covariance matrix using simulation based methods. Researchers focus exclusively on the

estimation of the mean and variance parameters thereby assuming a diagonal structure to the covariance matrix Σ of the taste parameters. We will show how this problem can be easily overcome by the use of the Laplace approximation method we propose in this paper. Later on in this section we will also show how ignoring the covariances may lead to biased results and unreliable policy analysis if the taste parameters in the true data generating process are correlated.

We propose estimating the model parameters (b, Σ) by non-linear least squares. Let s_{ij} be the observed market share of firm j in market i . We can construct the approximation of the expected share using the Laplace approximation as described in Section 2, $\hat{s}_{ij}(b, \Sigma) = E_{\beta}(s_{ij})$. This will be a non-linear function in the model parameters b and Σ and can be implemented using either Proposition 1 or Proposition 2. The implementation of Proposition 2 is immediate and only involves the use of matrix functions. We can then proceed to estimate the model parameters by least squares or weighted least squares which can improve efficiency:

$$\left(\hat{\beta}, \hat{\Sigma}\right) = \operatorname{argmin}_{\beta, \Sigma} \sum_{i=1}^N \sum_{j=1}^K (s_{ij} - \hat{s}_{ij}(b, \Sigma))^2. \quad (32)$$

The optimization can be achieved using a Newton type constrained optimization routine. Some parameters may require linear constraints (e.g. if the optimization is performed over variance parameters, then $(\sigma^2)_p > 0$ for all taste parameters β_p). The optimization needs to ensure that the estimated covariance matrix is positive definite at each step, for example by employing an appropriate re-parameterization or the Cholesky decomposition

This can be achieved by an appropriate penalization at the edges of the allowable domain. The model can also be estimated by minimum chi-square techniques or by maximum likelihood given our evaluation of the expected shares. Simulation results suggest no significant performance differences between these different methods of implementation.

Table 1: Estimation of the one variable random coefficients model. $N = 1000, K = 6$.

Mean Bias	Quadrature	Fixed Point Laplace	Laplace	Halton
b	0.00778	0.00492	0.00555	0.00269
σ^2	0.04957	0.02348	0.00634	-0.02011

MSE	Quadrature	Fixed Point Laplace	Laplace	Halton
b	0.01003	0.00297	0.00281	0.00301
σ^2	0.09833	0.06641	0.08357	0.07381

In Table 1 we estimate a random coefficients model with a single taste parameter using the methods discussed above. The covariate is drawn from a mixture distribution of a normal and a uniform random variable. This particular construction is performed in order to correct for unreliable estimates that have been reported when only normal covariates are being used. Since the model only requires univariate integration we can also perform numerical integration. We use a second order Newton-Coates algorithm to perform the integration by quadrature for each expected share. Additionally we compute estimates using the two versions of the Laplace approximation of the expected share as described in Section 2 in Propositions 1 and 2 respectively. The results labelled as “Fixed Point Laplace” compute the optimal expansion points $\tilde{\beta}_{ij}$ using iterative fixed point techniques. The results labelled “Laplace” approximate this fixed point calculation using the analytic expression of Proposition 2. We also compute estimates using Halton sequences as implemented by Whinston, Small and Yan (2005). We perform 500 draws for each observation.

The results in Table 1 show that all four methods produce comparable results. Interestingly though, numerical integration tends to be outperformed by either of the approximation methods presented here. In particular the Laplace approximation we proposed performs very similarly to the simulated estimation based on Halton sequences both in terms of mean bias and mean squared error. This result was confirmed in additional sim-

ulations were the number of taste parameters was increased. The Laplace approximation introduced in this paper outperforms the method of simulated moments in terms of computational time. Even in this simple one dimensional example the Laplace method runs about three times faster than the corresponding estimation using Halton sequences.

We have found no significantly different performance results between the Laplace approximation using the fixed point calculation and that using the approximation to the optimal expansion point. The Laplace approximation of Proposition 2 nevertheless outperformed all other methods in terms of computational time, being 3 to 5 times faster than the simulation approach.

