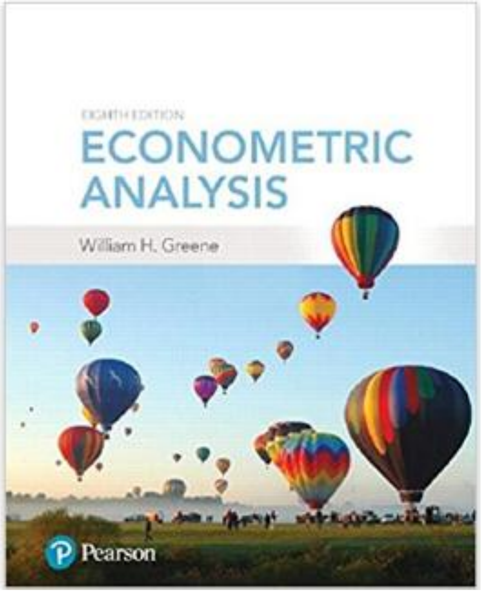
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 23 – Simulation Based Estimation

Settings

- Conditional and unconditional log likelihoods
  - Likelihood function to be maximized contains unobservables
  - Integration techniques
- Bayesian estimation
  - Prior times likelihood is intractible
  - How to obtain posterior means, which are open form integrals
- The problem in both cases is "…how to do the integration?"

A Conditional Log Likelihood

Conditional (on random v) density,

$$f(y_i \mid \theta, x_i, v_i)$$

Unconditional density:

$$f(y_i \mid \theta, \alpha\ x_i) = \int_{v_i} f(y_i \mid \theta, x_i, v_i) h(v_i \mid \alpha) dv_i$$

Log likelihood function

$$\text{log-L}(\theta, \alpha) = \sum_{i=1}^{n} \log \int_{v_i} f(y_i \mid \theta, x_i, v_i) h(v_i \mid \alpha) dv_i$$

Integral does not exist in closed form. How to do the maximization?

Application - Innovation

- Sample = 1,270 German Manufacturing Firms
- Panel, 5 years, 1984-1988
- Response: Process or product innovation in the survey year? (yes or no)
- Inputs:
  - Imports of products in the industry
  - Pressure from foreign direct investment
  - Other covariates
- Model: Probit with common firm effects
- (Irene Bertschuk, doctoral thesis, Journal of Econometrics, 1998)

Likelihood Function for Random Effects

- Joint conditional (on $u_i=\sigma v_i$) density for obs. i.

$$f(y_{i1},\ldots,y_{iT}\,|\,v_i)=\prod_{t=1}^{T} g(y_{it}\,|\,v_i)=\prod_{t=1}^{T}\Phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\sigma v_i)]$$

- Unconditional likelihood for observation i

$$L_i=\int_{u_i}\prod_{t=1}^{T} g(y_{it}\,|\,v_i)h(v_i)dv_i$$

- How do we do the integration to get rid of the heterogeneity in the conditional likelihood?

Obtaining the Unconditional Likelihood

- The Butler and Moffitt (1982) method is used by most current software
  - Quadrature (Stata –GLAMM)
  - Works only for normally distributed heterogeneity

Hermite Quadrature

$$\int_{-\infty}^{\infty} f(x,v)\exp(-v^2)dv \approx \sum_{h=1}^{H} f(x,v_h)W_h$$

Adapt to integrating out a normal variable

$$f(x) = \int_{-\infty}^{\infty} f(x,v)\frac{\exp(-\frac{1}{2}(v/\sigma)^2)}{\sigma\sqrt{2\pi}}dv$$

Change the variable to $z = (1/(\sigma\sqrt{2}))v$,

$v = (\sigma\sqrt{2})z$ and , $dv=(\sigma\sqrt{2})dz$

$$f(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} f(x,\lambda z)\exp(-z^2)dz,\ \lambda=\sigma\sqrt{2}$$

This can be accurately approximated by Hermite quadrature

$$f(x) \approx \sum_{h=1}^{H} f(x,\lambda z)W_h$$

Example: 8 Point Quadrature

Nodes for 8 point Hermite Quadrature
Use both signs, + and -

0.381186990207322000,
1.15719371244677990
1.98165675669584300
2.93063742025714410

