



E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Shrinkage Estimation in the Random Parameters
Logit Model

Tong Zeng
Georgia Southern University

R. Carter Hill
Louisiana State University

Working Paper 2014-11
http://faculty.bus.lsu.edu/workingpapers/pap14_11.pdf

*Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
<http://www.bus.lsu.edu/economics/>*

Shrinkage Estimation in the Random Parameters Logit Model

Tong Zeng ^{a,*1} and R. Carter Hill^{b,†1}

^aFinance and Economics Department, Georgia Southern University, Statesboro, GA 30458

^bDepartment of Economics, Louisiana State University, Baton Rouge, LA 70803

Abstract

We explore the properties of a Stein-like shrinkage estimator that combines the fully correlated and uncorrelated Random Parameters Logit model (RPLM). Monte Carlo experiments show that shrinkage and pretest estimators can improve upon the fully correlated RPLM estimator.

Keywords: Pretest Estimator, Positive-part Stein-like Estimator, Likelihood Ratio Test, Random Parameters Logit Model

JEL Classification: C12, C13, C35

1 Introduction

The random parameters logit model (RPLM) is a generalization of the conditional logit model for multinomial choices. The conditional logit model is derived from

*Corresponding author, Email: tzeng@georgiasouthern.edu, Tel: +1 912 478 1842

†Email: eohill@lsu.edu, Tel: +1 225 578 1490

an assumption that the errors in the underlying random utility functions for each choice alternative are statistically independent and identically distributed (iid) extreme value type I. This leads to the property known as the Independence of Irrelevant Alternatives (IIA): The ratio of the probability of two alternatives remains constant no matter how many choices there are. This is widely regarded to be a very restrictive assumption.

The key feature of the RPLM is that response parameters can vary randomly, following a chosen distribution, across the population from which samples are drawn. The random coefficients capture individual heterogeneity and the model does not suffer from the IIA assumption. The random coefficients can be correlated in the RPLM as generally expected in reality, because the unobservable preference of each individual is used to evaluate the attributes of all alternatives in each choice situation. Estimation is by maximum simulated likelihood (MSL), which is described by Greene (2012, 603-654).

In this paper we explore a problem that can exist in any correlated random parameters model. Let y_n , $n = 1, \dots, N$ be an observable outcome variable from a density $f(y_n|x_n, \beta_n)$, where x_n is a vector of K explanatory variables and β_n are random parameters with mean $\bar{\beta}$ and covariance matrix Σ . Using MSL we estimate the population parameters $\bar{\beta}$ and Σ . Allowing the random parameters to be correlated introduces potentially many new parameters, $K(K-1)/2$ covariance terms, that are difficult to estimate.

Most applied researchers will test the significance of the covariance parameters before deciding to rely on the fully correlated random parameter model instead the model in which the parameters are random but uncorrelated. We explore whether a pretesting strategy improves postestimation inference. We also explore the use of a Stein-like shrinkage estimator as an alternative to pretesting. This estimator shrinks the estimates from the fully correlated random parameter model towards the estimates of the uncorrelated random parameter model. In numerical experiments using the RPLM we find that both the pretest estimator and shrinkage estimators

have improved mean squared error (MSE) relative to the MSL estimator of the fully correlated parameter model. Last, we analyze the share of the population putting a positive or negative value on the alternative attributes, and the Monte Carlo mean estimates of direct elasticity with fully correlated RPLM estimates and pretest and shrinkage estimates. Based on our Monte Carlo experiment results, pretest and shrinkage estimates provide more accurate estimates on both of them than the fully correlated RPLM estimates.

2 The Random Parameters Logit Model

The RPLM is described in Train (2009, 134-150). Consider individual n facing M alternatives. The random utility associated with alternative i is $U_{ni} = \beta'_n x_{ni} + \varepsilon_{ni}$, where x_{ni} are K observed explanatory variables for alternative i , ε_{ni} is an iid type I extreme value error which is independent of β_n and x_{ni} . The random coefficients β_n can be regarded as being composed of a mean $\bar{\beta}$ and deviations $\tilde{\beta}_n$. The RPLM decomposes the unobserved part of the utility into the extreme value term ε_{ni} and the random part $\tilde{\beta}'_n x_{ni}$. Conditional on β_n the probability that individual n chooses alternative i is of the usual logistic form, $L_{ni}(\beta_n) = e^{\beta'_n x_{ni}} / \sum_i e^{\beta'_n x_{ni}}$. Assume that β_n is multivariate normal¹ with mean vector $\bar{\beta}' = (\bar{\beta}_1, \dots, \bar{\beta}_k)$ and covariance matrix Σ with elements σ_{jk} . Denoting the MVN density $f(\beta|\theta)$, where θ contains the unknown mean and covariance parameters, the probability that individual n chooses alternative i is