Once we allow for multiple taste parameters we can ask the question whether these taste parameters are correlated with each other. Consider a model with 3 taste parameters, drawn from a distribution with mean $(b_1, b_2, b_3)'$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. In many cases of interest there is no a priori reason to constrain the covariance matrix of this distribution to be diagonal. We can allow for correlations between taste parameters by setting the off-diagonal elements of the covariance matrix equal to $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ for $-1 < \rho_{ij} < 1$. The parameter ρ_{ij} measures the strength of the correlation between the different taste parameters. The full covariance matrix which needs to be estimated in this case is:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \quad (33)$$

We use the Laplace approximation method to estimate all 9 parameters and report results for mean bias and MSE in Table 2. We were not able to estimate the same parameters using the method of simulated moments with Halton sequences. The algorithm failed to converge for Halton sequences under different model parameters and different starting

Table 2: Estimation of the three variable random coefficients model with covariances. $N = 2000, K = 6$.

Laplace	Mean Bias	MSE
b_1	0.01167	0.00233
b_2	0.00679	0.00201
b_3	-0.00371	0.00298
σ_1^2	-0.06889	0.09499
σ_2^2	-0.08245	0.07016
σ_3^2	0.03880	0.03180
ρ_{12}	0.04918	0.00774
ρ_{13}	0.04317	0.00350
ρ_{23}	-0.00702	0.00551

values.

Computational issues involving the use of simulated moments seem to have prevented empirical work involving the estimation of the full covariance matrix. We now wish to explore to what extent this may bias the results. To this purpose we estimate the same model as in the above example but ignore the covariances. Thus the true model has $\rho_{ij} \neq 0$ but we only estimate the restricted model where we assume $\rho_{ij} = 0$ for all $i, j, i \neq j$.

The results are presented in Table 3. We were able to obtain estimates of the restricted model using both the new Laplace approximation we propose and by using the simulation approach involving Halton sequences. Once again both methods produce comparable results. While the estimates of the mean parameters $(b_1, b_2, b_3)'$ seem to be sufficiently robust to the misspecification of the covariance matrix, the estimates of the variance parameters $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ seem to be strongly affected by the non-inclusion of the covariance terms in the optimization. The size of the bias is model dependent and we have found an absolute value of the bias between 30-60% in most simulations. Additionally, it seems that negative correlations which are falsely excluded bias the results much more than positive ones.

The failure to include the correlations between taste parameters may also lead to incor-

Table 3: Estimation of the three variable random coefficients model without covariances. The true model contains covariances but these are not estimated. $N = 2000, K = 6$.

	Mean Bias Laplace	Mean Bias Halton	MSE Laplace	MSE Halton
b_1	0.02037	0.01003	0.00321	0.00256
b_2	0.01582	0.00778	0.00201	0.00258
b_3	0.00651	0.00212	0.00122	0.00197
σ_1^2	-0.01032	-0.21102	0.10192	0.18226
σ_2^2	-0.50883	-0.43340	0.32381	0.27094
σ_3^2	-0.12991	-0.14967	0.03900	0.09577

rect policy recommendations. Thus, consider the three variable described above where the true data generating process has non-zero correlation terms and a full covariance matrix.

We can interpret the model as follows.

We label the first variable as “price” and consider the policy experiment whereby the government has to decide whether to impose a 10% tax on a specific good. The tax is fully passed on to the consumers in the form of a 10% price increase. There are $K = 6$ competing firms in each market producing differentiated brands of the good on which the tax was imposed. We wish to simulate the ex post effect of the tax on the market shares of each firm. In order to do so we collect a sample of observations consisting of the market shares of each firm in different markets and the product characteristics of the differentiated good produced by brand and market. We estimate the random coefficients model with a full covariance matrix which allows for correlations between taste parameters. We also estimate the same model but limit ourselves to estimating a diagonal covariance thus restricting the correlations to be zero and also derive the logit estimates of the means corresponding to the case where the taste parameters are assumed to be constant in the population. We can use these estimates to simulate the distribution of market shares of each firm across the markets and compare them to the initial distribution of market shares before the tax was

implemented. We present the resulting distributions in Figure 2.

