Quantification of qualitative data: the case of the Central Bank of Armenia

Martin Galstyan[1] and Vahe Movsisyan[2]

Overview

The effect of non-financial organisations' and consumers' attitudes on economic activity is a subject of great interest to both policymakers and economic forecasters. The Business Tendency and Consumer Surveys of the Organisation for Economic Co-operation and Development (OECD) are important sources of information about household opinion and non-financial organisations' expectations concerning current economic conditions and future developments.

The information collected in such surveys is mainly qualitative because respondents are asked to assign qualities, rather than quantities, to the variables of interest. For example, in a business tendency survey, respondents are asked to assign qualities to the value of their order books, such as "higher than normal", "normal" or "below normal".

It is generally much easier for respondents to give qualitative rather than quantitative information. As a result, the questionnaires can be completed quickly and results of the surveys can be published earlier than the results of traditional statistical surveys. This is one of the main advantages of qualitative surveys.

Thus, before computing the final indices, it is necessary to quantify the qualitative data collected, and it is very important to select the right quantification methods.

Quantification of qualitative data

Quantitative analysis is the numeric representation and manipulation of qualitative observations for the purpose of describing and explaining the event that those observations reflect.

The analysis of how economic agents form their expectations about economic variables has been treated as one crucial issue in explaining many important economic trends.

While there is a vast literature on this topic, no consensus has been achieved among researchers on how to quantify the expectation survey data.

There are at least four main approaches to converting the results of qualitative surveys to standard quantitative variables.

The first method is some variant of the probabilistic approach. The principle behind this approach is that the respondents reply that the value of the reference variable x can be described by a certain statement (eg x stays stable) if it lies between two known thresholds

[1] Central Bank of Armenia, Statistics Department, Vazgen Sargsyan str 6, Yerevan 0010, Republic of Armenia. E-mail: m.galstyan@cba.am

[2] Central Bank of Armenia, Statistics Department, Vazgen Sargsyan str 6, Yerevan 0010, Republic of Armenia. E-mail: vmovsisyan@cba.am

(eg ± 5 per cent around its initial value). Thus, by assuming that the functional form of the underlying probability distribution of opinions and expectations about x is known, the average value of x can be expressed as a function of the given thresholds. A measure of heterogeneity of opinions and operators' uncertainty can also be derived within the same analytical framework.

The second approach is based on regression techniques aimed at estimating the value of x underlying each qualitative answer. This method requires the regression of a standard quantitative measure of x against the time series of percentage of people who gave each qualitative answer.

The third method regards the percentages of each qualitative answer as a function of a common "latent measure" of x observed by respondents. The usual multivariate techniques can help in estimating the dynamics or the sectoral variations (but not the absolute level) of the latent factor affecting the opinions and expectations expressed by the interviewed operators.

The time series of percentages of answers collected in qualitative surveys are closely correlated. In the first place, this fact implies that the latent variable approach is possibly sound and reliable. However, it also suggests that even very sophisticated methods, based on complicated transformations of original percentages, tend to produce indicators that follow the common trend and cycles that can be easily deduced by any time series of percentages, or a simple combination thereof. This explains and justifies the widespread use of the "balance" between the percentages of "optimistic" and "pessimistic" answers.

The fourth method is the calculation of diffusion and composite indices. This method is a summary measure designed to facilitate the analysis and forecast of business cycles combining the behaviour of a group of economic indicators which represent different economic activities such as production and employment.

Diffusion indices aggregate the directions of change of a selected series to detect a business cycle phase, while composite indices aggregate the percentage changes of a selected series to detect the volume of a business cycle.

The empirical verification of performance of the various methods is mixed. Generally, no one procedure outperforms another, even if some authors have pointed out the sharp inefficiency of balance statistics and others have noted that dynamic regression models are generally superior. Table 1 provides a swift comparison of different approaches.