Weights for 8 point Hermite Quadrature

0.661147012558199960,
0.207802325814899990,
0.0170779830074100010,
0.000199604072211400010

Butler and Moffitt's Approach
Random Effects Log Likelihood Function

$$\log L = \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{T} g\left[y_{it}, \left(\mathbf{x}_{it}'\boldsymbol{\beta}^0 + v_i \right) \right] \right\} h(v_i) dv_i$$

Butler and Moffitt: Compute this by Hermite quadrature

$$\int_{-\infty}^{\infty} f(v_i)h(v_i)dv_i \approx \sum_{h=1}^{H} f(z_h)w_h \text{ when } \mathrm{h}(v_i) = \text{normal density}$$

z_h = quadrature node; w_h = quadrature weight

$z_i = \sigma v_i, \sigma$ is estimated with $\boldsymbol{\beta}^0$

The Simulated Log Likelihood

$$\log L = \sum_{i=1}^{N} \log \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{T} g\left[y_{it}, \left(\mathbf{x}_{it}'\boldsymbol{\beta}^0 + \sigma v_i \right) \right] \right\} \phi(v_i) dv_i$$

where v_i is the normally distributed effect.

Use the law of large numbers:

let $v_{i1},...,v_{iR}$ = a random sample of R draws from the standard normal population.

$$\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T} g\left[y_{it}, \left(\mathbf{x}_{it}'\boldsymbol{\beta}^0 + \sigma v_{iR} \right) \right] \xrightarrow{P} \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^{T} g\left[y_{it}, \left(\mathbf{x}_{it}'\boldsymbol{\beta}^0 + \sigma v_i \right) \right] \right\} \phi(v_i) dv_i$$

Monte Carlo Integration

$$\frac{1}{R}\sum_{r=1}^{R} f(u_{ir}) \xrightarrow{P} \int_{u_i} f(u_i)g(u_i)du_i = E_{u_i}[f(u_i)]$$

(Certain smoothness conditions must be met.)

Drawing u_{ir} by 'random sampling'

$u_{ir} = t(v_{ir}),\ v_{ir} \sim U[0,1]$

E.g., $u_{ir} = \sigma\Phi^{-1}(v_{ir}) + \mu\ for\ N[\mu,\sigma^2]$

Requires many draws, typically

hundreds or thousands

Generating Random Draws

Most common approach is the "inverse probability transform"

Let u = a random draw from the standard uniform (0,1).

Let x = the desired population to draw from

Assume the CDF of x is $F(x)$.

The random draw is then $x = F^{-1}(u)$.

Example: exponential, θ. $f(x)=\theta\exp(-\theta x)$, $F(x)=1-\exp(-\theta x)$

Equate u to $F(x)$, $x = -(1/\theta)\log(1-u)$.

Example: Normal(μ,σ). Inverse function does not exist in closed form. There are good polynomial approximations to produce a draw from N[0,1] from a U(0,1). Then $x = \mu+\sigma v$.

Drawing Uniform Random Numbers

Computer generated random numbers are not random; they are Markov chains that look random.

The Original IBM SSP Random Number Generator for 32 bit computers.

SEED originates at some large odd number

d3 = 2147483647.0 $(2^{31} - 1)$

d2 = 2147483655.0 $(2^{31} + 7)$

d1=16807.0 (a strange number)

SEED=Mod(d1*SEED,d3) ! MOD(a,p) = a - INT(a/p) * p

X=SEED/d2 is a pseudo-random value between 0 and 1.

Problems:

(1) Short period. Based on 32 bits, so recycles after $2^{31} - 1$ values

(2) Evidently not very close to random. (Recent tests have discredited this RNG)