If we estimate any of the misspecified models by using either the logit estimates of Equation 2 or the random coefficients logit estimates of Equation 3 under the assumption of no correlation we would reach very different conclusions from the case when we take into account the full covariance matrix between taste parameters. Thus we can see how ignoring the correlations may lead to incorrect policy recommendations when the random coefficients model is used to estimate the distribution of taste parameters.

4 Conclusion

In this paper we have introduced a new analytic approximation to the choice probability in a random coefficients logit model. The approximation was derived using a multivariate extension of the Laplace approximation for sub-asymptotic domains. The expression results in a non-linear function of the data and parameters which can be conveniently estimated using non-linear least squares.

This new method of estimating random coefficients logit models allows for the estimation of correlations between taste parameters. The estimation of a full covariance matrix seems to have eluded many previous implementations of the random coefficients logit model employing simulations of the underlying taste distributions.

Simulation results show that our new method performs extremely well, both in terms of numerical accuracy and computational time. We also provide an example of the importance of estimating correlations between taste parameters through a tax simulation where very different policy implications would be reached if the estimated model is misspecified by restricting the correlations to be zero.

In this paper we have focused on the case of Normal preferences. Harding and Hausman (2007) show how the Laplace approximation procedure described in this paper can also be

applied to more general preference specifications that allow for skewness or multimodality in addition to correlations between taste parameters.

A Appendix

This appendix states the multivariate Laplace approximation theorem. For additional discussions of the theorem and applications to statistics see Muirhead (2005) and Jensen (1995). A proof is given in Hsu (1948).

Laplace Approximation Theorem. Let D be a subset of R^p and let f and g be real-valued functions on D and T a real parameter. Consider the integral

$$I = \int_{\beta \in D} f(\beta) \exp(-Tg(\beta)) d\beta \quad (34)$$

- a) g has an absolute minimum at an interior point $\tilde{\beta}$ of D ;
- b) there exists $T \geq 0$ such that $f(\beta) \exp(-Tg(\beta))$ is absolutely integrable over the domain D ;
- c) all first and second order partial derivatives of $g(x)$, $\frac{\partial g}{\partial \beta_i}$, $\frac{\partial^2 g}{\partial \beta_i \partial \beta_j}$, for $i = 1 \dots p$ and $j = 1 \dots p$ exist and are continuous in the neighborhood $N(\tilde{\beta})$ of $\tilde{\beta}$.
- d) there is a constant $\gamma < 1$ such that $\left| \frac{\exp(-g(\beta))}{\exp(-g(\tilde{\beta}))} \right| < \gamma$ for all $x \in D \setminus N(\tilde{\beta})$
- e) f is continuous in a neighborhood $N(\tilde{\beta})$ of $\tilde{\beta}$.

Then for large T , we have:

$$\tilde{I} = \left(\frac{2\pi}{T} \right)^{p/2} [\det(H(\tilde{\beta}))]^{-1/2} f(\tilde{\beta}) \exp(-Tg(\tilde{\beta})), \text{ where } H(\tilde{\beta}) = \frac{\partial^2 g(\tilde{\beta})}{\partial \tilde{\beta} \partial \tilde{\beta}'} \quad (35)$$

and

$$I = \tilde{I}(1 + O(T^{-1})) \quad \text{as } T \rightarrow \infty. \quad (36)$$

In Section 2 we let $f(\beta) = 1$ and $g(\beta) = \frac{1}{2}(\beta - b)' \Sigma^{-1}(\beta - b) + \log \left(\sum_{k=1}^K \exp(x'_{ijk} \beta) \right)$.

This is sometimes referred to as an *exponential form Laplace approximation*.

Moreover we use the observation of Butler (2002) that in many cases of interest this approximation performs very well even in sub-asymptotic cases where T remains small. In our case $T = 1$.