Table 1

Quantification methods

Method	Main assumptions	Advantages	Drawbacks
Probabilistic	• The functional form of opinions about the relevant variable *x* is known. • Respondents reply "*x* did (will) remain stable" if *x* lies between two given thresholds. Additional assumptions are required in a polychotomous case.	• The results depend only on the observed percentages of answers and only to a minor extent on the assumptions about the probability distribution of the variables and the thresholds assumed by respondents.	• The time series of results may be very volatile if some special combinations of answers occur. • Other information on *x* is completely neglected, even when it is available. • The treatment of polychotomous questions may be complicated.

Table 1 (cont)

Quantification methods

Method	Main assumptions	Advantages	Drawbacks
Regression	• Respondents also attach to each qualitative answer a reference value of *x*. • Reference values can be estimated by using regression models.	• It is very general, regardless of the wording of questions and the number of answers authorised. • Integration into standard econometric models is straightforward.	• A reference quantitative time series is needed. • Estimation can be flawed by multicollinearity and numerical convergence problems.
Latent factor	• A single common "latent factor" drives each percentage of answers.	• It is very general. • In principle, no extraneous information is needed. However, they can be exploited as well. • The same quantified indicator may be used in both preliminary analysis and econometric modelling.	• Very short time series of answers cannot be treated.
Diffusion and composite indices	• Diffusion indices measure one half of the respondents reporting "no change" and all respondents reporting • "positive" answers. The reason why a group of indicators combined into a composite indicator should be more reliable over a period of time than any of its individual components is related to the nature and causes of business cycles.	• Easy to compute • The performance of individual indicators will then depend on the causes behind a specific cycle. Some indicators will perform better in one cycle and others in a different cycle. It is therefore necessary to have signals for many possible causes of cyclical changes, and to use all potential indicators as a group.	• These indices are more volatile than indices constructed with the methods listed above.

Of the methods described, we use a diffusion and composite indices method for construction of the consumer confidence index (CCI), the economic activity index (EAI) and the business climate index (BCI).

Purposes and tasks of the surveys

In response to the widespread belief that consumers' opinions and expectations influence the direction of the economy, a growing number of studies have set out to analyse the relationship between consumer attitudes and economic variables.

Taking this into consideration, the estimation of household expectations regarding the economy, as an ultimate private sector driver of market economies, is an important factor in the organisation and implementation of macroeconomic policies.

For observation of household perspectives on the current economic situation and estimation of their expectations regarding future shifts in the economy, the Statistics Department of the Central Bank of Armenia (CBA) has conducted quarterly consumer confidence surveys since the first quarter of 2005. The main purpose of the surveys is to estimate consumer behaviour in the light of their expectations of current and future economic conditions, and to calculate the CCI.

In order to achieve these goals, the following tasks are performed:

- Analysis of household opinions regarding the overall economic situation (current and future)
- Analysis of household opinions regarding their material security (current and future)
- Calculation of indices of current and future conditions.

The Statistics Department of the CBA also conducts business tendency surveys, also known as economic activity and business climate surveys. The main purpose of these surveys is to ask managers of non-financial organisation about the current status of their business and their plans and expectations for the near future. These surveys provide information that is valuable to the respondents themselves and to economic policymakers and analysts. Although they do not provide precise information on levels of output, sales, investment or employment, they can be used to predict changes in these aggregates, and for that reason, they are particularly useful for analysing the business cycle.

Survey methodology

Consumer confidence survey

The survey is conducted in the second month of each quarter, with time-independent samples of households, and covers all Armenian households.

In order to facilitate the analysis of the evolution of phenomena over time, starting with the next survey, a part of the sample will comprise households interviewed in previous surveys (panel households). Panel households will represent about 40 per cent of the sample.

Data are collected from households by means of telephone interviews. The survey sample size ranges from 1,700 to 2,000. The sampling of survey has stratified one stage sample design without replacement:

The whole universe was divided into administrative subdivisions called strata. The city of Yerevan was divided into communities (strata) and its regions into districts.

The sample units are selected randomly from each stratum. The sample size of each stratum is proportional to its population.

Survey questions are drafted with the aim of eliciting useful information without imposing an undue burden on respondents. The questions are generally qualitative and have a three-point scale of response (increase, stable, decrease). Quantitative questions are also included, but generally confined to demographic aspects of households. The questionnaires also contain a question about household income, but because of the sensitivity of this question, it is suggested that households situate their income within one of the given ranges.

