Warsaw School of Economics–SGH
Institute of Econometrics
Department of Applied Econometrics

ISSN 2084-4573

Department of Applied Econometrics Working Papers
Warsaw School of Economics–SGH
ul. Madalinskiego 6/8
02-513 Warszawa, Poland

Working Paper No. 3-12

Information content of survey data: applications of entropy and dissimilarity measures

Emilia Tomczyk
Warsaw School of Economics–SGH, Poland

Information content of survey data: applications of entropy and dissimilarity measures*

Emilia Tomczyk
Institute of Econometrics, Warsaw School of Economics, Poland

Abstract

This paper evaluates information content of survey data by the means of entropy and dissimilarity measures. Similarities between *a priori* information (expectations) and *a posteriori* information (realizations) are assessed for four variables originating from Polish business tendency survey and analyzed across two ownership sectors and four size categories. The measures employed include Shannon empirical entropy, Kullback-Leibler relative entropy, and Chomątowski-Sokołowski dissimilarity coefficient.

Results of empirical analysis allow to conclude that enterprise size does not significantly affect entropy values even though some size effects are observed in case of production and general business conditions variables. Public and private enterprises are not differentiated by interpretation of expectations horizon and therefore can be studied on the aggregated level without loss of important information on forecast horizon patterns. Additionally, previous findings (see E. Tomczyk, 2011) are generally confirmed on longer sample. Production time series are found to be characterized by the highest entropy, and prices data – the lowest; entropy of production is also found to be the least variable. In public enterprises, concentration of answers to the survey questions is higher and also more variable than in private sector.

Keywords: tendency surveys, expectations, entropy, dissimilarity of structures

JEL classification: C83, D84

* This paper has been submitted for the 31st CIRET Conference in Vienna.

1. Introduction

The second law of thermodynamics (the Law of Entropy) has been introduced to mainstream economics by H. Theil (1967) and N. Georgescu-Roegen (1971) who provided both theoretical background and inspiration for empirical applications. Early examples of these consist of assessments of social liberty and measurement of political and economic freedom (see D. Gabor, A. Gabor, 1979); this line of research continues with recent analyses of economic development and convergence, particularly in new member states of European Union (see S. Stattev, S. Raleva, 2006). A related subject, brought up by H. Theil in his *Economics and Information Theory* (1967), proposes entropy-based measures of social and economic inequality, later employed also to describe racial and income diversity, social isolation etc. Theil inequality index based on Shannon's definition of empirical entropy offers an advantage over the widely used Gini index: it can be decomposed to distinguish between-group and within-group inequalities. Theil's coefficient has been offered as replacement for Gini index by J. E. Foster and A. Sen (1997) and employed, for example, to evaluate wage inequalities in research conducted by Polish Ministry of Labour and Social Policy (I. Marcinkowska *et al.*, 2008). Krugman specialization index, based on related idea of similarities of structure, has been used by European Central Bank (ECB, 2004) to assess sector specialization and concentration.

A key branch of current economic applications of entropy focuses on sustainability, defined by the World Commission on Environment and Development (1987) as a call for continued economic expansion without environmental degradation. Numerous publications address issues of economic sustainability and related fields of ecological and environmental economics, inter-temporal allocation of resources, and environmental management; for a summary of macro- and microeconomic aspects of sustainability, see S. Borghesi (2008). Other economic applications of entropy include topics as diverse as cross-country demand analysis (H. Theil, D. Chen, 1996), description of technological evolution (K. Frenken, A. Nuvolari, 2004), analysis of learning processes (V. A. Singh *et al.*, 2010) and financial reporting (R. A. Nehmer, 2011). Advocates of more general use of entropy in economics – among them, E. T. Jaynes (1991) and A. Raine *et al.* (2006) – call for extensive reformulation of economic theory in terms of the second law of thermodynamics.

A less encompassing – but nevertheless promising – interpretation of entropy, so far relatively neglected in economic literature, calls for employing empirical measures of entropy of a probability distribution to assess information content of survey data. To the best of my knowledge, an effort to evaluate information gain from change of structure from its *a priori* to *a posteriori* form as a measure of degree of similarity (or dissimilarity) of structures defined on the basis of survey data has been first undertaken in E. Tomczyk (2011) where *a priori* structure is defined by fractions of respondents expressing expectations, and *a posteriori* structure – by fractions of respondents declaring observed changes in economic variables (realizations). Among the few Polish publications describing economic applications of entropy measures only one has previously mentioned the possibility of employing relative entropy to evaluate expert forecasts (see H. Kowalczyk, 2010). This paper follows analysis initiated in E. Tomczyk (2011) on employing measures of entropy and dissimilarity to evaluate information content of survey data.

