Empirical study to segment firms and capture dynamic business context using LCA

Chakrabarty, Subhajit and Nag, Biswajit

November 2013

The Empirical Economics Letters, 12(11): (November 2013) ISSN 1681 8997

Empirical study to segment firms and capture dynamic business context using LCA

Subhajit Chakrabarty
Associate Professor, Auro University, India (Corresponding author)
Email: subhajitchakrabarty@hotmail.com
Biswajit Nag
Associate Professor, Indian Institute of Foreign Trade, India

Abstract: *The usual methods of segmenting firms are insufficient as they do not consider hidden (unobserved) groupings and do not consider the dynamic market context such as in the apparel industry. An empirical analysis was done using latent class analysis on a cross-section survey of 334 Indian apparel exporting firms. Five latent classes were found by empirical estimation – (i) very old manufacturers in tier 1 cities with large turnover, (ii) manufacturers in tier 2 and 3 cities, (iii) small merchants from the quota-system period dealing in some high fashion, (iv) new firms dealing in some high fashion and women's garments, (v) new firms not in high fashion. These latent classes are found valid in market context and hence this method can be further explored. An incentive policy structure for the target latent groups in the industry can be better designed from the results.*

Keywords: Segmentation, classification, clusters, policy, garments

JEL: F10, F12, F14

1. INTRODUCTION

Segmentation in a market is usually done from perspective of consumer demand. These methods may also be useful to segment firms from perspective of policy (such as Government policy). Convenient methods to group the firms could be product-wise, geographical (location-based), size-wise (capacities), market-wise (exporting to European market etc), customer-wise (business-to-business scenario etc), age-wise (year of starting operations) and so on. What about

unobserved combinations of old firms (vs. new firms), niche / boutique firms (vs. traditional firms), merchants (vs. manufacturers), large firms (vs. small firms)? Our motivation is to find whether we can find latent (unobserved) groupings among these firms and whether this can capture the dynamic business context from a policy perspective.

1.1 Context of the empirical study - Indian apparel exporting firms

The apparel manufacturers and merchants are commonly classified on the basis of products because of the traditional differentiating factors such as design, fabric and process for these products. Policymakers segment them in terms of export turnover (such as star category) and in terms of manufacturer or merchant (manufacturers required many licenses). But a policy perspective will require more or different variables to consider (as against the perspective of consumer demand) such as turnover, location in rural or urban places and so on. The current methods of segmenting are, therefore, insufficient as they do not consider latent (hidden) groupings.

In the context of apparel trade, there has been a period prior to 2006 when exports to US (the biggest importing country for Indian apparels) were determined by quotas; this period is often called the Multi-Fibre Agreement period (MFA period or 'quota' period). During this period, there emerged in India, merchant firms who would trade in these quotas or export in their own name. Such merchant firms need also to be identified.

1.2 Background of LCA

The classical theory of latent class analysis was first proposed by Paul Lazarsfeld in 1950. The theoretical framework was first laid out by T.W.Anderson in 1954. L.A.Goodman in 1974 found an iterative method to solve the latent class model through maximum likelihood equations. This method, in its general form, was introduced by Dempster, Laird and Rubin in 1977, now called the EM-algorithm. (Laird 1978) has provided a useful discussion of literature on various Bayes methods for analysing contingency tables. He used the EM-algorithm for estimation of the parameters. (Andersen 1982)

The latent class model seeks to stratify the cross-classification table of observed ("manifest") variables by an unobserved ("latent") unordered categorical variable. Conditional upon values of this latent variable, responses to all of the manifest variables are assumed to be statistically independent; an assumption referred to

as “conditional" or “local" independence. The model probabilistically groups each observation into a “latent class," which in turn produces expectations about how that observation will respond on each manifest variable. The latent class model is actually a type of finite mixture model, as the unobserved latent variable is nominal (membership of a class) (Agresti, 2002, p. Ch.13).

