Markov Latent Class Analysis and Its Application to the Current Population Survey in Estimating the Response Error[1]

Bac Tran and Franklin Winters
U.S. Census Bureau, 4700 Silver Hill Road, Washington, D.C. 20233-1912

Key Words: Index of inconsistency, gross difference rate, misclassification error, response error, reinterview survey

Abstract

The Census Bureau uses reinterview programs in Current Population Survey (CPS) to identify problematic questions. The index of inconsistency and the gross difference rate are statistical tools which are used to measure response error.

Latent Class Analysis (LCA) offers a new way of estimating response error in longitudinal surveys. LCA does not require the reinterview process. Therefore, it saves cost, time, and reduces the burden on the respondents. Furthermore, it can estimate the error probabilities, as well as response bias.

This paper documents a simulation study which was conducted to investigate the properties of the Markov Latent Class Analysis (MLCA) technique. We applied MLCA to the CPS data and found that the results from MLCA agreed with the current reinterview method up to 84.6 percent for the Employed category, 84.4 percent for the Unemployed category, and 91.9 percent for the Not In Labor Force category in estimating the response error when the reinterview replicated the original interview.

I. Background

The Current Population Survey (CPS), sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), is a household sample survey conducted monthly by the U.S. Census Bureau. It is the nation's primary source of labor force statistics for the entire population. The CPS sample is a multistage stratified sample of approximately 60,000 housing units. Each month the CPS interview starts from the week of the 19th, and is completed within two weeks. The CPS focuses on the labor force status (Employed, Unemployed, and Not in Labor Force) of the working-age population and demographic characteristics of workers and non workers.

In order to evaluate the quality of the data, the Census Bureau uses a reinterview-based, test-retest approach. The CPS reinterview program has been in place since 1954, and it serves two main purposes: Quality Control (QC) to monitor the work of the field representatives (FRs) and response error (RE) to evaluate the data quality. RE is one type of non-sampling error. In this paper, we focus on RE reinterview. Here afterward when we describe reinterview, it refers to the Response Error reinterview. The reinterview sample is a systematic random sample from the eligible sample units interviewed each month. Each month, reinterview is conducted on one percent of the original interview households.

Response Model

The model used in this section was described in detail in Flanagan's dissertation. A brief description is given below.

Response= True + Error

$x_i = \mu_i + \epsilon_i$

where x_i is the response from i^{th} sample unit, μ_i is the 'true' answer and ϵ_i is the response error. The census total is:

$$X = \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \mu_i + \sum_{i=1}^{N} \epsilon_i,$$ where N denotes the size of the population.

$$MSE(X) = E\left(X - \sum_{i=1}^{N} \mu_i\right)^2 = [Bias(X)]^2 + Var(X)$$

where

$$Var(X) = Var\left(\sum_{i=1}^{N} \epsilon_i\right) = \sum_{i=1}^{N} Var(\epsilon_i) + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}}^{N} Cov(\epsilon_i,\epsilon_j)$$

The first term divided by N, $\frac{1}{N}\sum_{i=1}^{N} Var(\epsilon_i)$, is called the ***simple response variance*** (SRV). It is the average variance of response error over the population. The second term $\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i\neq j}}^{N} Cov(\epsilon_i,\epsilon_j)$ can represent any situation that would cause correlation from person to person in the response errors.

Reinterview Model

Estimates of SRV are useful to see how extensive the response error is for a particular question. We design the reinterview with the purpose of estimating the SRV. With reinterview we are able to measure the data quality via the index of inconsistency for each question category. The index of inconsistency is defined as the ratio of SRV to the Var(X).

[1]This paper (chapter) reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in progress). The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Using the model described above we have:
$x_{1i} = \mu_i + \epsilon_{1i}$ (for the original interview)
$x_{2i} = \mu_i + \epsilon_{2i}$ (for the reinterview)
Assuming the two error terms are independent, identically distributed (iid) random variables, then the following is used as an unbiased estimate of the SRV (Flanagan 2001):

$$\hat{\sigma}_{\epsilon}^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_{1i}-x_{2i})^2}{2}$$

where n is the sample size of the reinterview sample.