Three reasons may be offered for attempting a study of information content of survey data. First, measures of similarity may provide the means to evaluate accuracy of expectations (predictions) in the business tendency survey. The larger the similarity between expectations and realizations, the better predictive power of expectations. Second, measures of information content may help to identify the actual forecast horizon used by respondents and determine whether public and private enterprises, or enterprises of various sizes, interpret it differently. This result may, in turn, be useful in other formal analyses of expectations, for example establishing appropriate number of lags in econometric models with expectations or defining dependent variables in quantification models. Third, use of entropy measures in analyzing survey data extends the set of methods available for formal analysis of expectations and realizations reported in business tendency surveys.

The structure of the paper is as follows. In section 2, measures of entropy, information content and dissimilarity of structures are described (following notation introduced in E. Tomczyk, 2011). Business tendency survey data are presented in section 3. Empirical results are reported in sections 4 (measures of entropy and dissimilarity of structures) and 5 (analysis of firm size effects). Section 6 concludes.

2. Measures of entropy, information content and dissimilarity of structures

Following E. Wędrowska (2009), let us define structure S^n as a vector $S^n = [s_1, s_2, \dots, s_n]^T \in R^n$ which elements s_i ($i = 1, 2, \dots, n$) fulfill two conditions:

$$0 \le s_i \le 1, \tag{1}$$

$$\sum_{i=1}^{n} s_i = 1. \tag{2}$$

Structure S^n is therefore fully described by a vector of fractions (structure elements). Information content of a message is defined in information theory in relation to the probability that a given message is received from the set of all possible messages: the less probable it is, the more information it carries. Empirical measure of entropy has been defined by C. E. Shannon (1948) as

$$H(S^n) = \sum_{i=1}^{n} s_i \log_2 \frac{1}{s_i}. \tag{3}$$

Basic properties of entropy measure $H(S^n)$ can be summarized as follows (see A. Rényi, 1961):

- $H(S^n) = H_{min} = 0$ if one of the elements s_i ($i = 1, 2, \dots, n$) is equal to 1 and all the remaining structure elements are equal to 0 (that is, distribution is concentrated in one element of structure only),
- $H(S^n) = H_{max} = \log_2 n$ if all structure elements s_i are equal (that is, $s_1 = s_2 = \dots = s_n$; distribution of structure elements is uniform).

$H(S^n)$ can be therefore interpreted as measure of concentration of elements s_i of structure S^n and used in empirical setting to evaluate information content of a structure.

In economic practice, extent of changes detected between assumed (*a priori*) and observed (*a posteriori*) structures may be of even greater interest than degree of uncertainty associated separately with *a priori* and *a posteriori* structures. Relative entropy (Kullback-Leibler divergence) can be used to evaluate the size of change between *a priori* structure S_p^n and *a posteriori* structure S_q^n:

$$I\left(S_q^n : S_p^n\right) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i}. \tag{4}$$

Again, basic properties of relative entropy include the following (see A. Rényi, 1961; E. Wędrowska, 2009):

- $I(S_q^n : S_p^n) = I_{min} = 0$ if both structures are identical (that is, $S_p^n = S_q^n$),
- $I(S_q^n : S_p^n)$ increases to infinity with the size of differences between the structures.

Relative entropy measures expected amount of "new" information provided by *a posteriori* structure. Accordingly, it is also known as information gain and interpreted as degree of change or dissimilarity between assumed (*a priori*) and observed (*a posteriori*) structures. The larger relative entropy is, the less similar the structures are.

In empirical setting, it is more convenient to apply a standardized coefficient defined on interval [0, 1] to facilitate interpretations and comparisons. S. Chomątowski and A. Sokołowski (1978) introduce a similarity coefficient to classify industrial production data into clusters, and provide a related dissimilarity measure that can be used to evaluate extent of change from *a priori* to *a posteriori* structure:

$$P\left(S_q^n : S_p^n\right) = 1 - \sum_{i=1}^{n} \min(q_i, p_i)\,. \tag{5}$$

Basic properties of the Chomątowski-Sokołowski coefficient can be summarized as follows:

- $P(S_q^n : S_p^n) = P_{min} = 0$ if both structures are identical (that is, $S_p^n = S_q^n$),
- $P(S_q^n : S_p^n)$ approaches 1 if as dissimilarities between structures increase.