The latent class model can suffer from the local dependence problem in which the latent class nodes are based on local branches and not the entire tree. Improvements in algorithms were made by considering hierarchical latent class (HLC) model or latent structure (LS) model. A better approach is a general graph structure for the manifest (observed) variables so as to tackle the local dependence problem through searching for the dominant nodes and pruning others (Chen, Hua, & Liu, Generalized Latent Class Analysis based on model dominance theory, 2009).

1.3 Objectives

The first objective of our study was to conduct an empirical exercise for identifying latent classes among firms in the dynamic business context from the policy perspective – using Indian apparel exporting firms as a case. The second objective was to interpret the latent groupings in the empirical exercise.

2. THEORETICAL FRAMEWORK OF LCA AND APPLICABILITY

2.1 Theoretical model

Let us take a 4-dimensional contingency table with I x J x K x L random variables denoted as

$\{ X_{ijkl} \}$, i = 1 to I, j = 1 to J, k = 1 to K, l = 1 to L.

The cell probabilities are denoted by π_{ijkl} . The parameters of π_{ijkl} depend on an unobservable latent variable denoted as θ. This latent variable may be a single-value or a vector.

For marginal variables of the contingency table, let A, B, C, D correspond to i,j,k,l indices respectively. Therefore the marginal probability of falling in category i, j, k, l is respectively given by

$\pi_i^A(\theta)$, $\pi_j^B(\theta)$, $\pi_k^C(\theta)$, $\pi_l^D(\theta)$.

Assumption of local independence implies that:-

Equation 1: Product of marginal probabilities

$$\prod_{ijkl}(\theta) = \prod_{i}^{A}(\theta).\prod_{j}^{B}(\theta).\prod_{k}^{C}(\theta).\prod_{l}^{D}(\theta).$$

Typically latent structure models assume continuous distribution while latent class models assume point distribution. Let $\phi(\theta)$ be an m-point distribution in which m distinct points (θ_1, … θ_m) take all the probability mass.
For the point distribution with class category v, let $\pi_{iv}^{A}, \pi_{jv}^{B}, \pi_{kv}^{C}, \pi_{lv}^{D}$ denote $\pi_i^A(\theta_v)$, $\pi_j^B(\theta_v)$, $\pi_k^C(\theta_v)$, $\pi_l^D(\theta_v)$ respectively (v = 1 to m, for m points in the distribution).
The latent class model is:-

Equation 2: Basic latent class model

$$\prod_{ijkl} = \sum_{v=1}^{m} \prod_{iv}^{A} \prod_{jv}^{B} \prod_{kv}^{C} \prod_{lv}^{D} \phi_v$$ (Andersen, 1982).

In the I x J x K x L x m dimensional contingency table X_{ijklv}, the cell probability is given by:-

Equation 3: Cell probability

$$\mathrm{E}[X_{ijklv}] = n.\prod_{iv}^{A} \prod_{jv}^{B} \prod_{kv}^{C} \prod_{lv}^{D} \phi_v$$

As per the EM algorithm, there are two steps – one for expectation and the other for maximization.
In the E-step, we estimate variable $p_{iv}^{A(n)}$ such that

Equation 4: E-step

$$p_{iv}^{A(n)} = \sum_{j} \sum_{k} \sum_{l} X_{ijkl} \prod_{ijklv}^{*}$$

In which

$$\prod_{ijklv}^{*} = \prod_{iv}^{A} \prod_{jv}^{B} \prod_{kv}^{C} \prod_{lv}^{D} \phi_v \Big/ \sum_{v=1}^{m} \prod_{iv}^{A} \prod_{jv}^{B} \prod_{kv}^{C} \prod_{lv}^{D} \phi_v$$

In the M-step,

Equation 5: M-step

$$\prod_{iv}^{A} \phi_v = p_{iv}^{A(n)}$$

We can now re-compute the values through repeating the E and M steps using the previous values. Then, we obtain stable values of the parameters π_{iv}^{A} , π_{jv}^{B} , π_{kv}^{C} , π_{lv}^{D} , ϕ_v of the model through repeated E-M steps (Andersen, 1982).