In reinterview we assume independence between the original interview and the reinterview. We also assume that the reinterview replicates the original interview. Under these two assumptions, we estimate the index of inconsistency, a measure of response error. Table 1 is a 2x2 table for a dichotomous response.

Table 1

Reinterview Response	Original Response		
	Subtotal	Yes	No
Subtotal	n	a + c	b + d
Yes	a + b	a	b
No	c + d	c	d

For a 2x2 table, the index is defined as:

$$\text{Index} = \frac{(b+c)/n}{\frac{1}{2}(p_1\, q_2 + p_2\, q_1)}$$

where $p_1 = (a+c)/n$, $p_2 = (a+b)/n$, $q_1 = 1-p_1$, $q_2 = 1-p_2$.

The numerator of the index is the Gross Difference Rate (GDR). One half of GDR estimates the Simple Response Variance. The Net Difference Rate (NDR) = $p_2 - p_1$ is used to test the replication assumption. The significance test statistic for the NDR is:

$X^2 = \frac{(|b-c| - 1)^2}{b+c}$. X^2 has the χ^2 distribution with 1 degree of freedom (Fleiss 1981). A statistical difference from 0 of NDR, equivalently $X^2 > 3.841$, indicates that the reinterview did not replicate the original interview.

We use the following rule of thumb to interpret the index of inconsistency.

Index	Value	Interpretation
I< 20	Low	Usually not a major problem
20<= I < 50	Medium	Somewhat problematic
50 <= I	High	Very problematic

We conducted research using Markov Latent Class Analysis (MLCA), and found that MLCA could be an alternate way to measure the quality of data (estimating the index of inconsistency and response error probabilities) without conducting the RE reinterview.

The concept behind latent class analysis for the CPS is that the true labor force status cannot be observed directly but we can nevertheless represent it with latent variable(s). The observed (manifest) data are just the indicators of the unobserved (latent) variables. The relationship between manifest variables and latent variables is made explicit through mathematical models. In this paper we describe the measurement model in which the relationship between observed variables and latent variables will be represented by the response probabilities. If a subject moves from one state to another state (a chain) over the time, then the process will be defined by the transition probabilities. A Markov chain is a chain in which a state of a subject at time t depends on the state at time t-1and does not depend on the states at time t - 2 or earlier. A Markov chain is said to be stationary if the transition probabilities do not change over time. Response (or transition) probabilities are said to be homogeneous under some subpopulations if they are the same under those subpopulations. Markov Latent Class Analysis (MLCA) Model is a model which has Markov chain. In this paper we present the labor force status by a Markov chain of three consecutive months.

Biemer and Bushery (2000) applied MLCA to CPS data for the first three months in each year of 1993, 1995, 1996. They found that the model fit to the data had the following properties:
transition probabilities were nonhomogeneous and non-stationary, response probabilities were nonhomogeneous and stationary (see section II). We utilize their findings for our further investigation. First, we validated the theory by a simulation. Second, we applied the theory to the CPS data from January 1996 to December 1999, found the best model to the data, then compared the indices computed from the model parameters to the reinterview's indices. At last, we conducted a second simulation to see how well MLCA performed if the Markov assumption was violated.

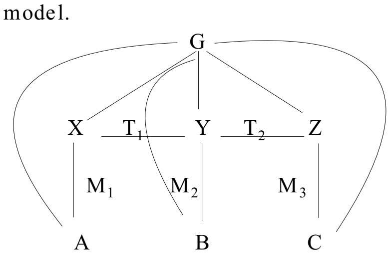
II. Methodology

We ran a simulation study to verify the validity of the MLCA method. We found that the method was reliable in estimating the parameters of the CPS simulated data when the Markov assumption was met. We applied the method to the CPS data. Then we compared response error estimated from the MLCA parameters with those of the RE reinterview when the RE reinterview replication was satisfied. In this study we used CPS data from 1996 to 1999.