To the best of my knowledge, entropy and dissimilarity measures described above were first used to analyze survey data in E. Tomczyk (2011). This paper updates and extends analysis of information content of expectations and realizations reported by survey respondents.

3. Data

Empirical part of this paper is based on qualitative business tendency surveys conducted by the Research Institute for Economic Development (RIED) at the Warsaw School of Economics. Launched in 1986 to monitor manufacturing industry, currently they cover households, farming sector, exporters, construction industry, and banking sector as well. In the monthly survey addressed to industrial enterprises (see Table 1), respondents are asked to evaluate both current situation (as compared to last month) and expectations for the next 3–4 months by assigning them to one of three categories: increase / improvement, no change, or decrease / decline. Aggregated survey results are regularly published in Research Institute for Economic Development bulletins (see RIED, 2012).

Table 1. Monthly RIED questionnaire in industry

		Observed within last month	Expected for next 3–4 months
01	Level of production (value or physical units)	up unchanged down	will increase will remain unchanged will decrease
02	Level of orders	up normal down	will increase will remain normal will decrease
03	Level of export orders	up normal down not applicable	will increase will remain normal will decrease not applicable
04	Stocks of finished goods	up unchanged down	will increase will remain unchanged will decrease
05	Prices of goods produced	up unchanged down	will increase will remain unchanged will decrease
06	Level of employment	up unchanged down	will increase will remain unchanged will decrease
07	Financial standing	improved unchanged deteriorated	will improve will remain unchanged will deteriorate
08	General situation of the economy regardless of situation in your sector and enterprise	improved unchanged deteriorated	will improve will remain unchanged will deteriorate

Source: the RIED database

Four survey questions have been selected for empirical analysis, namely, those pertaining to changes in production (question number 01), prices (05), employment (06) and general business conditions (08). These variables have been analyzed previously by various statistic and econometric tools (see E. Tomczyk, 2005) as they can be easily compared with official government statistics for the purpose of quantitative analysis. *A priori* structure is defined as percentages of respondents who expect increase / no change / decline, and *a posteriori* structure as percentages of respondents who observe increase / no change / decline three and four months later. This definition fulfills conditions (1) and (2) and therefore can be employed to evaluate information content and dissimilarity between expectations and observed realizations declared by Polish industrial enterprises in business tendency surveys. Since forecast horizon is not precisely defined in RIED surveys, both alternatives (k = 3 and k = 4) are taken into consideration when calculating measure of dissimilarity $P(S_q^n : S_p^n)$.

Basic sample available for analysis includes 184 observations from March 1997 to June 2012. However, since expectations have to be matched with observed realizations to calculate the measure of dissimilarity $P(S_q^n : S_p^n)$, length of time series is reduced either by three (for 3-month forecast horizon) or by four observations (for 4-month forecast horizon). Additionally, general business conditions data starts one month later than the rest of the survey (that is, in April 1997). For clarity of presentation, all results are reported for the core time period of August 1997 to February 2012 (175 observations). All four variables (production, prices, employment and general business conditions) are analyzed across two ownership types (public and private, as well as aggregated in "all enterprises" class) and four size categories. Even though RIED survey classifies respondents in five size classes:

I. up to 50 employees,
II. 51 to 250 employees,
III. 251 to 500 employees,
IV. 501 to 2000 employees,
V. over 2001 employees,

category V contains a very small number of very large firms, and in many cases this number is insufficient to calculate appropriate entropy and dissimilarity measures. Table 2 presents percentages of nonempty "size V" cells in calculations based on 175 data points in sample covering period from August 1997 to February 2012.

Table 2. Number of observations available for category V (firms with over 2001 employees)

	expectations	**realizations**
production	82%	80%
prices	61%	66%
employment	29%	33%
general business conditions	62%	48%

Source: own calculations on the basis of RIED data

Table 2 makes it clear that only 29–82 % of observations would be available if category V is included in the empirical examination. Due to difficulties with comparisons across variables and sectors, category V has been omitted from further analysis. For the remaining categories, if there are single values missing (for example, 171 or 173 observations available out of the sample total of 175), entropy and dissimilarity measures are calculated on the basis of nonempty cells.

4. Empirical results for measures of entropy and dissimilarity of structures

Table 3 presents summary statistics for entropy measure $H(S^n)$ defined by formula (3), calculated for all four variables, separately for expectations and observed changes, across ownership sectors.