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta \quad (1)$$

For estimation purposes we use Cholesky's decomposition and write $\Sigma = AA'$, where A is lower triangular. The parameter means $\bar{\beta}_k$ and elements of A are the objects of estimation. The parameters of the fully correlated RPLM (FCRPLM),

¹Other choices are possible. See Train (2009, 136)

are

$$\theta_f = (\bar{\beta}_1, \dots, \bar{\beta}_k, a_{11}, \dots, a_{kk}, a_{21}, \dots, a_{k,k-1}) \quad (2)$$

where a_{kk} are diagonal elements of A and a_{jk} , $j < k$, are below the diagonal. If the random coefficients in the RPLM are uncorrelated, denoted UCRPLM, then θ is

$$\theta_u = (\bar{\beta}_1, \dots, \bar{\beta}_k, a_{11}, \dots, a_{kk}) \quad (3)$$

where $\sigma_k^2 = a_{kk}^2$.

3 Stein-Like Shrinkage Estimation

The positive-part Stein-like estimator (θ^+) is a stochastically weighted convex combination of the MLE from an unrestricted model and a restricted MLE subject to J constraints. In our case the unrestricted MLE comes from the FCRPLM estimates ($\hat{\theta}_f$) and the restricted MLE from the UCRPLM estimates ($\hat{\theta}_u$)

$$\theta^+ = c\hat{\theta}_u + (1 - c)\hat{\theta}_f \quad (4)$$

where $c = 1 - I_{(a, \infty)}(u)(1 - a/u)$ and $I_{(a, \infty)}(u)$ is the indicator function of a test statistic u for the null hypothesis that the coefficient covariance matrix is diagonal, or equivalently that the Cholesky elements in A below the diagonal are zero. The constant a controls the amount of shrinkage towards the UCRPLM estimates. The shrinkage estimator θ^+ becomes the UCRPLM estimator θ_u when the test statistic u is less than the value of a . The larger the value of a , the more weight that is given to the UCRPLM estimates. Kim and Hill (1995) show that if the number of constraints $J > 2$, then under information weighted quadratic loss the risk of the shrinkage estimator is smaller than the risk of the unrestricted maximum likelihood estimator for any $c > 0$. Common choices for the shrinkage constant are $a = 2(J - 2)$

and $a = J - 2$. In our case $J = K(K - 1)/2$ is the number of covariance terms constrained to zero when obtaining the UCRPLM estimates.

With test statistic u , the pretest estimator θ^* is:

$$\theta^* = \begin{cases} \theta_u & \text{if } u \leq c_\alpha \\ \theta_f & \text{if } u > c_\alpha \end{cases} \quad (5)$$

where c_α is the critical value of chi-square distribution with J degrees of freedom and significance level α . With the given of degrees of freedom, the critical value c_α is determined by the level of test significance α , which is between 0 and 1. When $\alpha = 0$, pretest estimator θ^* becomes UCRPLM estimator θ_u . When $\alpha = 1$, pretest estimator θ^* is FCRPLM estimator θ_f .

4 Monte Carlo Experiments

4.1 Design

In our experiments the number of choice alternatives is $M = 4$ and the number of individuals is $N = 200$. Each individual is assumed to be observed once. The four explanatory variables for each individual and each alternative x_{ni} are generated from independent log-normal distributions $\ln N(1, 0.25)$. The coefficients for each individual β_n are generated from multivariate normal distribution $N(\bar{\beta}, \Sigma)$, with $\bar{\beta}_k = 1$, $k = 1, \dots, 4$. The variance of each random coefficient is $\sigma_k^2 = 1$, $k = 1, \dots, 4$. The covariance elements $\sigma_{jk} = \rho$, $j, k = 1, \dots, 4$. The correlation ρ takes the values 0, 0.2, 0.4, 0.6, 0.8. The values of x_{ni} and β_n are held fixed over the $N_{SAM} = 1000$ Monte Carlo samples in each experiment.