A. MLCA Model for CPS

MLCA uses mathematical models for characterizing the latent variables. It is commonly used in the analysis of attitudinal measures from survey research (McCutcheon 1987). In our CPS model we used data from three time periods (three consecutive months). The observed variable was labor force status: Employed (E), Unemployed (U), Not in Labor Force (NLF). The grouping variable used was the proxy variable. Figure 1 illustrates the graphical model.

Figure 1



where T_1, T_2 are transition matrices, M_1, M_2, M_3 are response matrices, G is a grouping variable. In our model G has two levels: self and proxy. A response was considered as a 'self-response' if the respondent was the same for three consecutive months.

X, Y, and Z are latent variables which, under each level g of G, were defined as:

$$X = \begin{cases} 1 \text{ if person was employed in time period 1} \\ 2 \text{ if person was unemployed in time period 1} \\ 3 \text{ if person was NLF in time period 1} \end{cases}$$

with analogous definitions for Y and Z for periods 2, and 3 respectively.

A, B, and C are observed (reported) variables which, under each level g of G, were defined as:

$$A = \begin{cases} 1 \text{ if person was employed in time period 1} \\ 2 \text{ if person was unemployed in time period 1} \\ 3 \text{ if person was NLF in time period 1} \end{cases}$$

with analogous definitions for B and C for periods 2, and 3 respectively.

With the above illustration, we have:

$$\pi_{x,y,z\,|\,g} = \pi_{x\,|\,g}\,\pi_{y\,|\,g,x}\,\pi_{z\,|\,g,x,y} \qquad (1)$$

The left hand side is P(X=x,Y=y,Z=z|G=g), and the right hand side is:

$$\pi_{x\,|\,g} = \Pr(X=x|G=g)$$
$$\pi_{y\,|\,g,x} = \Pr(Y=y|\ X=x,\ G=g)$$
$$\pi_{z\,|\,g,x,y} = \Pr(Z=z|\ X=x,\ Y=y,\ G=g)$$

Therefore,

$$\pi_{g,a,b,c} = \pi_g \cdot \pi_{abc|g} \qquad (2)$$

$$= \sum_{x,y,z} \pi_g\ \pi_{abc|gxyz}\ \pi_{xyz|g} \qquad (3)$$

$$= \sum_{x,y,z} \pi_g\ [\pi_{a|gx}\ \pi_{b|gy}\ \pi_{c|gz}]\ [\pi_{x|g}\ \pi_{y|gx}\ \pi_{z|gxy}] \quad (4)$$

$$= \sum_{x,y,z} \pi_g\ [\pi_{a|gx}\ \pi_{b|gy}\ \pi_{c|gz}]\ [\pi_{x|g}\ \pi_{y|gx}\ \pi_{z|gy}] \quad (5)$$

$$= \sum_{x,y,z} \pi_g\ (\pi_{a|gx}\pi_{x|g})\ (\pi_{b|gy}\ \pi_{y|gx})\ (\pi_{c|gz}\ \pi_{z|gy}) \quad (6)$$

$\pi_{a|gx}$, $\pi_{b|gy}$, $\pi_{c|gz}$ are called misclassification probabilities.

$\pi_{y|gx}$, $\pi_{z|gy}$ are called transition probabilities. From (3) to (4), local independence assumption is assumed. From (4) to (5), the Markov assumption is assumed. These are the two assumptions for our CPS MLCA model.