Table 3. Summary statistics for entropy measures

	production					
	expectations			**realizations**		
	all	**public**	**private**	**all**	**public**	**private**
min	1.2650	1.1956	1.2851	1.3289	1.2638	1.2790
max	1.5515	1.5775	1.5679	1.5702	1.5803	1.5751
avg	1.4680	1.4526	1.4720	1.5044	1.4926	1.5070
median	1.4720	1.4670	1.4741	1.5089	1.5093	1.5160
std dev	0.0504	0.0755	0.0559	0.0429	0.0594	0.0474
	prices					
	expectations			**realizations**		
	all	**public**	**private**	**all**	**public**	**private**
min	0.7713	0.6295	0.7566	0.7674	0.5700	0.8155
max	1.3140	1.3768	1.2962	1.2962	1.4447	1.2882
avg	0.9992	0.9915	1.0022	1.0343	1.0276	1.0339
median	0.9980	0.9772	1.0008	1.0314	1.0100	1.0356
std dev	0.1045	0.1417	0.1090	0.1037	0.1674	0.1033

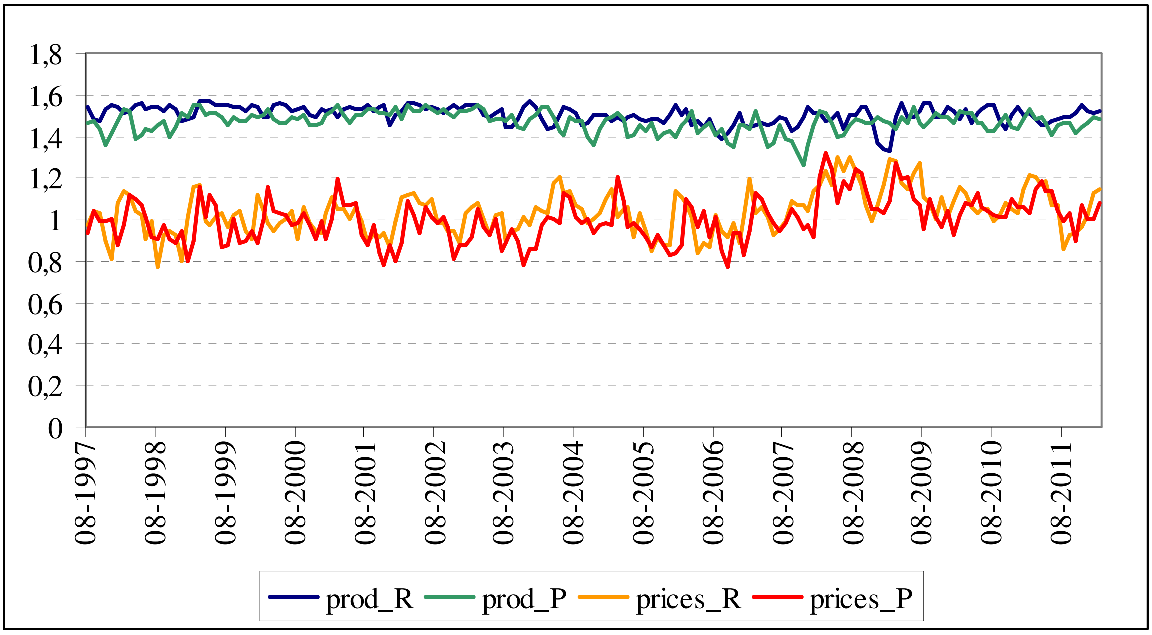
	employment					
	expectations			realizations		
	all	public	private	all	public	private
min	0.9398	0.8669	0.8676	1.0486	0.8241	1.0404
max	1.3266	1.4246	1.3859	1.3940	1.5469	1.3901
avg	1.1785	1.1687	1.1787	1.2417	1.2240	1.2523
median	1.1872	1.1723	1.1805	1.2468	1.2321	1.2509
std dev	0.0736	0.0945	0.0888	0.0674	0.1174	0.0712
	general business conditions					
	expectations			**realizations**		
	all	**public**	**private**	**all**	**public**	**private**
min	0.8663	0.8540	0.8778	0.6238	0.7749	0.5354
max	1.4407	1.4902	1.4796	1.3857	1.4478	1.4308
avg	1.2935	1.2537	1.3148	1.1790	1.1387	1.2035
median	1.3018	1.2660	1.3299	1.1945	1.1522	1.2267
std dev	0.0807	0.1136	0.0857	0.1073	0.1153	0.1226

