
Compiling Respondent Burden Items: A Composite Index Approach

Daniel K. Yang

CE Survey Methods Symposium

2015

U.S. Bureau of Labor Statistics

Content does not represent BLS policy

Gemini Project of Redesign

- ❑ Primary goal: redesign the CE Surveys to improve data quality, through a verifiable reduction in measurement error.
- ❑ Expectation: a reduction in burden (or not inducing extra) on respondents could contribute to data quality.
- ❑ **How to best assess respondents'** perceived level of burden is still an open question.

Burden Items

- Research Section: at the end of the final interview wave, respondents were asked a series of research questions, including ten questions that ask respondents to assess their perceived burden.
- In psychometric approach, multivariate items can be deployed to measure different behavior aspects or dimensions of respondents.

Composite Burden Index

- An index based on these multiple items will often be developed by psychologists to reflect latent constructs of a **respondent's behavior or perception.**
- Objective: our purpose is to develop a composite burden index for CE to implement and to track perceived respondent burden over time.

Data Sample

- Burden item data were collected between October, 2012 (the 4th quarter) and March, 2013 (the 1st quarter) in the CE Research Section (not on going).
- Exclude households with missing values in any of the burden item questions.
- Final sample total is 3,247 households: 1,636 in 2012 4th quarter and 1,611 in 2013 1st quarter, respectively.

Burden Items

- Burden items asked respondent's feeling about:
 - ▶ The length of the interview.
 - ▶ How interesting was this survey to respondent?
 - ▶ How difficult or easy for respondent you to answer the questions?
 - ▶ Too many interviews or seem like a reasonable number?
 - ▶ How burdensome was this survey to respondent?
 - ▶ Sensitivity?
 - ▶ Agree to take a similar survey in the future?
 - ▶ Willing to extend survey for another 15 minutes?
 - ▶ Amount of effort that respondent put into answering survey?
 - ▶ Trust Census to safeguard respondent's information?

Multivariate Index

- An index is generally produced by combining measurements of related items with specified rules or equations into a numeric score which reflects the construct of interest.
- Must account for data structure.
- This situation requires multivariate techniques.

Criteria for Methods

Criterion	Desired Property
Variance Contribution	Consistently estimate the proportion of variance explained
Multivariate Capacity	Suitable for multivariate burden items
Likert Scale Capacity	Suitable for Likert Scale (categorical) burden items
Items Utilization	Use all burden items information (suppose all are important)
Independent Components	Produce independent (perpendicular) components from burden items
Composite Scores	Able to compute composite scores of burden items
Assumptions	Few assumptions to check
Model Subjectivity	Involve none or the least subjective model
Computation Time	Reasonable computing time
Application	Easily understood for implementation

Multivariate Techniques

- Multiple Correspondence Analysis (MCA)
 - ▶ Visualize how rows and columns are associated in two-way contingency tables.
- Multidimensional Scaling (MDS)
 - ▶ Transform multivariate data into lower dimensional space.
- Propensity Scores (PS)
 - ▶ Provide a univariate measure of information from multiple variables under assumptions.
- Classification and Regression Tree (CART)
 - ▶ Partition households to distinguish groups.

Principal Component Analysis (PCA)

- Most common method to aggregate multivariate items by allocating weights, and summarize into a composite measure.
- Disadvantages of PCA on Likert Scales:
 - ▶ Violation of the normal assumption
 - ▶ Inaccurate covariance or correlation estimation
 - ▶ May not preserve the order of categorical variables
- **Spearman's Rank Correlation PCA**
 - ▶ Free of Normal assumption
 - ▶ Required levels of Likert Scales: ≥ 5 , 7 preferred