Equation (6) provides the expected sample frequency in cell (g, a, b, c) through the parameters on the right hand sides. Under the two assumptions, in this paper the E-M algorithm was used to estimate the model parameters. The iterations from the algorithm will converge to the solution when the log likelihood function of GABC table does not exceed some convergence criterion (log likelihood for the model, Van de Pol. 1986). The likelihood function of the GABC table is

$$\Pr(G=g,\ A=a,\ B=b,\ C=c) = k \prod_{g,a,b,c} \pi_{g,a,b,c}^{f_{gabc}}$$

where k is the multinomial constant, and f_{gabc} is the observed frequency in cell (g, a, b, c).

B. Theory Validation

We simulated fifty CPS data sets of sample size 50,000 each. Figure 1 could be used as a visual tool to understand the creation of simulated data. In the simulation process, X and the set of probabilities (transition and misclassification) were created first. They were used to simulate the CPS data. In addition, a variability (the diagonal of the transition and misclassification probabilities) of 5 percent was built into the simulation. Also, the data were created such that transition and misclassification probabilities were stationary but non-homogeneous. We developed some software, which treated the conditions on the probabilities as specified, to estimate the parameters based on the observed frequencies. The parameters estimated by our program were compared with the true values: X, transition and misclassification probabilities. The results showed 100 percent agreement. Therefore, we concluded that the theory works for CPS simulated data.

III. Application to the CPS

We used CPS data from 1996 to 1999. In each year, we created three-time-period data by combining data from three consecutive months. We compared all possible models which fit the data. We found that the model for CPS had the conditions: non-stationary, non-homogeneous transition probabilities, stationary but non-homogeneous response probabilities. This is consistent with the findings from Biemer and Bushery (2000). We used PANMARK software to estimate the CPS model parameters.

A. Estimation of Index from MLCA Parameters

To estimate the index of inconsistency from the model we need to construct the interview-reinterview table from the misclassification probabilities. This is equivalent to estimating $\pi^{A\ A'}$ where A, and A' denote the labor force classification for the original and reinterview, respectively. Under each level of X, x= 1, 2, 3, we assume that

$$\pi^{A\ A'|X=x} = \Pr(A=i,\ A'=j|\ X=x) = \Pr(A=i|X=x)*\Pr(A'=j|X=x)$$
$$= \Pr(A=i|X=x)*\Pr(A=j|X=x)$$

The second equality means that the errors are the same between the original and reinterview.

Therefore,

$$\pi^{A\ A'} = \sum_{x=1}^{3} Pr(X=x) * \pi^{A\ A'|X=x}$$

B. Results

We compared the indices from the reinterview with those estimated from the misclassification probabilities of the model. We found that when reinterview replicated the original interview (NDR was not significantly different from 0), the indices obtained from the model parameters agreed with those of reinterview up to 84.5 percent for the Employed category, 84.4 percent for the Unemployed category, and 91.9 percent for the NLF category. Table 4 in the Appendix shows the comparison of the reinterview indices to those computed from the MLCA model parameters for 1996.

We compare our estimates of the CPS classification probabilities with similar estimates from the previous paper (Biemer and Bushery). The results are summarized in Table 2.

Table 2

Classification		Biemer & Bushery (The first three month of each year)	Our Estimates (From Jan 1996 to Dec 1999)
True	**Reported**		
Emp	**Emp**	98.77 **(1993)** 98.73 **(1995)** 98.73 **(1996)**	98.74
	Unemp	0.34 **(1993)** 0.49 **(1995)** 0.37 **(1996)**	0.37
	NLF	0.89 **(1993)** 0.78 **(1995)** 0.79 **(1996)**	0.89
Unemp	**Emp**	7.06 **(1993)** 7.86 **(1995)** 8.57 **(1996)**	9.87
	Unemp	81.81 **(1993)** 76.09 **(1995)** 74.42 **(1996)**	71.38
	NLF	11.13 **(1993)** 16.04 **(1995)** 17.00 **(1996)**	18.75
NLF	**Emp**	1.41 **(1993)** 1.11 **(1995)** 1.13 **(1996)**	1.26
	Unemp	0.75 **(1993)** 0.69 **(1995)** 0.87 **(1996)**	0.72
	NLF	97.84 **(1993)** 98.20 **(1995)** 98.00 **(1996)**	98.03