2.2 Applicability of LCA

Market segmentation is the most common application of latent class analysis (Lockshin and Cohen 2011) (Green, Carmone and Wachspress 1976) and has uses across many branches / sectors. For example, applicability to fashion product consumers has been shown (Kim and Lee 2011). Intra-industry heterogeneity has been analysed using LCA (DeSarbo, Wang and Blanchard 2010). Scale development and testing is an important area where LCA has been found useful (Kreuter, Yan and Tourangeau 2008). Trade has been analysed using LCA / LCR (Audretsch, Sanders and Zhang 2011). A study on segmenting electrical distribution firms for Government policy (Cullmann 2012) is a work using LCA which is close to our paper.

3. METHODOLOGY

3.1 Instrument and method

A questionnaire was used to seek information from about 7500 exporters about the products they were dealing currently. These exporters are registered with an export promotion body sponsored by the Government of India. 334 responded properly. The options for product-type were Men's wear, Women's wear, Kid's wear, Made-ups, Industrial wear, High Fashion, Accessories, Any other. The other details obtained which were confirmed with the data already available in database are Year of commencing operations, Type of exporter – Manufacturer or Merchant, Level of operations – Export turnover greater than Rs 5 Crore or not, Primary location (city or town) and Live exporter or ceased.

Accordingly, the observed (manifest) variables taken in the model are given below:-

Profile-related manifest variables:-

- TYPE (manufacturer or merchant)
- MEM (turnover of over Rs 5 crore or less)
- CITY (tier classification of city)
- YEARS (years since started business)

Product-related manifest variables:-

- GENTS (men's wear)
- WOMEN (women's wear)
- KIDS (kid's wear)
- HIGHFASHION (high fashion/ boutique wear)
- MADEUPS (made-ups)
- ACCESSORIES (accessories to garments)
- MIXED (other categories not listed).

Latent Class Analysis was done through the poLCA package in R (Linzer & Lewis, 2011). AIC values were checked for determining the number of classes (Nylund, Asparouhov, & Muthén, 2007).

4. RESULTS

4.1 Number of latent classes

The results indicated presence of five latent classes as indicated from the AIC values given below:-

Table 1: Number of classes

No. of Classes	AIC	BIC	CHI-SQ
2	3514.277	3617.178	4660.397
3	3482.265	3638.522	3328.122
4	3454.692	3664.305	3276.108
5	3421.416	3684.385	3861.723
6	3423.913	3740.238	2958.944

Based on the lower AIC value, we take the number of classes as five. This is also validated from the business context of the groupings.

4.2 Profile of the latent classes

The profiles of the latent classes are detailed in Table 2.

Table 2: Profile of latent classes

CLASS 1	CLASS 2	CLASS 3	CLASS 4	CLASS 5
Tier 1 cities	Tier 1 & 2 cities	Mostly Tier 1 cities	Tier 3 and Tier 2 cities	Tier 1 and 2 cities
Very old firms (>25 yrs)	Mostly new firms (post-MFA)	Mostly MFA-period firms		Relatively new firms (post-MFA)
Large turnover (>Rs 5 Cr)		Most small turnover		
Manufacturers		Mostly Merchants	Manufacturers	
Women's also to a large extent	Women's			
Kid's also to a large extent				Kid's
Men's also to a large extent			Mostly men's	
No High Fashion	Little-bit High Fashion	Some High Fashion	No High Fashion	No High Fashion
No Made-ups				
No Accessories		Mostly Accessories		

Based on the profiles, the following latent classes emerged:-
(i) very old manufacturers in tier 1 cities with large turnover (Class 1),
(ii) manufacturers in tier 2 and 3 cities (Class 4),
(iii) small merchants from the MFA-period dealing in some high fashion (Class 3),
(iv) new firms dealing in some high fashion and women's garments (Class 2) and
(v) new firms not in high fashion (Class 5).