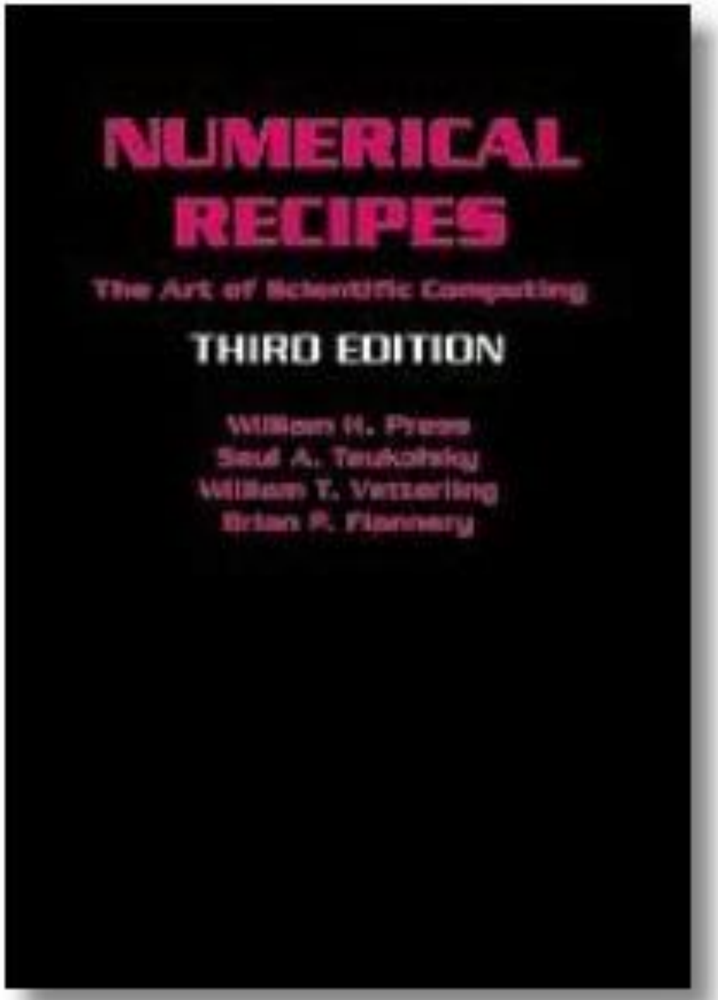
(3) Current state of the art is the Mersenne Twister. (Default in R, Matlab, etc.)

Period = 2^{20000} Passes (DieHard) randomness tests

Poisson with mean = 4.1
Table

X=Cases	Prob[X=x]	Prob[X<=x]
0	0.016573	0.01657
1	0.067948	0.08452
2	0.139293	0.22381
3	0.190368	0.41418
4	0.195127	0.60931
5	0.160004	0.76931
6	0.109336	0.87865
7	0.064040	0.94269
8	0.032820	0.97551
9	0.014951	0.99046
10	0.006130	0.99659
12	0.000781	0.99966
13	0.000246	0.99990
14	0.000072	0.99997
15	0.000020	0.99999
16	0.000005	1.00000

Uniform Draw = .72159
Poisson Draw = 4


NUMERICAL
RECIPES
The Art of Scientific Computing
THIRD EDITION
William H. Press
Saul A. Teukolsky
William T. Vetterling
Brian P. Flannery

Quasi-Monte Carlo Integration Based on Halton Sequences

Coverage of the unit interval is the objective,

not randomness of the set of draws.