C. Goodness of Fit

Out of forty CPS data sets from January 1996 to December 1999, twenty three of them (with 12 degrees of freedom each) had small χ^2. However, if using dissimilarity index, a model fit criterion suggested by Vermunt (1997), then all of them had the indices smaller than 0.0024. Models having dissimilarity indices smaller than 0.05 (i.e., 5 percent model error) are considered to fit the data well (Vermunt 1997).

IV. Violation of the Markov Assumption

From forty CPS data sets (1996-1999), we created forty simulated CPS data sets. We started with a stationary Markov chain $T_1= T_2=(t_{ij})$ where $t_{ij}= Pr(Y=i|X=j)= Pr(Z=i|Y=j)$. Then we created T_2 by modifying t_{ij} by λt_{ij} under certain values of X=x. In other words, the transition matrix (T_2) is modified to be T'_2 as a function of X, $t'_{ij}= \lambda t_{ij} =Pr(Z=i|Y=j, X=x)$. The more λ is away from 1, the more the model is away from the Markov assumption (λ=1). In our study we specified λ= 0.2, 0.4, 0.6, 0.8. In addition, we focused on the Unemployed category (X=2), which had the smallest counts and had the inconsistency over time. The results from the simulation showed that the difference between Pr('Employed'|True 'Employed') as well as Pr('Not in Labor Force'|True 'Not In Labor Force') of the Markov model and those of the violated ones are very small, about 0.003 and 0.002 respectively. However, for the Pr('Unemployed'|True 'Unemployed'), the difference varies depending on how far the λ is away from 1. This is because the relatively small counts of the Unemployed category to the sample size (about 3 percent) becomes more and more dispersed over the other categories when λ is more and more away from 1.

The figures in the second row of Table 3 are the average difference between Pr('Unemployed'| true 'Unemployed') corresponding to each value of λ to those when Markov assumption met (λ=1), i.e. the average of

d= $|Pr_{\lambda}$('Unemp' | true 'Unemp') - $Pr_{\lambda=1}$('Unemp'| true 'Unemp')|

Table 3: **Comparison of d Under Non-Markov Transitions**

λ	λ= 0.2	λ= 0.4	λ= 0.6	λ= 0.8	λ= 1 (Markov)
d	0.072	0.079	0.069	0.039	0

As described above, in our simulation we tried to distort t_{ij} when X= 2 (Unemployed). Table 5 gives examples of how the counts look like when λ varies. From Table 5, we can compare and see that CPS data do not fluctuate that much. In addition, when applying MLCA to CPS data we found high agreements between reinterview indices and MLCA indices; therefore, the situations where $\lambda < 0.8$ are not likely to occur in CPS data.

V. Pros and Cons

The MLCA offers many advantages which our traditional cannot. First of all, MLCA does not require reinterview data. This saves cost, time, and respondent burden. Respondent burden could cause a non-response problem for subsequent interviews. Not only can MLCA estimate the SRV, but also can estimate the bias and misclassification probabilities. However, MLCA requires multiple interviews (longitudinal data). It deals with the marginal frequencies, not with individual level data. MLCA also requires a set of assumptions. In this study, local independence and Markov assumptions are employed. A small portion of the population which become unemployed and stay unemployed for a long time could violate the Markov assumption. To further investigate this, we would consider the Mover-Stayer Markov Latent Class model (MSMLC) to investigate the heterogeneity, which we do not present in this paper. MSMLC model classify the sample population into subgroups in which the statuses of the people are pretty much unchanged in one subgroup, but changed in the other.