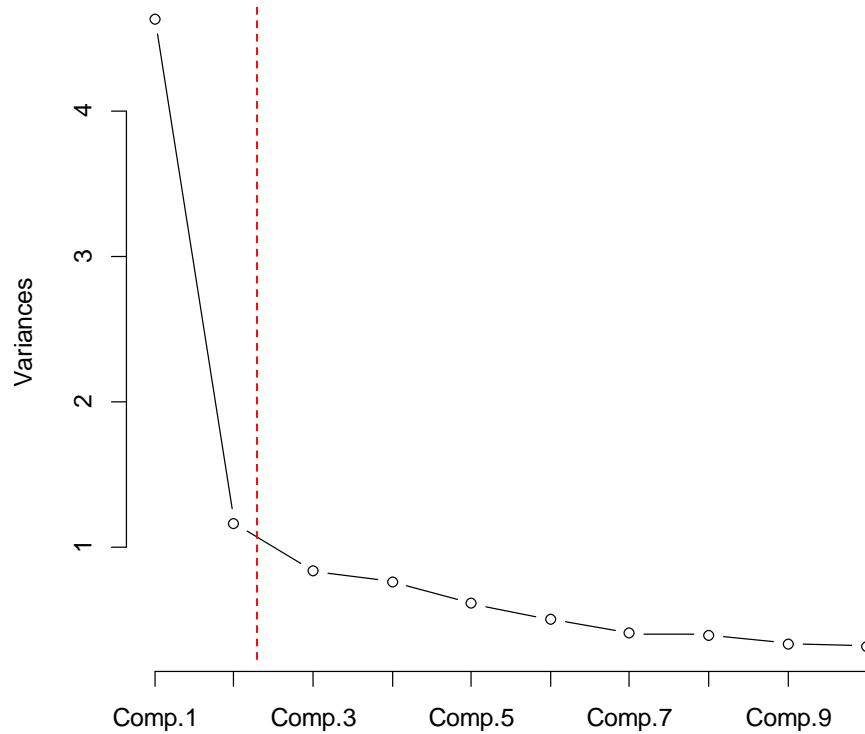
Recommend Polychoric PCA

- Kolenikov and Angeles (2004): PCA on a polychoric correlation (estimates from two observed ordinal variables for continuous Normal latent variables) matrix provides the most desirable performance. In terms of CE burden items:
 - ▶ Provide consistent estimates of proportion of variance explained from the components.
 - ▶ Handle multivariate burden item inputs and capable for Likert Scale.
 - ▶ Utilize all burden items in the analysis.
 - ▶ Produce independent components.
 - ▶ Produce loadings of all burden items for each independent component to compute a composite score.
 - ▶ Minimize efforts to check assumptions and model subjectivity.
 - ▶ Use well-known statistical procedures to implement.

Develop overall composite burden index scores by Polychoric PCA

- Compute a Polychoric correlation matrix for the burden item data.
- Compute PCA on this Polychoric correlation matrix.
- Determine the number of principal components to be selected.

Screeplot of Principal Components in R



Overall composite burden index scores by Polychoric PCA (cont.)

- Propose proportional weighted summary to compute the overall composite burden index scores based on Polychoric PCA:

$$\sum_{i=1}^k p_i \times X \times S_i, S_i = \rho \times PC$$

- vs. a simplified way is to compute a simple summation.