Halton sequences --- Markov chain

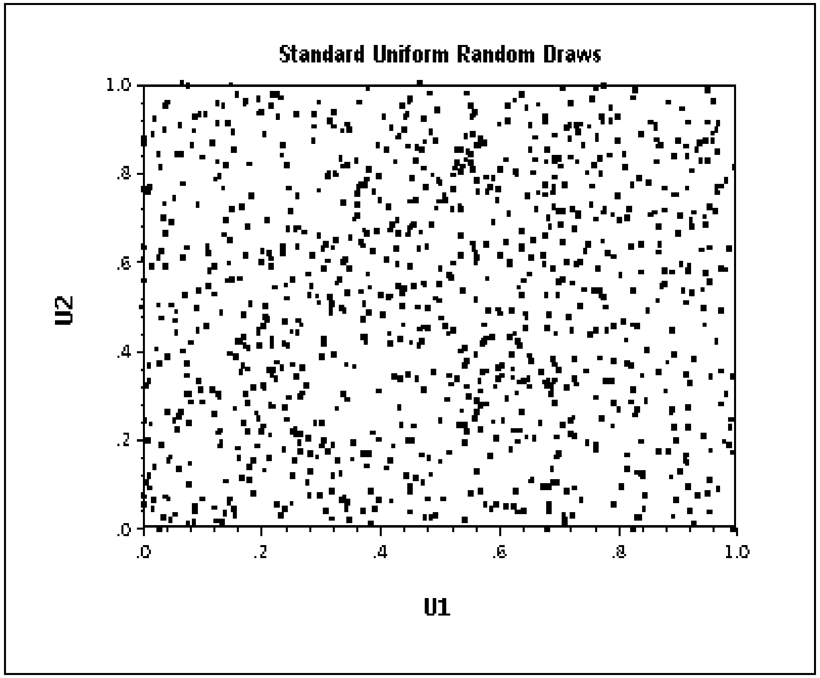
p = a prime number,

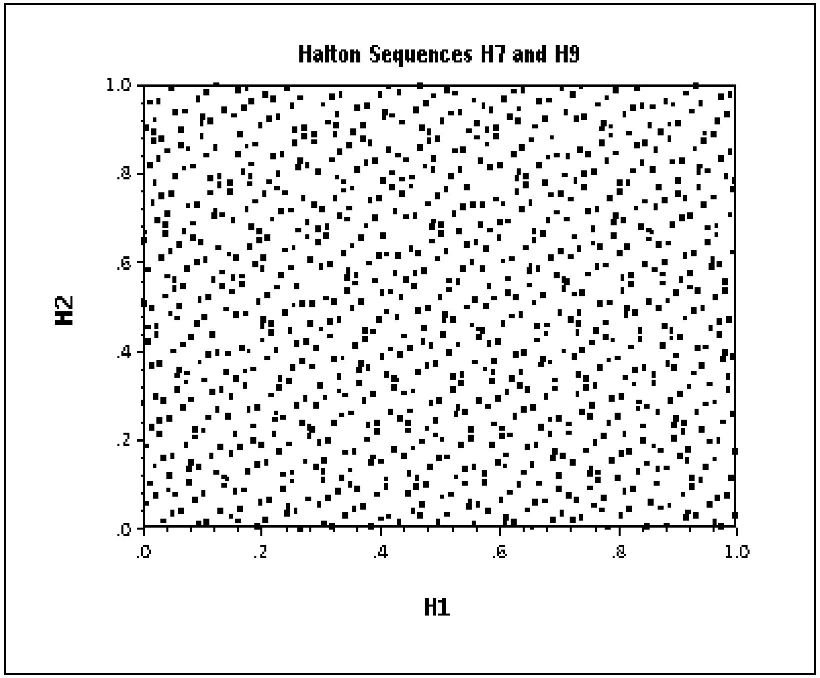
r= the sequence of integers, decomposed as $\sum_{i=0}^{I} b_i p^i$

H(r|p) $= \sum_{i=0}^{I} b_i p^{-i-1}$, r = r_1,... (e.g., 10,11,12,...)

For example, using base p=5, the integer r=37 has b_0 = 2, b_1 = 2, and b_2 = 1; ($37 = 1\text{x}5^2 + 2\text{x}5^1 + 2\text{x}5^0$). Then H(37|5) = $2\times5^{-1} + 2\times5^{-2} + 1\times5^{-3}$ = 0.448.

Halton Sequences vs. Random Draws





Requires far fewer draws – for one dimension, about 1/10. Accelerates estimation by a factor of 5 to 10.

Panel Data Estimation
A Random Effects Probit Model

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} + u_i, \; t = 1,...,T, \; i = 1,...,N,$$

$$y_{it} = \mathbf{1}(y_{it}^* > 0), \; \text{(observation mechanism)}$$

$$[\varepsilon_{i1}, \varepsilon_{i2},...,\varepsilon_{iT}]' \sim N[\mathbf{0},\mathbf{I}], \; \boxed{u_i \sim N[0,\sigma^2] \perp (\mathbf{x}_{it}, \varepsilon_{it})}$$

$$Var[...] = (1+\sigma^2)\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \rho & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{bmatrix}, \; \rho = \frac{\sigma^2}{1+\sigma^2}$$