Individual choices y_{ni} are determined by comparing the utility of each alternative:

$$y_{ni} = \begin{cases} 1 & \beta'_n x_{ni} + \varepsilon_{ni} > \beta'_n x_{nj} + \varepsilon_{nj} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Our simulation and RPLM estimation were carried out in NLOGIT 5.0. Based on our Monte Carlo experiment results and Bhat(2001), we use 100 Halton draws to simulate choice probabilities during MSL estimation. The positive-part Stein-like and pretest estimators were calculated based on the likelihood ratio (LR), Lagrange multiplier (LM) and Wald test statistics with 25%, 5% and 1% significance level. Because the empirical percentile values of LR test are closer to the related critical values than those of LM and Wald tests, we only provide the results based on the LR test statistic. Using Monte Carlo experiments to study the RPLM, especially with correlated parameters, is numerically challenging. Key elements that are worth mentioning are (i) for the uncorrelated parameter model conditional logit estimates were used as starting values; (ii) for the correlated parameter model the estimates from (i) were used as starting values; (iii) samples for which convergence was not achieved were discarded, only 0.3% of the results are unconverged in our Monte Carlo experiments.

4.2 Results

To study how the pretest and shrinkage estimators reduce the estimation risk of the FCRPLM estimators, we calculate the MSEs of the estimated parameters mean, variance, covariance with the pretest, shrinkage and FCRPLM estimators respectively. First, we compare the MSE of the fully correlated estimators and those of UCRPLM estimators, where MSE is the Monte Carlo average of the squared error loss $\sum_{k=1}^4 (\hat{\beta}_k - \bar{\beta}_k)^2$. In Table 1, the MSEs of UCRPLM estimators are all smaller than those of FCRPLM estimators. The risk of the estimated parameters mean with the FCRPLM is more than twice that of the UCRPLM. The MSEs of the estimated variance with the UCRPLM are about 25% of those with the FCRPLM. With nonzero correlation ρ , the MSEs of estimated covariance parameters based on the FCRPLM are much bigger than those based on the UCRPLM. When $\rho = 0.2$ and 0.4, the ratios of MSEs of estimated covariance elements are relatively smaller compared to the results for higher correlations. This implies that when the specifi-

cation error is small, the FCRPLM, which is the correct model, has a much larger relative MSE for parameter covariance elements than the UCRPLM.

Table 1: The Ratios of UCRPLM Estimator MSE to the FCRPLM Estimator MSE

ρ	$\hat{\beta}$	$\hat{\sigma}_{kk}$	$\hat{\sigma}_{jk}$
0.0	0.454	0.217	0.000
0.2	0.450	0.251	0.010
0.4	0.385	0.221	0.025
0.6	0.327	0.141	0.044
0.8	0.329	0.231	0.091

In Table 2, we compare the MSEs of LR based pretest and shrinkage estimators to those of FCRPLM estimators. All Table 2 ratios are less than one. The pretest and shrinkage estimators all perform better than the FCRPLM estimators. With a smaller level of test significance α , the UCRPLM estimator θ_u is more frequently chosen as the pretest estimator and the pretest estimator has smaller MSE. However, compared to the shrinkage estimators, the LR based pretest estimators with $\alpha = 0.01$ have larger MSEs than the shrinkage estimators with the shrinkage constant $a = 2(J - 2) = 8$, especially for the estimated covariance elements, which have the smallest ratio values.

The covariance elements reveal important information about the joint effect of alternative attributes on people' decisions. Two highly positively correlated random coefficients imply that people are attracted and motivated by both of the related attributes. In our Monte Carlo experiments, the shrinkage estimators with higher shrinkage constant a outperform estimators with less shrinkage and most of the pretest estimators.