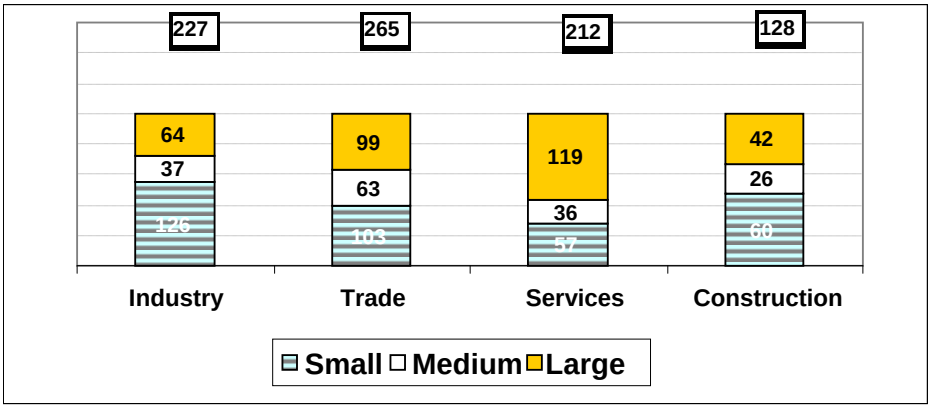
Economic activity and business climate surveys

As indicated, the main purpose of the quarterly EAIs and BCIs is to analyse the expectations and perspectives of economic agents concerning each branch of the economy (industry, construction, trade and services).

For the economic activity and business climate surveys, the sample of non-financial organisations is constituted by a non-probability sampling method: cutting off the tail. The sample comprises the largest organisations that account for at least 80 per cent of the gross profit of a particular branch or segment of the economy. Thus, for the first quarter of 2009, the survey sample consists of 832 companies (Figure 1).

Figure 1

Structure of non-financial organisations sample



The survey is conducted by telephone, letter and facsimile.

Furthermore, as the four surveyed branches account for the largest share of Armenia's GDP (Figure 2), it is also important to analyse the correlation between the EAI, BCI and value added of the respective branches.

Figure 2

Weights of four surveyed branches in GDP

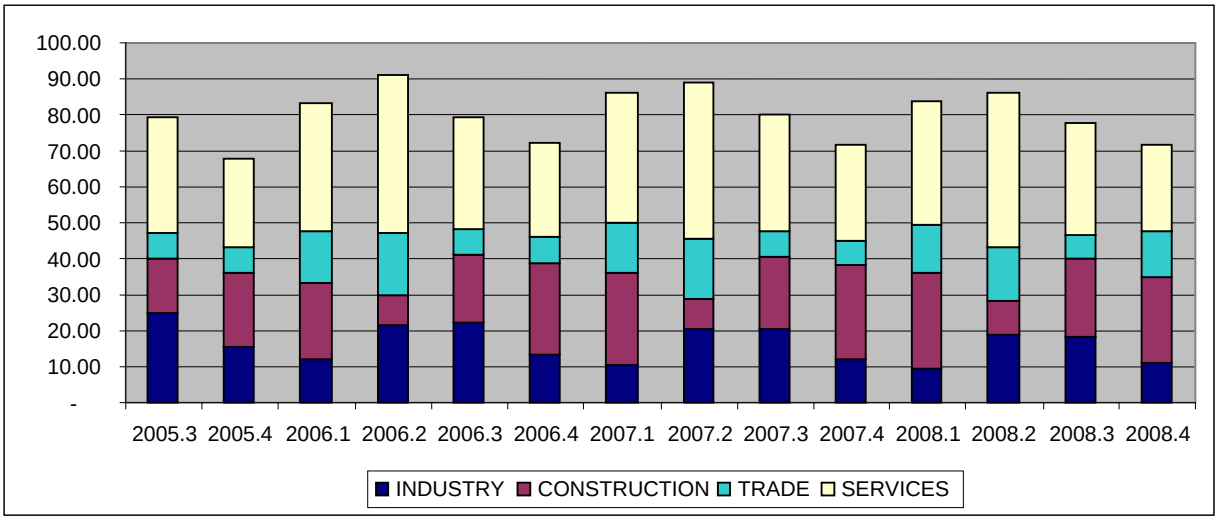


Figure 2 shows that the four branches together account for about 80 per cent of GDP.