Source: own calculations on the basis of RIED data

In line with results obtained on the basis of a sample shorter by 20 months (see E. Tomczyk, 2011), very high values of entropy are observed in case of production. For a structure composed of three categories (in this case, expected or observed decrease / no change / increase), the maximum value of measure of entropy is $H_{max} = \log_2 n = \log_2 3 = 1.5850$. Maximum values obtained for production in public sectors, where numbers are particularly high (1.5775 for expectations and 1.5803 for realizations) are close to the upper limit, and average values (1.4680 for expectations and 1.5044 for realizations) are also very high in comparison to entropy of prices, employment and general business conditions. The closer empirical entropy of a structure to its maximum value, the more uniform the structure is, and therefore the less informative *a priori* structure becomes in relation to *a posteriori* structure. In case of production, therefore, *a priori* structures of both expectations and observed changes in production are not informative in light of *a posteriori* distributions observed 3–4 months later.

On the other hand, values of entropy decrease as one of the elements of a structure approaches 1 and uncertainty associated with distribution of outcomes decreases. The lowest entropy values are observed for prices; that is, *a priori* structures of both expectations and observed changes in prices are relatively informative in light of *a posteriori* distributions observed 3–4 months later. Values of entropy for both border cases – production and prices – are presented in Figure 1. Apart from the clear difference between (high) production and (low) price entropy values, relatively high variability of price entropy and low variability of production entropy is also noticeable, and confirmed by descriptive statistics presented in Table 3.

Figure 1. Comparison of entropy values for production expectations (prod_P), production realizations (prod_R) and prices price expectations (prices_P) and prices realizations (prices_R)



Source: own calculations on the basis of RIED data

Generally, empirical results presented in Table 3 and Figure 1 confirm previous findings obtained on the basis of a shorter sample. For three variables (production, prices and employment), expectations exhibit lower averages and medians, across all ownership sectors, than realizations; in case of general business conditions, the opposite is true. Comparison of public and private sector enterprises leads to the conclusion that public firms generally exhibit lower average entropy and higher variation than private firms. This pattern is broken only in case of variability of observed general business conditions which is slightly higher for private enterprises.

After adding enterprise size effects into consideration, results do not change. When disaggregated into size categories, entropy of production remains the highest and least variable and entropy of prices – the lowest and most variable. Therefore size of a firm does not seem to influence the degree to which either expectations or observed changes are relatively informative in light of *a posteriori* distributions observed 3–4 months later.

Though perhaps disappointing on the empirical level, lack of differentiation of entropy measures with enterprise size should be perhaps seen as favorable result since statistics published by Polish Central Statistical Office are based on a different set of size categories:

- up to 9 employees,
- 10 – 49 employees,
- 50 – 249 employees,
- 250 – 999 employees,
- over 1000 employees.

Two size classifications – RIED's and CSO's – are not directly comparable, and if entropy measures were found to depend on enterprise size, further empirical analysis of size effects would be hindered by incompatibility of the two classifications.

Next, empirical values and dynamics of dissimilarity measure $P(S_q^n : S_p^n)$ are employed to evaluate similarities between expectations and realizations expressed in business tendency surveys as they reflect structure change from its *a priori* to *a posteriori* state. Following E. Tomczyk (2011), differences between $P(S_q^n : S_p^n)$ are calculated for k = 3 and for k = 4. Negative values signify that structures of realizations and expectations for the three-month forecast horizon (k = 3) are more similar than structures of realizations and expectations for the four-month forecast horizon (k = 4). Minimum, maximum, average and median values of differences in dissimilarity statistics, along with standard deviation as measure of variation, are presented in Table 4.