Log Likelihood

$$\log L(\beta,\sigma) = \sum_{i=1}^{n} \log \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\sigma v_i)]\phi(v_i)dv_i$$

$$\rho = \frac{\sigma^2}{1+\sigma^2}$$

Quadrature

$$\log L(\beta,\sigma) \approx \sum_{i=1}^{n} \log \sum_{h=1}^{H} \prod_{t=1}^{T} W_h \Phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\sigma z_h)]$$

$\mathrm{W_h}$ = quadrature weight, $\mathrm{z_h}$ = quadrature node

Simulated

$$\log L(\beta,\sigma) \approx \sum_{i=1}^{n} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \Phi[(2y_{it}-1)(\boldsymbol{\beta}'\mathbf{x}_{it}+\sigma \hat{v}_{ir})]$$

$\hat{\mathrm{v}}_{\mathrm{ir}}$ = rth draw from standard normal for individual i.

$(\hat{\mathrm{v}}_{\mathrm{i1}},...,\hat{\mathrm{v}}_{\mathrm{iR}})$ are reused for all computations of function or derivatives.

Application: Innovation

Bertschek and Lechner applied the GMM estimator to an analysis of the product innovation activity of 1,270 German manufacturing firms observed in five years, 1984 - 1988, in response to imports and foreign direct investment. [See Bertschek (1995).] 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,
$x_{3,it}$ = Import share = ratio of industry imports to (industry sales plus imports),
$x_{4,it}$ = Relative firm size = ratio of employment in business unit to employment in the industry (times 30),
$x_{5,it}$ = FDI share = Ratio of industry foreign direct investment to (industry sales, 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,

Application: Innovation

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

$i = 1,...,1270$, $t = 1984,...,1988$.

y_{it} = 1 if a product innovation was realized by German manufacturing firm i in year t, 0 otherwise,

$x_{2,it}$ = Log of industry sales in DM,

$x_{3,it}$ = Import share = ratio of industry imports to (industry sales plus imports),

$x_{4,it}$ = Relative firm size = ratio of employment in business unit to employment in the industry (times 30),

$x_{5,it}$ = FDI share = Ratio of industry foreign direct investment to (industry sales, 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

Table 2. Estimated Random Effects Models

	Random Effects			
	Quadrature Estimator		Simulation Estimator	
Variable	Estimate	Std.Error	Estimate	Std.Error
Constant	-2.839**	0.533	-2.884**	0.543
log Sales	0.244**	0.0522	0.249**	0.0510
Rel Size	1.522**	0.257	1.452**	0.281
Imports	1.779**	0.360	1.796**	0.360
FDI	3.652**	0.870	3.724**	0.831
Prod.	-2.307	1.911	-2.321**	0.151
Raw Mtl	-0.477*	0.202	-0.469*	0.186
Inv Good	0.331**	0.0952	0.331**	0.0915
ρ	0.578**	0.0189	0.578**[a]	0.0231

[a] Based on estimated standard deviation of the random constant of 1.1707 with estimated standard error of 0.01865. **$(1.1707^2 / (1 + 1.1707^2) = 0.578)$**

* Indicates significant at 95% level, ** indicates significant at 99% level based on a two tailed test.

Quadrature vs. Simulation

- Computationally, comparably difficult
- Numerically, essentially the same answer. MSL is consistent in R
- Advantages of simulation
  - Can integrate over any distribution, not just normal
  - Can integrate over multiple random variables. Quadrature is largely unable to do this.
  - Models based on simulation are being extended in many directions.
  - Simulation based estimator allows estimation of conditional means → essentially the same as Bayesian posterior means

A Random Parameters Model

$$\text{Prob(Innovation)} = \Phi(\beta_{1i}\text{FDI} + \beta_{2i}\text{Imports}$$
$$\beta_3 + \beta_4\text{logSales} + \beta_5\text{Employment} + \beta_6\text{Productivity})$$

$$\begin{bmatrix}\beta_{1i}\\ \beta_{2i}\end{bmatrix} \sim N\left[\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right]$$

and four fixed (nonrandom) parameters.