Tables 2 and 3 describe all the steps in implementing the diffusion and composite indices method for construction of the consumer confidence, economic activity and business environment indices.

Table 2

CCI construction

Phase 1	**Balance of current conditions for each community**	$BA^{current}{}_i = \sum_{j=1}^{3} (Answer^{current}{}_{pos} + \frac{1}{2} Answer^{current}{}_{neu})$ $BA^{current}{}_i$ – balance of current conditions for i-th community $Answer^{current}{}_{pos}$ – positive answers to each *current* question $Answer^{current}{}_{neu}$ – neutral answers to each *current* question
	Balance of future conditions for each community	$BA^{future}{}_i = \sum_{j=1}^{3} (Answer^{future}{}_{pos} + \frac{1}{2} Answer^{future}{}_{neu})$ $BA^{future}{}_i$ – balance of current conditions for i-th community $Answer^{future}{}_{pos}$ – positive answers to each *future* question $Answer^{future}{}_{neu}$ – neutral answers to each *future* question
Phase 2	**Total balances for each community**	$BA_i = \frac{1}{2}\left(BA^{current}{}_i + BA^{future}{}_i\right)$ BA_i – average balance of answers to all questions for i-th community
Phase 3	**Weighted balance of all households**	$WBA = \sum_i W_i^{pop} BA_i$ WBA – weighted balance of answers for all households W_i^{pop} – weight of population in i-th community
Phase 4	**Index computation (CCI)**	$Index = \frac{WBA_1}{WBA_0} \cdot 100$ 1 and 0 refer to current and base period, respectively.

Table 3

Three-step weighted method of EAI and BCI construction

Phase 1	**Balance of answers**	$$BA_{ijq} = \sum_{k=1}^{p} W_{EmplK}\left(Answer_{pos} + \frac{1}{2} Answer_{neu}\right)$$ BA_{ijq} – balance of j-th segment of i-th branch for q-th question W_{EmplK} – weight of k-th organisation's employees in all organisations of that segment $Answer_{pos}$ – positive answers to each question $Answer_{neu}$ – neutral answers to each question
Phase 2	**Balance of answers for branch**	$$WBA_{iq} = \sum_{j=1}^{n} BA_{ij} \cdot RW_{ij}$$ WBA_{iq} – weighted balance of i-th branch questions for q-th question RW_{ij} – weight of profit of the j-th segment in i-th branch
Phase 3	**Average balance of answers**	$$WBA_{iAverage} = \frac{\sum_{q=1}^{4} WBA_{iq}}{4}$$ $WBA_{iAverage}$ – average balance of answers to all questions included in i-th branch
Phase 4	**Balances of answers for the overall economy**	$$TBA = \sum_{i=1}^{4} WBA_{iAverage} \cdot VA_i$$ TBA – weighted balance of answers for the overall economy VA_i – weight of i-th branch value in cumulative value of all four branches
Phase 5	**Index computation (EAI, BCI)**	$$Index = \frac{TBA_1}{TBA_0} \cdot 100$$ 1 and 0 refer to current and base period, respectively.

To investigate the possible relationships between economic activity and business climate questions, we need to analyse the correlations between the branches' variables (questions) and the value added for each branch.

Correlation matrices are presented in the appendices.

We have to underline that here, in correlation analysis, the variables concerning the future expectations of respondents are taken with a "+ 1" lag: for example, in industry, the expectation of respondents (taken at quarter t) regarding the demand for their products for

the $t+1$-th quarter is correlated with the actual growth of industry value added for the $t+1$-th quarter.

As we can see, in both industry and construction, almost all individual economic activity questions have strong positive correlations with the quarterly growth rate for that branch.

In trade, only one question (VolumeC) is significantly correlated (0.719) with the growth of trade value added.