VI. Conclusion

Our primary goal of this research was to investigate the validity of MLCA through simulation data. The results showed that the method was robust for the simulated data, then we applied it to estimate the CPS labor force classification error. The results showed that our estimates are pretty close to those published by Biemer and Bushery (2000). Also, when comparing the indices of inconsistency estimated from the MLCA model to the traditional method, we found that they matched 84 percent when the NDR in the reinterview data was not significantly different from 0. Lastly, we did a second simulation to violate the Markov assumption to see how well MLCA performs when the model assumption is violated. The results showed that the MLCA method was in the 3.9 percent difference range (Table3) when Markov assumption was violated at λ= 0.8 under the Unemployed category. We also found that the MLCA estimates of correct misclassification probabilities for group 1 (self) were lower than those of group 2 (proxy). This is consistent with our reinterview research on the consistency of the responses across the two groups, self and proxy. In addition, the misclassification probabilities were high on the 'Unemployed' group. This also agreed with our previous studies (Bushery and McGovern). When $\lambda < 0.8$ the average difference between the two misclassification probabilities (P('Unemployed'| True 'Unemployed')), Markov and Non-Markov, was 0.08 at the most when Markov assumption is violated badly. However, based on the CPS data over time, this is unlikely to occur. We believe that the labor force questions in our questionnaire follow the Markov chain. This implies that the respondents would not take into account their previous responses when providing their current employment status. However, there is a small portion of the population which could violate the Markov assumption (section V), then mover-stayer model could be employed to investigate the heterogeneity for this issue. In summary, we concluded that MLCA performed pretty well in estimating the error probability in CPS. It should be considered as an alternate method for evaluating the CPS data quality.

References

Agresti, A. (2002). Categorical Data Analysis. New York: Wiley

Biemer, P., Bushery, J. (2000). "On the Validity of Markov Latent Class Analysis For Estimating Classification Error in Labor Force Data," Survey Methodology, Vol. 26, No. 2, pp. 139-152.

Bushery, J., McGovern, P. (1999) Data Mining the CPS Reinterview: Digging Into Response Error. Internal U.S. Census Bureau, Washington, D.C.

Flanagan, P. (2001). "Measurement Errors in Survey Response," PH.D. Dissertation.

Fleiss, J. L. (1981). Statistical Methods for Rates and Proportions. New York: Wiley

Hagenaars, J.A. (1990). Categorical Longitudinal Data: Log-Linear Panel, trend, and Cohort Analysis. Newbury Park, CA: Sage.

Hui, S.L and S.D. Walter (1980). "Estimating the Error Rates of Diagnostic Test," Biometrics, 36, 167-171.

McCutcheon, A.L. (1987) Latent Class Analysis. Newbury Park, CA: Sage.

McCutcheon, A.L. and Hagenaars, J.A. (1977). "Simultaneous Latent Class Models for Comparative Social Research." In R. Langeheine and J. Rost (eds.), Applications of Latent Traits and Latent Class Models. New York: Waxman, pp. 266-77.

Mitchell Dayton (1998) Latent Class Scaling Analysis. Thousand Oaks, CA: Sage.

Van de Pol, F. and De Leeuw, J. (1986). "A Latent Markov Model to Correct for Measurement Error," Socialogical Methods and Research, Vol. 15, pp.118-141.

Van de Pol, F., R. Langeheine, and Wil de Jong (1999). PANMARK 3: Panel Analysis using Markov Chains-A Latent Class Analysis program (users' manual). Voorburg, The Netherlands: University of Kiel.

Vermunt, J. K. (1997). Log-Linear Models For Event Histories, Thousand Oaks, CA: Sage.

Notes: PANMARK 3 program (Van de Pol F., 1999) was used to estimate all of the models presented in this paper.