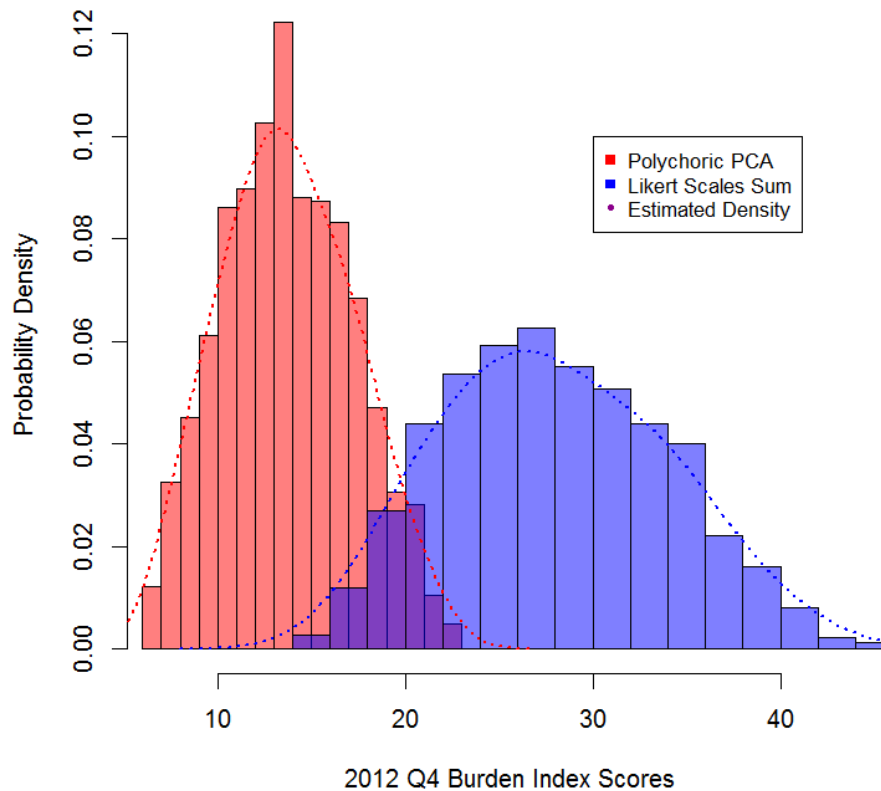
Evaluation of the Overall Composite Burden Index Scores

- Distributional properties.
- Weighted means of the overall composite burden index scores and proportions of characteristics.
- Association of overall composite burden index scores vs. characteristics of interest.

Distributional Properties (1)

- Distributional Properties
 - ▶ For simplicity, let us assume simple random sampling without accounting for complex designs and weights.
 - ▶ Superimposed histograms with estimated density overlays and unweighted descriptive statistics.