From Table 4 it is obvious that in industry, construction and trade, the branch economic activity analysis better describes the appropriate branch of the economy than business climate analysis. In particular, in construction, the coefficient of correlation between the weighted balances of economic activity questions and the construction growth rate equals 0.842 (it is significant at the 0.01 level).

Table 4

Correlation matrix of EA and BC weighted balances of answers (WBA) and growth of branches

	EA WBA	BC WBA
Industry growth	0.719**	0.360
Construction growth	0.842**	0.540*
Trade growth	0.668**	0.240
Services growth	0.476	0.480

** The correlation is significant at the 0.01 level (2-tailed). * The correlation is significant at the 0.05 level (2-tailed).

Conclusion

To summarise the analysis, we can conclude that the household and corporate sector surveys conducted by the CBA since 2005 are important sources of information that describes the main sectors of the national economy. The indices, calculated by the diffusion and composite index method, particularly the economic activity indices of industry, construction and trade, can be used as leading growth indicators for the corresponding sectors' value added. This shows that the selected quantification method (diffusion and composite index method) works well for the intended purposes.

We should mention that we have short time series of composite indices (data for 14 quarters), and this can lead to overestimating the reliability of the results.

In the future, with enlargement of the survey database, it will be possible to obtain more reliable estimates of indicators that can be used for preliminary forecasting of the development of the Armenian economy.

Appendix 1:
Correlation matrix (industry)

	Balance of economic activity questions				**Balance of business environment questions**				**Balance of other questions**		
	VolumeC	VolumeF	StockF	DemandF	RisksC	RisksF	SubindC	SubindF	PriceC	WageF	EmployeeF
VolumeC	1										
VolumeF	0.777**	1									
StockF	0.575*	0.578*	1								
DemandF	0.807**	0.905**	0.633*	1							
RisksC	0.45	0.37	0.755**	0.39	1						
RisksF	0.30	0.13	0.543*	0.23	0.657**	1					
SubindC	0.592*	0.34	0.527*	0.37	0.679**	0.606*	1				
SubindF	0.585*	0.51	0.526*	0.680**	0.677**	0.521*	0.657**	1			
PriceC	0.44	0.48	0.730**	0.48	0.634*	0.48	0.33	0.604*	1		
WageF	0.555*	0.26	0.34	0.24	0.519*	0.605*	0.846**	0.51	0.27	1	
EmployeeF	0.606*	0.33	0.634*	0.50	0.597*	0.682**	0.781**	0.700**	0.43	0.782**	1
Ind growth	**0.753****	**0.726****	**0.23**	**0.647***	**0.10**	**–0.18**	**0.49**	**0.53**	**0.30**	**0.38**	**0.25**

** The correlation is significant at the 0.01 level (2-tailed). * The correlation is significant at the 0.05 level (2-tailed).

Appendix 2:
Correlation matrix (construction)

	Balance of economic activity questions			**Balance of business environment questions**				**Balance of other questions**		
	VolumeC	VolumeF	DemandF	RisksC	RisksF	SubindC	SubindF	PriceC	WageF	EmployeeF
VolumeC	1									
VolumeF	0.870**	1								
DemandF	0.843**	0.931**	1							
RisksC	–0.436	–0.270	–0.212	1						
RisksF	0.068	0.281	0.058	–0.179	1					
SubindC	0.565*	0.555*	0.450	–0.044	–0.072	1				
SubindF	0.611*	0.779**	0.787**	–0.125	0.361	0.294	1			
PriceC	0.446	0.760**	0.646**	–0.039	0.390	0.412	0.604*	1		
WageF	0.678**	0.826**	0.896**	–0.005	–0.072	0.320	0.758**	0.624*	1	
EmployeeF	0.764**	0.926**	0.952**	–0.232	0.145	0.409	0.781**	0.760**	0.879**	1
Const growth	**0.764****	**0.833****	**0.785****	**–0.464**	**0.443**	**0.044**	**0.800****	**0.515**	**0.622***	**0.819****

** The correlation is significant at the 0.01 level (2-tailed). * The correlation is significant at the 0.05 level (2-tailed).