B Appendix

In some applications we may wish to allow for a mixture of fixed and random coefficients. We can partition the $p \times 1$ dimensional vector of taste parameters into two subvectors b^0 and β^1 of lengths p_0 and p_1 respectively, where $p_0 + p_1 = p$. The vector b^0 contains the fixed (unknown) parameters corresponding to the non-random coefficients of the model, while the vector β^1 captures the random coefficients. Furthermore, we can assume that β^1 is Normally distributed with mean b^1 and variance Σ . The results derived in this paper extend to the case of a model specification with both random and fixed coefficients by performing the integration over the random coefficients while treating the fixed coefficients as constant for the purpose of deriving the Laplace approximation.

We now re-state Proposition 2 for the case with both fixed and random coefficients, $\beta = (b^0, \beta^1)$. The unknown parameters to be estimated are (b^0, b^1, Σ) , where b^1 is the vector of mean parameters of the random coefficients β^1 and Σ is the corresponding covariance matrix of β^1 .

Proposition 2.1: We can approximate $E_\beta(s_{ij}) = E_\beta\left\{\left(\sum_{k=1}^K \exp(x'_{ijk}\beta)\right)^{-1}\right\}$ by

$$E_\beta(s_{ij}) \cong \sqrt{\frac{|\tilde{\Sigma}_{ij}|}{|\Sigma|}} \exp\left\{-\frac{1}{2}(\tilde{\beta}_{ij}^1 - b^1)' \tilde{\Sigma}_{ij}^{-1}(\tilde{\beta}_{ij}^1 - b^1)\right\} \left(\sum_{k=1}^K \exp(x'_{ijk}\tilde{\beta}_{ij})\right)^{-1}, \quad (37)$$

where $\tilde{\beta}_{ij} = (b^0, \tilde{\beta}_{ij}^1)$ and $\tilde{\beta} = (b^0, b^1)$ and

$$\tilde{\beta}_{ij}^1 = b^1 + [\Sigma^{-1} + H_{ij}(b^*)_{b^*=\tilde{b}}]^{-1} J'_{ij}(\tilde{b}) \quad (38)$$

$$\tilde{\Sigma}_{ij}^{-1} = \Sigma^{-1} + H_{ij}(b^*)_{b^*=\tilde{b}_{ij}}, \quad (39)$$

and

$$J_{ij}(b^*) = \sum_{k=1}^K \left\{ x'_{ijk} \frac{\exp(x'_{ijk}b^*)}{\sum_{k=1}^K \exp(x'_{ijk}b^*)} \right\} \quad (40)$$

$$H_{ij}(b^*) = \frac{\sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk}b^*)}{\sum_{k=1}^K \exp(x'_{ijk}b^*)} - \frac{\left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk}b^*) \right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk}b^*) \right]}{\left[\sum_{k=1}^K \exp(x'_{ijk}b^*) \right]^2}. \quad (41)$$

In Section 2 we assert one of the conditions required for the existence of a Laplace approximation with a unique expansion point, the concavity of the function $-g(\beta)$. The Lemma below proves this result.

Lemma: The function $g(\beta)$ is convex, where

$$g(\beta) = \frac{1}{2}(\beta - b)' \Sigma^{-1}(\beta - b) + \log \left(\sum_{k=1}^K \exp(x'_{ijk}\beta) \right) \quad (42)$$

Proof: $g(\beta)$ is the sum of two convex functions, a quadratic form in β and the function $g_1(\beta) = \log \left(\sum_{k=1}^K \exp(x'_{ijk}\beta) \right)$. The Hessian of this function is given by $H_{ij}(\beta)$ defined in equation 21 above. In order to see that $H_{ij}(\beta) \geq 0$ notice that,

$$\left[\sum_{k=1}^K \exp(x'_{ijk}b) \right]^2 H_{ij}(b) = \sum_{k=1}^K \exp(x'_{ijk}b) \sum_{k=1}^K x_{ijk} x'_{ijk} \exp(x'_{ijk}b) - \quad (43)$$

$$- \left[\sum_{k=1}^K x_{ijk} \exp(x'_{ijk}b) \right] \left[\sum_{k=1}^K x'_{ijk} \exp(x'_{ijk}b) \right]. \quad (44)$$