$$\log L(\beta,\sigma) = \sum_{i=1}^{n} \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty}$$

$$\prod_{t=1}^{T} \Phi[(2y_{it}-1)(\boldsymbol{\beta}_0'\mathbf{x}_{it}^0 + (\beta_1+\sigma_1 v_{i1})x_{it}^1 + (\beta_2+\sigma_2 v_{i2})x_{it}^2)]\phi(v_{i2})\phi(v_{i1})dv_{i2}dv_{i1}$$

$$\rho = \frac{\sigma^2}{1+\sigma^2}$$

Simulated

$$\log L(\beta,\sigma) \approx \sum_{i=1}^{n} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} G(v_{i1,r} v_{i2,r})$$

$\hat{v}_{ik,r}$ = rth draw from standard normal for individual i and variable k.

$(\hat{v}_{i1,r},...,\hat{v}_{i1,R},...)$ are reused for all computations of function or derivatives.

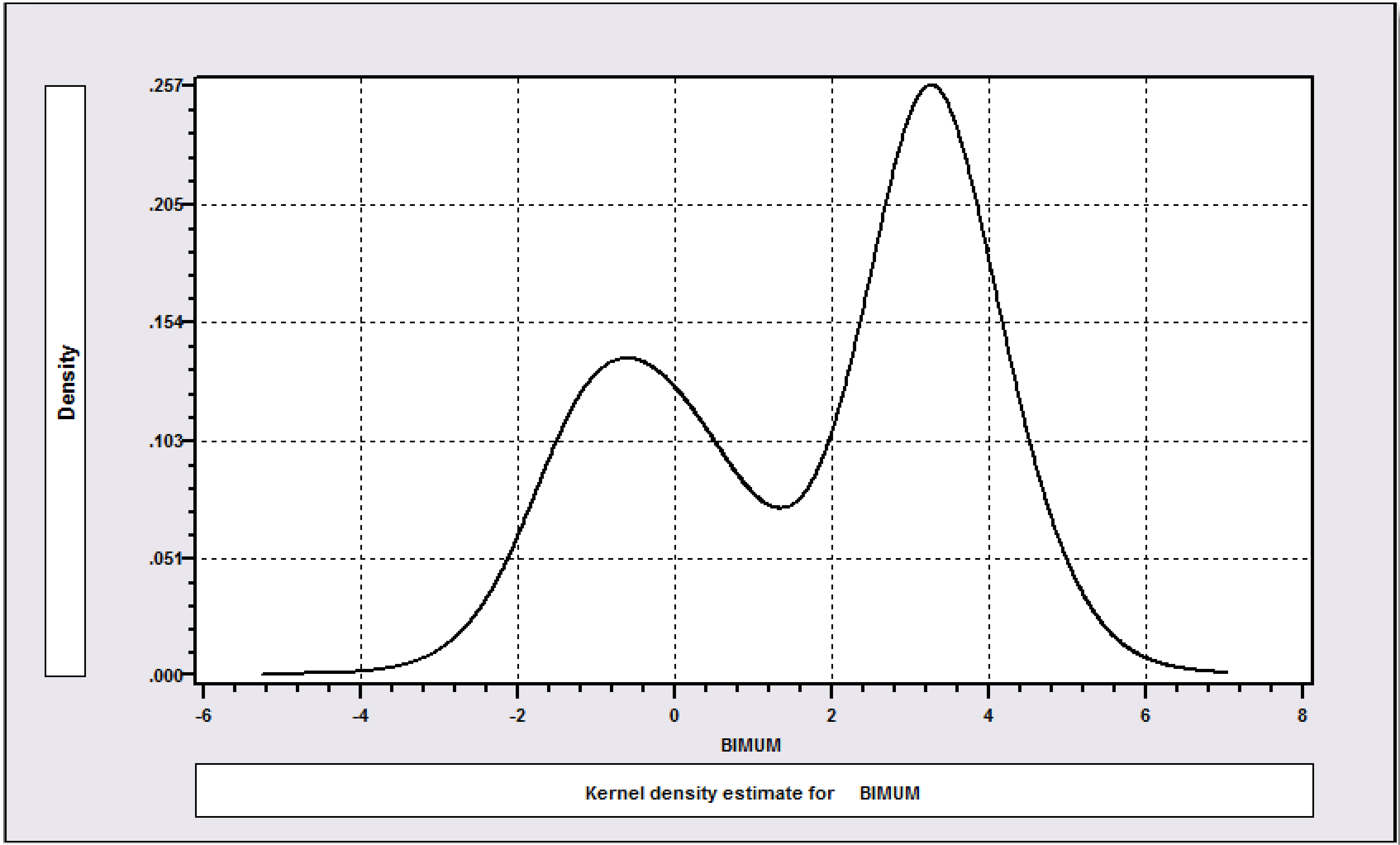
Estimates of a Random Parameters Model

```
----------------------------------------------------------------------
Probit   Regression Start Values for IP
Dependent variable                     IP
Log likelihood function      -4134.84707
Estimation based on N =   6350, K =   6
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.30420     8281.69414
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
Constant|   -2.34719***       .21381      -10.978   .0000
   FDIUM|    3.39290***       .39359        8.620   .0000      .04581
    IMUM|     .90941***       .14333        6.345   .0000      .25275
LOGSALES|     .24292***       .01937       12.538   .0000     10.5401
      SP|    1.16687***       .14072        8.292   .0000      .07428
    PROD|   -4.71078***       .55278       -8.522   .0000      .08962
--------+-------------------------------------------------------------
```