Appendix 3:
Correlation matrix (trade)

	Balance of economic activity questions			**Balance of business environment questions**				**Balance of other questions**		
	VolumeC	VolumeF	DemandF	RisksC	RisksF	SubindC	SubindF	PriceC	WageF	EmployeeF
VolumeC	1									
VolumeF	0.380	1								
DemandF	0.584*	0.727**	1							
RisksC	0.024	0.160	0.025	1						
RisksF	0.354	0.059	–0.009	0.617*	1					
SubindC	0.650**	0.523*	0.634*	0.415	0.348	1				
SubindF	0.154	0.673**	0.619*	0.237	–0.040	0.363	1			
PriceC	0.059	0.375	0.038	0.330	–0.058	0.063	0.290	1		
WageF	0.114	0.111	0.340	–0.505	–0.381	0.091	0.267	0.065	1	
EmployeeF	0.374	0.123	0.617*	–0.139	–0.157	0.518*	0.260	–0.105	0.629*	1
Trade growth	**0.719****	**0.358**	**0.233**	**–0.215**	**0.500**	**0.413**	**–0.011**	**–0.031**	**0.274**	**–0.043**

** The correlation is significant at the 0.01 level (2-tailed). * The correlation is significant at the 0.05 level (2-tailed).

Appendix 4:
Correlation matrix (services)

	Balance of economic activity questions			**Balance of business environment questions**				**Balance of other questions**		
	VolumeC	VolumeF	DemandF	RisksC	RisksF	SubindC	SubindF	PriceC	WageF	EmployeeF
VolumeC	1									
VolumeF	0.661**	1								
DemandF	0.732**	0.865**	1							
RisksC	0.553*	0.527*	0.532*	1						
RisksF	–0.041	–0.079	0.145	0.032	1					
SubindC	0.546*	0.394	0.628*	0.527*	0.410	1				
SubindF	0.343	0.627*	0.554*	0.082	0.234	0.386	1			
PriceC	0.492	0.437	0.384	0.467	0.394	0.413	0.441	1		
WageF	–0.107	0.228	–0.046	0.210	–0.379	0.144	–0.080	–0.136	1	
EmployeeF	0.042	0.013	–0.132	0.053	0.314	–0.313	0.067	0.262	–0.223	1
Serv growth	**0.384**	**0.412**	**0.467**	**–0.239**	**0.319**	**0.110**	**0.594***	**0.381**	**–0.516**	**0.090**

** The correlation is significant at the 0.01 level (2-tailed). * The correlation is significant at the 0.05 level (2-tailed)

.

Appendix 5:
Abbreviations of variables

VolumeC	Volume change (current)
VolumeF	Volume change (future)
StockF	Stock change (future)
DemandF	Demand change (future)
RisksC	Risks change (current)
RisksF	Risks change (future)
SubindC	Economic situation of segment (current)
SubindF	Economic situation of segment (future)
PriceC	Price change (current)
WageF	Average wage change (future)
EmployeeF	Employees change (future)
Ind growth	Quarterly growth rate of industry value added
Const growth	Quarterly growth rate of construction value added
Trade growth	Quarterly growth rate of trade value added
Serv growth	Quarterly growth rate of service value added

References

Abeyasekera, Savitri (2000): “Quantitative analysis approaches to qualitative data: why, when and how”, Statistical Services Centre, University of Reading.

Enrico D’Elia (2005): “Using the results of qualitative surveys in quantitative analysis”, Working Paper no 56.

Henzel, Steffen, and Timo Wollmershäuser (2005): “An alternative to the Carlson-Parkin method for the quantification of qualitative inflation expectations: evidence from the *Ifo World Economic Survey*”, Ifo Institute for Economic Research (Ifo), University of Munich, Working Paper no 9, June.

Nilsson, Ronny (2000): “Confidence indicators and composite indicators”, paper for presentation at the Centre for International Research on Economic Tendency Surveys (CIRET) Conference in Paris, 10–14 October.

Schenker, Rolf (2008): “Methods of quantifying qualitative data: a survey of recent project research and results”, ETH (Swiss Federal Institute of Technology), 29 July.