If we expand the right hand side of equation 48 and cancel the terms in $x_{ijk} x'_{ijk} (\exp(x'_{ijk}b))^2$ we can re-arrange this expression as:

$$\left[\sum_{k=1}^K \exp(x'_{ijk}b) \right]^2 H_{ij}(b) = \sum_{r=1}^{K-1} \sum_{s=r+1}^K (x_{ijr} - x_{ijs}) (x_{ijr} - x_{ijs})' \exp(x'_{ijr}b) \exp(x'_{ijs}b) \geq 0. \quad (45)$$

References

- BAJARI, P., H. HONG, AND S. RYAN (2005): “Identification and Estimation of Discrete Games of Complete Information,” Mimeo, MIT.
- BARNDORFF-NIELSEN, O. AND D. R. COX (1979): “Edgeworth and Saddle-point Approximations with Statistical Applications,” *Journal of the Royal Statistical Society B*, 41, 179–312.
- BUTLER, R. AND A. T. A. WOOD (2001): “Laplace Approximations for Hypergeometric Functions of Matrix Argument,” *The Annals of Statistics*, 30, 1155–1177.
- CARDELL, S. AND F. DUNBAR (1980): “Measuring the Societal Impacts of Automobile Downsizing,” *Transportation Research A*, 14, 423–434.
- CHIOU, L. AND J. WALKER (2005): “Identification and Estimation of Mixed Logit Models under Simulation Methods,” Mimeo, MIT.
- DANIELS, H. E. (1954): “Saddlepoint Approximations in Statistics,” *The Annals of Mathematical Statistics*, 25, 631–650.
- DE BRUIJN, N. (1981): *Asymptotic Methods in Analysis*, New York: Dover Publications.
- EFSTATIOU, M., E. GUTIERREZ-PENA, AND A. F. M. SMITH (1998): “Laplace Approximations for Natural Exponential Families with Cuts,” *Scandinavian Journal of Statistics*, 25.
- GLYNN, W. J. (1980): “Asymptotic Representations of the Densities of canonical Correlations and Latent Roots in Manova When the Population Parameters have Arbitrary Multiplicity,” *The Annals of Statistics*, 8, 958–976.
- HARDING, M. C. AND J. A. HAUSMAN (2007): “Flexible Parametric Estimation of the Taste Distribution in Random Coefficients Logit Models,” Mimeo, MIT.
- HAUSMAN, J. A. (1975): “Project Independence Report: A Review of U.S. Energy Needs up to 1985,” *Bell Journal of Economics*.
- HAUSMAN, J. A. AND D. A. WISE (1978): “A Conditional Probit Model for Qualitative Choice-Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences,” *Econometrica*, 403–426.
- HENSHER, D. AND W. GREENE (2003): “The Mixed Logit Model: The State of Practice,” *Transportation*, 133–176.
- HESS, J. AND J. POLAK (2005): “Mixed Logit Modelling of Airport Choice in Multi-Airport Regions,” *Journal of Air Transport Management*, 59–68.