Since one of the advantages of RPLM is providing the information on the share of population that places a positive or negative value on the alternative attributes, we also calculate the joint probability of the first two estimated parameters are less than zero. Compared to the results with UCRPLM and FCRPLM estimates in Table 3, the joint probability with FCRPLM estimates are closer to the true value with larger MSEs, except when $\rho = 0$. From Table 3, the pretest and shrinkage esti-

Table 2: The Ratios of LR Based Pretest, Shrinkage Estimator MSE to the FCRPLM Estimator MSE

ρ	Pretest Estimator($\hat{\beta}$)			Shrinkage Estimator ($\hat{\beta}$)	
	LR_25%	LR_5%	LR_1%	$a = 4$	$a = 8$
0.0	0.80	0.55	0.47	0.66	0.45
0.2	0.70	0.51	0.45	0.60	0.42
0.4	0.74	0.48	0.43	0.58	0.44
0.6	0.80	0.52	0.41	0.58	0.41
0.8	0.91	0.73	0.52	0.60	0.43
ρ	Pretest Estimator($\hat{\sigma}_{kk}$)			Shrinkage Estimator ($\hat{\sigma}_{kk}$)	
	LR_25%	LR_5%	LR_1%	$a = 4$	$a = 8$
0.0	0.76	0.43	0.26	0.52	0.25
0.2	0.60	0.34	0.25	0.45	0.23
0.4	0.77	0.45	0.35	0.52	0.30
0.6	0.81	0.49	0.26	0.52	0.24
0.8	0.81	0.63	0.40	0.59	0.37
ρ	Pretest Estimator($\hat{\sigma}_{jk}$)			Shrinkage Estimator ($\hat{\sigma}_{jk}$)	
	LR_25%	LR_5%	LR_1%	$a = 4$	$a = 8$
0.0	0.71	0.32	0.17	0.36	0.10
0.2	0.56	0.19	0.08	0.27	0.06
0.4	0.75	0.39	0.30	0.39	0.15
0.6	0.83	0.49	0.23	0.44	0.18
0.8	0.87	0.68	0.37	0.45	0.22

mates reduce the MSE of the joint probability estimator compared to the FCRPLM estimates. Even though the bias of the joint probability with pretest and shrinkage estimates are higher than UCRPLM and FCRPLM estimates, the difference is small in magnitude.

To analyze the sensitivity of the RPLM in response to a change in the level of alternative attribute, we calculate the mean estimates of direct elasticity with the true parameters $(\bar{\beta}, \Sigma_{\bar{\beta}})$, Table 4, and the Monte Carlo mean estimates of direct elasticity based on pretest, shrinkage estimates and FCRPLM estimates, Table 5. The first explanatory variable in each alternative $x_{i,j,1}$ is chosen to calculate them. Since the pretest estimator with smaller level of test significance has smaller MSE, we use the pretest estimator with $\alpha = 0.01$.

Comparing the results in Table 4 to Table 5, we find that the results with FCRPLM estimates are all higher than the true values. When $\rho > 0.2$, the results

Table 3: The Share of Population Putting Negative Value on the First Two Attributes of Each Alternative, $P(\beta_1 < 0, \beta_2 < 0)$

ρ	True Prob.	UCRPLM	FCRPLM	Pretest	Shrinkage	
					$a = 4$	$a = 8$
0.0	0.025	0.047	0.120	0.014	0.027	0.015
		[0.003]	[0.049]	[0.000]	[0.008]	[0.001]
		{0.022}	{0.095}	{-0.011}	{0.002}	{-0.010}
0.2	0.100	0.060	0.110	0.021	0.037	0.024
		[0.006]	[0.077]	[0.007]	[0.016]	[0.009]
		{-0.040}	{0.010}	{-0.079}	{-0.063}	{-0.076}
0.4	0.213	0.071	0.137	0.034	0.062	0.038
		[0.026]	[0.094]	[0.033]	[0.042]	[0.034]
		{-0.142}	{-0.076}	{-0.179}	{-0.151}	{-0.175}
0.6	0.334	0.084	0.210	0.052	0.123	0.066
		[0.069]	[0.133]	[0.082]	[0.087]	[0.080]
		{-0.250}	{-0.124}	{-0.282}	{-0.211}	{-0.268}
0.8	0.406	0.115	0.292	0.090	0.259	0.171
		[0.094]	[0.153]	[0.117]	[0.113]	[0.104]
		{-0.291}	{-0.114}	{-0.316}	{-0.147}	{-0.235}

Note: [] provides the MSE results, { } provides bias results.

with pretest and shrinkage estimators are closer to the true value than those based on the FCRPLM estimators. The shrinkage estimators with $a = 8$ have smaller bias of the Monte Carlo mean direct elasticity estimates than the pretest estimates and shrinkage estimates with $a = 4$. At the same time, the shrinkage and pretest estimators have smaller standard error of the Monte Carlo mean direct elasticity estimates than the FCRPLM estimates. Based on our Monte Carlo experiment results, the shrinkage and pretest estimates will give more reliable mean direct elasticity estimates than the FCRPLM estimates, especially with a larger shrinkage constant.