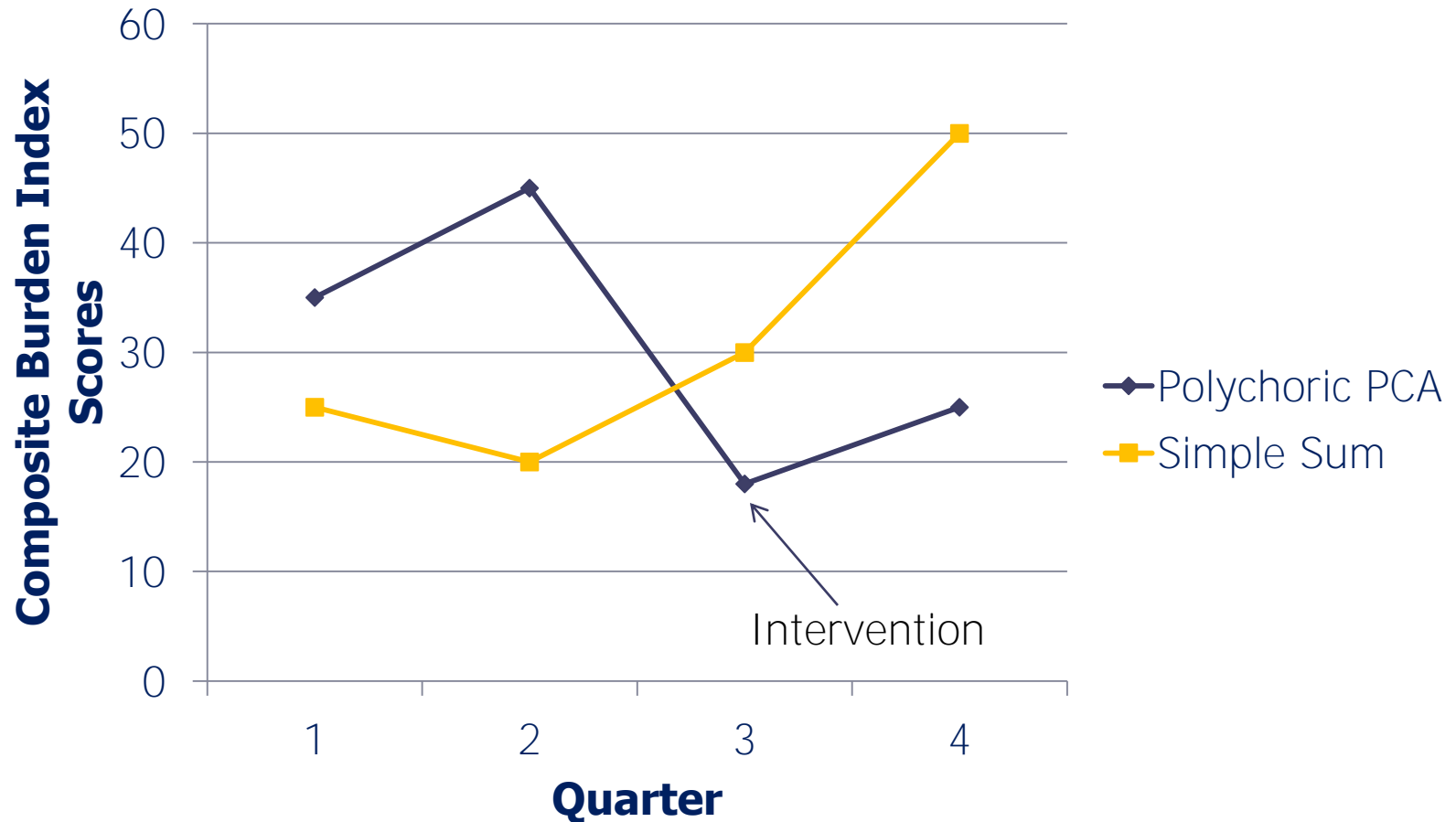
Figure 1. Histograms with estimated density overlays of burden index scores



Distributional Properties (2)

- Distributional Properties
 - ▶ For simplicity, let us assume simple random sampling without accounting for complex designs and weights.
 - ▶ Polychoric PCA: a narrower distribution with an approximate symmetric shape.
 - ▶ Simple summation: right skewed distribution with a wider spread.

Figure 2. Example of Composite Burden Index Scores Over Quarters



Evaluation of the Overall Composite Burden Index Scores

- Weighted means of the overall composite burden index scores and proportions of characteristics
 - ▶ In order to take into account the CE complex design, we conducted design-based model analyses with complex design weights and balanced repeated replication (BRR) with a Fay factor, $K = 0.5$ for variance estimation.
 - ▶ Characteristics: interview length, records usage, information booklet usage, interview mode and household size.

Table 1. Weighted Means of Composite Burden Scores and Interview Length

Variable	2012 4 th Quarter	2013 1 st Quarter
	Mean (SE)	Mean (SE)
Polychoric PCA Composite burden index scores	13.74 (0.12)	13.91 (0.14)
Household size	2.44 (0.05)	2.40 (0.043)
Interview length (minutes)	68.06 (1.13)	67.70 (1.20)

Note: SE – standard error.

Table 2. Weighted Characteristic Proportions

Characteristics	2012 4 th Quarter	2013 1 st Quarter
Records usage	Percentage (SE)	Percentage (SE)
Always or almost always	10.02 (0.64)	7.93 (0.69)
Most of the time	7.53 (0.42)	7.76 (0.53)
Occasionally	10.20 (0.55)	10.84 (0.62)
Never or almost never	21.99 (0.84)	23.73 (0.80)
Information booklet usage		
Always or almost always	11.93 (0.92)	3.80 (0.43)
Most of the time	4.33 (0.44)	3.51 (0.37)
Occasionally	3.42 (0.39)	8.25 (0.77)
Never or almost never	9.08 (0.72)	23.22 (0.84)
No Access	21.00 (0.94)	50.25 (0.53)
Interview mode		
Personal visit	33.44 (0.81)	31.47 (0.93)
Phone	16.31 (0.88)	18.78 (0.77)