4.3 Variables and the probabilities

The probabilities of the choices of the variables from the Latent Class Analysis are given in Table 3 and Table 4. Table 3 provides the results related to profile-related variables while Table 4 provides the results related to product-related variables.

Table 3: Probabilities of the choices of profile-related manifest variables

Variable TYPE

	Manufacturer	Merchant
Class 1	1.0000	0.0000
Class 2	0.6553	0.3447
Class 3	0.2286	0.7714
Class 4	1.0000	0.0000
Class 5	0.5941	0.4059

Variable MEM

	Turnover >= Rs 5 Crore	Turnover < Rs 5 Crore
Class 1	1.0000	0.0000
Class 2	0.2139	0.7861
Class 3	0.0408	0.9592
Class 4	0.3197	0.6803
Class 5	0.2238	0.7762

Variable CITY (Tier as per government city classification)

	Tier 1	Tier 2	Tier 3
Class 1	0.9069	0.0931	0.0000
Class 2	0.7969	0.2031	0.0000
Class 3	0.9132	0.0470	0.0398
Class 4	0.4204	0.1137	0.4659
Class 5	0.8608	0.1392	0.0000

Variable YEARS (Years since start of operation)

	< 8 years	Between 8 and 25 years	>25 years
Class 1	0.0000	0.3116	0.6884
Class 2	0.2594	0.5818	0.1588
Class 3	0.1260	0.8319	0.0422
Class 4	0.0554	0.7806	0.1640
Class 5	0.1721	0.7315	0.0963

We infer from the results in respect of variable TYPE that Class 1 and Class 4 represent manufacturers, of which Class 1 are also found to be the old and large manufacturers. Observing the results for variable MEM, we find that Class 1 represent exporters with high turnover while Class 3 represents those with low turnover. From the results for variable CITY, we find that Class 1 represent those located in tier 1 cities; they are the old manufacturers and the MFA-period merchant firms mostly. From the probabilities of the variable YEARS, we infer that Class 1 represents the old firms, Class 3 largely represents the MFA-period firms, while Class 2 represents the new firms, by and large.

Probabilities of choices of product-related variables are given in Table 4.

Table 4: Probabilities of the choices of product-related manifest variables

	GENTS (Men's wear)		WOMEN (Women's wear)		KIDS (Kid's wear)	
	Yes	No	Yes	No	Yes	No
Class 1	0.6390	0.3610	0.7793	0.2207	0.6423	0.3577
Class 2	0.0000	1.0000	1.0000	0.0000	0.1397	0.8603
Class 3	0.1674	0.8326	0.0272	0.9728	0.0000	1.0000
Class 4	0.8770	0.1230	0.7205	0.2795	0.7686	0.2314
Class 5	0.2936	0.7064	0.0000	1.0000	1.0000	0.0000

	HIGHFASHION		MADEUPS		ACCESSORIES	
	Yes	No	Yes	No	Yes	No
Class 1	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
Class 2	0.0581	0.9419	0.0000	1.0000	0.0298	0.9702
Class 3	0.3149	0.6851	0.0787	0.9213	0.5109	0.4891
Class 4	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
Class 5	0.0000	1.0000	0.0325	0.9675	0.0812	0.9188

	MIXED	
	Yes	No
Class 1	0.0000	1.0000
Class 2	0.2019	0.7981
Class 3	0.1837	0.8163
Class 4	0.0215	0.9785
Class 5	0.0000	1.0000

Class 3 firms and Class 2 firms deal in some in some high fashion (Class 3). Class 5 is of new firms not in high fashion. Class 1 only deals with traditional products (men's wear, women's wear, kid's wear) but not high fashion / accessories / mixed / made-ups. Class 4 represents new manufacturers (located in tier 2 and 3 cities) dealing in all traditional products but not high fashion / accessories / mixed / made-ups.