Table 4: The Mean Estimates of Direct Elasticity with True Parameters ($\bar{\beta}, \Sigma_{\bar{\beta}}$)

ρ	(P_1, x_{11})	(P_2, x_{21})	(P_3, x_{31})	(P_4, x_{41})
0.0	2.009	1.960	2.053	2.042
0.2	2.014	1.957	2.052	2.049
0.4	2.020	1.954	2.051	2.057
0.6	2.025	1.951	2.051	2.065
0.8	2.031	1.947	2.051	2.075

Table 5: The Monte Carlo Mean Estimates of Direct Elasticity Based on Three Estimates

ρ	FCRPLM Estimator				Pretest Estimator (with $\alpha = 0.01$)			
	(P_1, x_{11})	(P_2, x_{21})	(P_3, x_{31})	(P_4, x_{41})	(P_1, x_{11})	(P_2, x_{21})	(P_3, x_{31})	(P_4, x_{41})
0.0	2.058 (0.026)	1.995 (0.026)	2.108 (0.028)	2.108 (0.028)	1.779 (0.019)	1.720 (0.019)	1.805 (0.020)	1.799 (0.020)
0.2	2.169 (0.027)	2.097 (0.026)	2.216 (0.028)	2.219 (0.028)	1.842 (0.019)	1.785 (0.019)	1.872 (0.020)	1.864 (0.021)
0.4	2.297 (0.031)	2.219 (0.030)	2.347 (0.032)	2.356 (0.033)	1.885 (0.021)	1.828 (0.021)	1.915 (0.022)	1.907 (0.022)
0.6	2.408 (0.031)	2.324 (0.030)	2.463 (0.032)	2.486 (0.033)	1.873 (0.021)	1.819 (0.021)	1.904 (0.022)	1.896 (0.022)
0.8	2.568 (0.031)	2.475 (0.030)	2.629 (0.032)	2.673 (0.033)	1.879 (0.026)	1.828 (0.025)	1.914 (0.027)	1.911 (0.028)
ρ	Shrinkage Estimator (with $a = 4$)				Shrinkage Estimator (with $a = 8$)			
	(P_1, x_{11})	(P_2, x_{21})	(P_3, x_{31})	(P_4, x_{41})	(P_1, x_{11})	(P_2, x_{21})	(P_3, x_{31})	(P_4, x_{41})
0.0	1.882 (0.022)	1.823 (0.021)	1.918 (0.023)	1.915 (0.023)	1.801 (0.019)	1.742 (0.019)	1.829 (0.021)	1.824 (0.021)
0.2	1.956 (0.021)	1.894 (0.021)	1.992 (0.023)	1.989 (0.023)	1.866 (0.020)	1.807 (0.019)	1.896 (0.021)	1.890 (0.021)
0.4	2.032 (0.025)	1.969 (0.024)	2.070 (0.026)	2.068 (0.026)	1.914 (0.021)	1.856 (0.021)	1.946 (0.022)	1.939 (0.022)
0.6	2.082 (0.025)	2.018 (0.025)	2.122 (0.027)	2.127 (0.027)	1.922 (0.021)	1.866 (0.021)	1.956 (0.022)	1.950 (0.022)
0.8	2.206 (0.027)	2.137 (0.026)	2.254 (0.028)	2.273 (0.029)	1.961 (0.024)	1.906 (0.023)	1.999 (0.025)	2.001 (0.025)

Note: () provides the standard error results.

References

Bhat, C. R., (2001), Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B*, 35(7), 677-693.

Greene, William H. (2012), *Econometric Analysis*, Pearson Education, Inc.

Kim, M. and Hill, R. C., (1995), Shrinkage Estimation in Nonlinear Regression: the Box-cox Transformation. *Journal of Econometrics*, 66, 1-33.

Train, K. E., (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press.