RPM

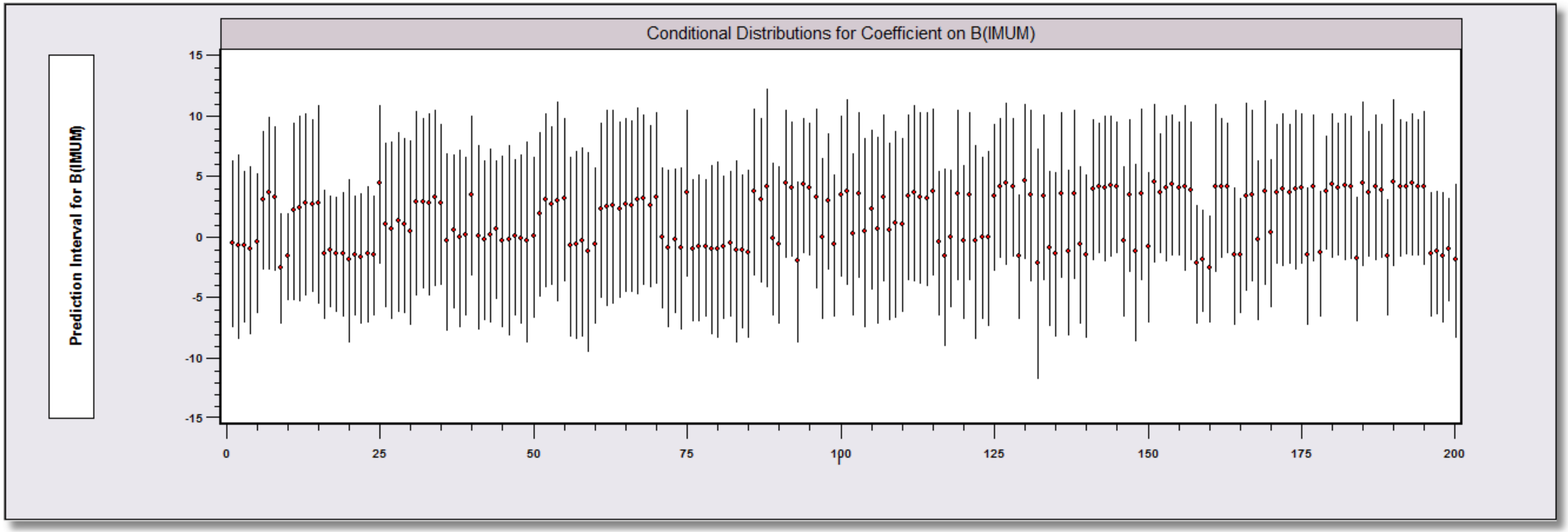
```
Random Coefficients  Probit   Model
Dependent variable                  IP
Log likelihood function     -4113.14958
Restricted log likelihood   -4134.84707
Chi squared [   2 d.f.]       43.39499
Significance level              .00000
McFadden Pseudo R-squared      .0052475
Estimation based on N =   6350, K =   8
Inf.Cr.AIC  =   8242.3 AIC/N =    1.298
Sample is  1 pds and   6350 individuals
PROBIT (normal)  probability model
```