4.4 Class Membership

The Class Membership probabilities are given in Table 5.

Table 5: Probabilities of membership of classes

	Class 1	Class 2	Class 3	Class 4	Class 5
Estimated class population shares	0.1757	0.1022	0.1523	0.3854	0.1844
Predicted class memberships (by modal posterior prob.)	0.1856	0.0958	0.1467	0.3982	0.1737

Class 4 has the highest membership probability, while Class 2 has the lowest membership probability. The classes are well distributed.

5. CONCLUSION

The major benefit of latent class analysis is that it can take in various types of variables including categorical ones. Another important benefit is that the normality assumption is not required.

Five latent classes were found from the case. These are unobserved groupings from observed variables derived from probabilities estimated through E-M algorithm. It was seen that most MFA-period firms were merchants. Pre-MFA firms were large manufacturers located in tier 1 cities and dealing in almost all products (men's, women's, kid's wear) but not high fashion and accessories. It was found that new firms were mostly moving towards tier 2 and tier 3 cities.

The benefit is that policymakers can use the groupings to target policy more properly. A firm can be associated with the respective latent group and the impact of the policy can be observed from the respective panel. Therefore, an incentive

structure can be better designed as this method can take into consideration the dynamic business context.

This study is relevant to any industry. Care has to be taken to take sufficient manifest variables so as to capture the dynamic business context. These variables can be qualitative in nature as there is no limitation for the latent class analysis method. Choice of the number of latent classes has to be validated from the business context. A latent class analysis, done in this fashion, is likely to reveal interesting groupings which were unobserved earlier, in any industry.

REFERENCES

Agresti, A., 2002, *Categorical Data Analysis.* Hoboken: John Wiley & Sons.

Andersen, E. B., 1982, Latent Structure Analysis: A Survey. *Scandinavian Journal of Statistics, 9*(1), 1-12.

Audretsch, D., Sanders, M., & Zhang, L., 2011, When You Export Matters. *European Economic Association Annual Meeting 2011.*

Chen, Y., Hua, D., & Liu, F., 2009, Genaralized latent class analysis based on model dominance theory. *International Journal on Artificial Intelligence Tools, 18*(5), 739-755.

Cullmann, A., 2012, Benchmarking and firm heterogeneity: a latent class analysis for German electricity distribution companies. *Empir Econ, 42*, 147–169.

DeSarbo, W. S., Wang, Q., & Blanchard, S. J., 2010, Exploring intra-industry competitive heterogeneity - The identification of latent competitive groups. *Journal of Modelling in Management, 5*(2), 94-123.

Green, P. E., Carmone, F. J., & Wachspress, D. P., 1976, Consumer Segmentation via Latent Class Analysis. *Journal of Consumer Research, 3*, 170-174.

Kim, Y.-H., & Lee, K.-H., 2011, Typology of Fashion Product Consumers: Application of Mixture-model Segmentation Analysis. *Journal of the Korean Society of Clothing and Textiles, 35*(12), 1440-1453.

Kreuter, F., Yan, T., & Tourangeau, R., 2008, Good item or bad—can latent class analysis tell?: the utility of latent class analysis for the evaluation of survey questions. *J. R. Statist. Soc. A, 171*(3), 723–738.

Laird, N. M., 1978, Empirical Bayes Methods for Two-Way Contingency Tables. *Biometrika, 65*(3), 581-590.

Linzer, D. A., & Lewis, J. B., 2011, poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software, 42*(10), 1-29.

Lockshin, L., & Cohen, E., 2011, Using product and retail choice attributes for cross-national segmentation. *European Journal of Marketing, 45*(7/8), 1236-1252.

Nylund, K. L., Asparouhov, T., & Muthén, B. O., 2007, Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural equation modeling, 14*(4), 535–569.