Evaluation of the Overall Composite Burden Index Scores

- Association of overall composite burden index scores vs. characteristics of interest
 - ▶ Again, design-based model analyses with complex design weights and balanced repeated replication (BRR) with a Fay factor, $K = 0.5$ for variance estimation to account for the CE complex design.
 - ▶ Household size is removed from model due to its colinearity.

Table 3. Weighted Regression Coefficient Estimates by Quarter and Differences

Effect Parameter	2012 4 th Quarter	2013 1 st Quarter	Difference
	Estimate: θ_1	Estimate: θ_2	$\theta_2 - \theta_1$
Intercept**	12.07 (0.29)**	12.93 (0.54)**	0.85 (0.53)
Interview length	0.009 (0.002)**	0.003 (0.003)	-0.006 (0.004)
Records usage**			
Always or almost always	-0.44 (0.24) [†]	0.004 (0.34)	0.45 (0.42)
Most of the time	-0.44 (0.28)	0.22 (0.31)	0.67 (0.39)
Never or almost never	0.63 (0.20)**	0.83 (0.23)**	0.21 (0.29)
Ref.: Occasionally			
Information booklet usage**			
Always or almost always	0.18 (0.28)	-0.86 (0.40)*	-1.05 (0.49)*
Most of the time	0.68 (0.33) [†]	-0.77 (0.40) [†]	-1.41 (0.45)**
Never or almost never	1.10 (0.27)**	0.65 (0.51)	-0.45 (0.49)
Occasionally	0.16 (0.33)	0.84 (0.51)	0.69 (0.54)
Ref.: No Access			
Interview mode**			
Phone	1.82 (0.28)**	1.10 (0.40)**	-0.72 (0.40)
Ref.: Personal Visit			

Legend: († p-value < 0.10, * p-value < 0.05, ** p-value < 0.01

Conclusions

- Recommended the Polychoric PCA to develop composite burden index.
- Advantages in terms of improved distribution.
- Household size was excluded because of collinearity and interview length may not be a good predictor.
- Verified significant associations between overall composite burden index scores and records usage, information booklet usage and interview mode.
- The significances of those three characteristics are stable among two quarters (between quarters coefficient estimates were not that different).

Recommend Future Steps

- An improvement in measurement precision does not **necessarily imply a reduction in respondents' perceived burden.**
- Polychoric PCA method provides an option to track **the changes of the overall respondents' perceived burden** after implementing the redesign changes.
- Will provide input for program managers during the decision making procedure of intervention, and keep tracking after the intervention.
- We tentatively plan to implement a burden index by 2017 as a baseline.

THANK YOU!

2015 Consumer Expenditure Survey (CE)
Microdata Users' Workshop Schedule



July 15-17, 2015

www.bls.gov

Acknowledgement

- Branch of Research and Program Development (BRPD), DCES.
 - ▶ Adam N. Safir, Chief
 - ▶ Laura P. Erhard, Economist
 - ▶ Lucilla Tan, PhD, Economist
- Office of Survey Methods Research (OSMR)
 - ▶ Wendy L. Martinez, PhD, Director, Mathematical Statistics Research Center (MSRC)
 - ▶ Brandon M. Kopp, PhD, Research Psychologist
 - ▶ Scott S. Fricker, PhD, Research Psychologist
 - ▶ Morgan S. Earp, PhD, Research Statistician

Contact Information

Daniel K. Yang

Research Mathematical Statistician
Office of Survey Methods Research (OSMR)

www.bls.gov/osmr/home.htm

yang.daniel@bls.gov

Disclaimer: Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the Bureau of Labor Statistics.