IP	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
Constant	-2.80391***	.24985	-11.22	.0000	-3.29360	-2.31422
LOGSALES	.28298***	.02276	12.43	.0000	.23837	.32760
SP	2.06212***	.14857	13.88	.0000	1.77093	2.35330
PROD	-8.16188***	.36419	-22.41	.0000	-8.87569	-7.44808
	Means for random parameters					
FDIUM	9.78159***	.65708	14.89	.0000	8.49373	11.06945
IMUM	1.82017***	.18017	10.10	.0000	1.46704	2.17329
	Scale parameters for dists. of random parameters					
FDIUM	11.9585***	.48085	24.87	.0000	11.0160	12.9009
IMUM	4.01600***	.11124	36.10	.0000	3.79798	4.23403



Density
.257
.205
.154
.103
.051
.000
-6
-4
-2
0
2
4
6
8
BIMUM
Kernel density estimate for BIMUM

Parameter Heterogeneity



CNNMoney Try Money FREE | Subscribe to Money | QUOTE | SYMBOL LOOK-UP | SEARCH | Web | CNN

Home | News | Markets | Technology | Commentary | Personal Finance

NEWS > Fun Money

SAVE | EMAIL | PRINT | SUBSCRIBE TO MONEY | RSS

Hollywood exports: helped by McDonald's?

Study: U.S. films do better overseas in countries with more of the fast-food chain's outlets.

December 5, 2005: 2:25 PM EST

NEW YORK (CNNMoney.com) - If Hollywood wants its films to burn up box offices overseas, it may want to keep the hamburger house McDonald's in mind, according to a study released Monday.

The study, produced by three professors at New York University's Stern School of Business, found that U.S. movies tended to have higher box office sales in countries with more "Golden Arches," taking income levels and population into account.

Using McDonald's as a measure of 'Americanization', professors C. Samuel Craig, Susan Douglas and William Greene looked at box office sales for the 50 top U.S. films from 1997 to 2002 in eight countries including Australia, Spain, Mexico, Germany and Argentina.

Included in the study's other findings was that certain film genres performed better than others, as action, animated and horror films typically outperformed family films.

The research, which was released in a statement Monday by NYU's Stern School of Business, is scheduled to be published in an upcoming issue of the Journal of International Marketing.

Fun Money
See more stories

Movie Model

The hypotheses involve variables at two levels, film and country. In addition, each film is unique and consequently it was also desirable to be able to account for film-specific heterogeneity. Consequently, a hierarchical random parameters regression model was formulated as follows:

$$B_{f,c} = \alpha_f + \beta_f B_{f,US} + \gamma_1 CD_c + \gamma_2 MACSPC_c + \gamma_3 English_c$$
$$+ \delta_{98} D_{1998} + \ldots + \delta_{02} D_{2002} + \sum_{g=1}^{12} \eta_g G_{f,g} + \varepsilon_{f,c}$$
$$\alpha_f = \alpha_0 + \alpha_1 \text{logIncome}_c + u_{\alpha,f} \qquad (1)$$

$$\beta_f = \beta_0 + \beta_1 \text{logIncome}_c + u_{\beta,f}$$

where "f" denotes film, $f = 1,\ldots,F = 299$ and "c" denotes country = UK, Australia, Germany, Austria, Argentina, Chile, and Mexico. In the primary equation, $B_{f,c}$ is the log of the per capita box office revenues for film f in country c; $B_{f,US}$ is the same for the United States. CD_c is our measure of the cultural distance of country j from the US. $MACSPC_c$ is the number of McDonald's restaurants per capita in country c in 2000; $English_c$ is a dummy variable for whether the country is an English-speaking country (UK, Australia). The variables $G_{f,g}$ are 12 dummy variables for primary genre; the 13th, Crime, is fixed as the basis. The disturbance, $\varepsilon_{f,c}$ is assumed to be normally distributed with mean zero and constant variance σ^2.