Appendix

Table 4: Comparison of MLCA and Traditional Estimates of the Index of Inconsistency in 1996

Month	Category	RI Index	RI LCL	RI UCL	MLCA Index	MLCA LCL	MLCA UCL
1	E	4.9684	2.7295	7.2074	5.8718	5.5932	6.1503
1	U	34.0909	15.5029	52.6789	36.8378	35.1675	38.5081
1	N	5.8389	3.3942	8.2837	7.7659	7.4403	8.0914
		.	.	.	.	.	.
2	E	7.6970	4.9192	10.4748	5.5558	5.2847	5.8269
2	U	34.2778	17.5876	50.9680	35.8220	34.1377	37.5062
2	N	9.4892	6.3755	12.6029	7.1093	6.7975	7.4211
		.	.	.	.	.	.
3	E	5.7721	3.3556	8.1885	5.1845	4.9221	5.4470
3	U	25.1595	13.6726	36.6464	36.9104	35.1807	38.6401
3	N	7.6592	4.8210	10.4973	6.7576	6.4530	7.0623
		.	.	.	.	.	.
4	E	6.0177	3.7809	8.2545	6.5632	6.2692	6.8573
4	U	37.7778	22.9767	52.5788	41.5334	39.6630	43.4039
4	N	9.7319	6.8787	12.5852	7.9152	7.5869	8.2434
		.	.	.	.	.	.
5	E	8.7895	6.0161	11.5630	6.7417	6.4443	7.0391
5	U	31.8905	19.1179	44.6630	42.0829	40.2198	43.9461
5	N	11.5752	8.3536	14.7969	8.9160	8.5689	9.2630
		.	.	.	.	.	.
6	E	7.1535	4.7421	9.5648	7.6966	7.3799	8.0132
6	U	31.9588	19.5869	44.3306	42.7358	40.9194	44.5521
6	N	10.0921	7.1819	13.0024	8.7573	8.4128	9.1018
		.	.	.	.	.	.
7	E	7.4629	4.9497	9.9760	7.5955	7.2789	7.9120
7	U	23.0744	12.8903	33.2586	42.3983	40.5978	44.1988
7	N	10.2966	7.2786	13.3146	8.7439	8.3969	9.0909
		.	.	.	.	.	.
8	E	7.8925	5.2935	10.4915	6.5743	6.2811	6.8675
8	U	36.1836	20.8457	51.5215	41.5367	39.6779	43.3955
8	N	11.1862	8.0689	14.3035	7.6844	7.3614	8.0073
		.	.	.	.	.	.
9	E	6.2632	4.0359	8.4906	6.3790	6.0927	6.6654
9	U	27.9967	16.2454	39.7480	37.7582	35.9582	39.5582
9	N	8.3059	5.7307	10.8810	7.3334	7.0209	7.6459
		.	.	.	.	.	.
10	E	5.5925	3.4525	7.7325	5.3031	5.0429	5.5633
10	U	54.4470	35.6741	73.2199	38.1908	36.3476	40.0340
10	N	8.9609	6.2331	11.6887	6.7594	6.4611	7.0577

Table 5: **The cross-tabulations of X*Y*Z when X= 2 (Unemployed):**

λ=1 (Markov)

		Z: E	Z: U	Z: N
	E	53.89	0.23	0.33
Y	U	42.52	437.26	80.14
	N	0.94	0.52	101.17

λ= 0.8

		Z: E	Z: U	Z: N
	E	43.11	11.01	0.33
Y	U	34.02	445.76	80.14
	N	0.75	0.71	101.17

λ= 0.6

		Z: E	Z: U	Z: N
	E	32.33	21.79	0.33
Y	U	25.51	454.27	80.14
	N	0.56	0.89	101.17

λ= 0.4

		Z: E	Z: U	Z: N
	E	21.56	32.56	0.33
Y	U	17.01	462.77	80.14
	N	0.38	1.08	101.17

λ= 0.2

		Z: E	Z: U	Z: N
	E	10.78	43.34	0.33
Y	U	8.5	471.28	80.14
	N	0.19	1.27	101.17