- HSU, L. C. (1948): “A Theorem on the Asymptotic Behavior of a Multiple Integral,” *Duke Mathematical Journal*, 623–632.
- JENSEN, J. L. (1995): *Saddlepoint Approximations*, Oxford: Clarendon Press.
- LERMAN, S. AND C. MANSKI (1981): “On the Use of Simulated Frequencies to Approximate Choice Probabilities,” in *Structural Analysis of Discrete Data with Econometric*, ed. by C. Manski and D. McFadden, Cambridge: MIT Press, 305–320.
- MANSKI, C. AND D. MCFADDEN, eds. (1981): *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: MIT Press.
- MCFADDEN, D. (1974): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, Amsterdam: Academic Press, 75–96.
- (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration,” *Econometrica*, 995–1026.
- MCFADDEN, D. AND K. TRAIN (2000): “Mixed MNL Models for Discrete Response,” *Journal of Applied Econometrics*, 447–470.
- MUIRHEAD, R. J., ed. (2005): *Aspects of Multivariate Statistical Theory*, New York: Wiley.
- MURRAY, J., ed. (1984): *Asymptotic Analysis*, New York: Springer-Verlag.
- PAKES, A. AND D. POLLARD (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 1027–1057.
- REVELT, D. AND K. TRAIN (1998): “Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level,” *Review of Economics and Statistics*, 647–657.
- SMALL, K. A., C. WINSTON, AND J. YAN (2005): “Uncovering the Distribution of Motorists’ Preferences for Travel Time and Reliability,” *Econometrica*, 1367–1382.
- TIERNEY, L. AND J. KADANE (1986): “Accurate Approximations for Posterior Moments and Marginal Densities,” *Journal of the American Statistical Association*, 82–86.
- TRAIN, E. K. (2003): *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

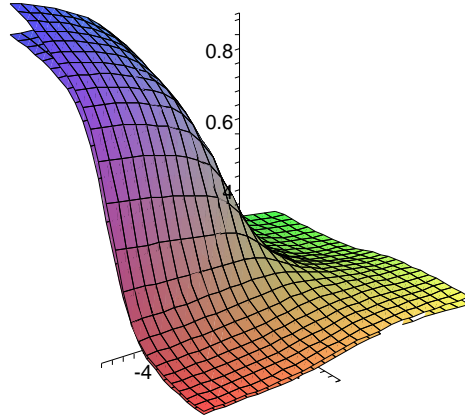


Figure 1: Comparison of expected share obtained by numerical integration and the corresponding Laplace approximation for a model with 2 covariates at fixed values of b and Σ .

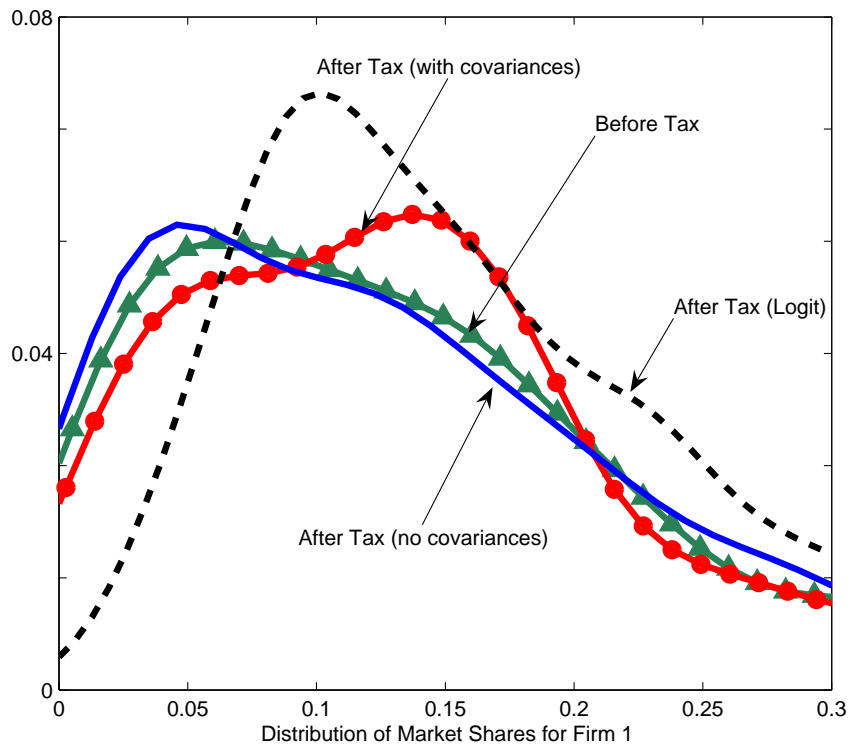


Figure 2: Market shares of Firm 1 before and after tax