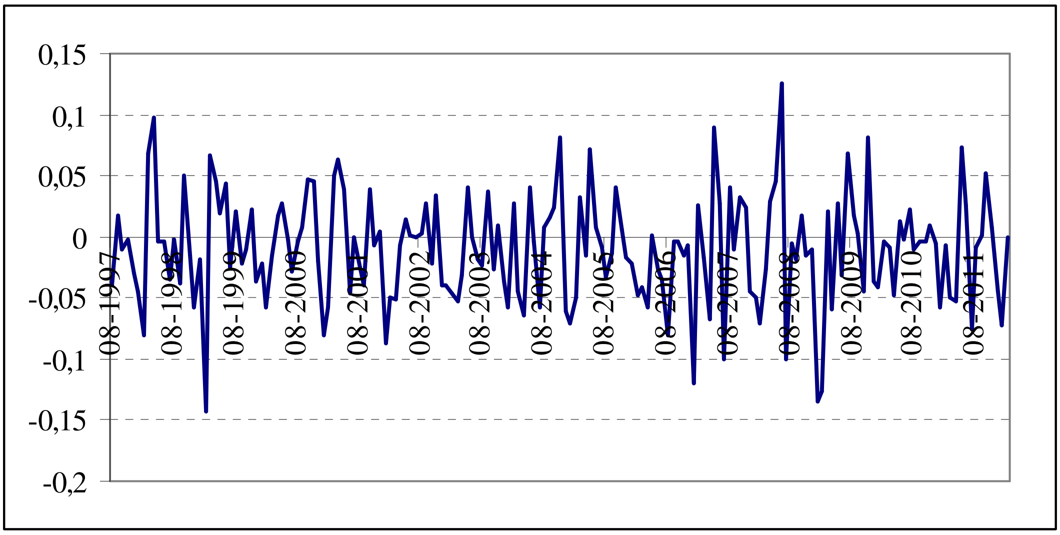
Table 4. Summary statistics for dissimilarity measures: difference between $k = 3$ and $k = 4$, all enterprises

	production	prices	employment	business
min	-0.1420	-0.1160	-0.1080	-0.1880
max	0.1250	0.1920	0.0700	0.1010
avg	-0.0102	-0.0065	-0.0048	-0.0085
median	-0.0080	-0.0050	-0.0020	-0.0060
std dev	0.0446	0.0389	0.0267	0.0454

Source: own calculations on the basis of RIED data

Again, results obtained earlier for a shorter sample (see E. Tomczyk, 2011) are not disproved by updated calculations presented in Table 4. Average and median values of dissimilarity statistics are, for all variables considered, slightly negative, which suggests that 4-month expectations are less compatible with observed changes than expectations interpreted in 3-month horizon. Still, differences between dissimilarity measures for 3-month and 4-month horizons remain very small and should not be considered as a proof of superiority of the shorter forecast horizon, particularly since standard errors are generally several times higher than averages or medians. To illustrate the issue, differences between dissimilarity measures of production for $k = 3$ and $k = 4$ are presented on Figure 2.

Figure 2. Difference between dissimilarity measures calculated for two forecast horizons ($k = 3$ and $k = 4$), question number 01 (production)



Source: own calculations on the basis of RIED data

Graph of differences between dissimilarity measures calculated for production, presented on Figure 2, is quantitatively similar to graphs corresponding to the other three variables, namely, it does not offer a clear answer as to which forecast horizon is supported by data. What is more, graphical analysis does not provide any information as to possible division of the sample into subsamples for which a pattern of preferred forecast horizon can be discerned.

In Tables 5 and 6, summary statistics for dissimilarity measures are reported separately for public and private enterprises.

Table 5. Summary statistics for dissimilarity measures: difference between $k = 3$ and $k = 4$, public enterprises

	production	**prices**	**employment**	**business**
min	-0.2610	-0.2100	-0.1340	-0.3040
max	0.1710	0.2020	0.1660	0.1570
avg	-0.0107	-0.0093	-0.0022	-0.0085
median	-0.0110	-0.0100	-0.0040	-0.0060
std dev	0.0602	0.0572	0.0473	0.0575

Source: own calculations on the basis of RIED data

Table 6. Summary statistics for dissimilarity measures: difference between $k = 3$ and $k = 4$, private enterprises

	production	**prices**	**employment**	**business**
min	-0.1600	-0.2020	-0.1270	-0.1650
max	0.1510	0.1730	0.0790	0.1500
avg	-0.0112	-0.0071	-0.0051	-0.0089
median	-0.0070	-0.0070	0.0000	-0.0090
std dev	0.0465	0.0398	0.0313	0.0509

Source: own calculations on the basis of RIED data

Sector-specific results are qualitatively similar to aggregated outcomes presented in Table 4. Average and median values of differenced dissimilarity measures remain negative, and no clear pattern can be discerned as to mean and variability measured by standard deviation. This result should be perhaps interpreted as favorable: most statistics available in Poland are not published separately for public and private sector, and finding, on the basis of Tables 5 and 6, that public and private enterprises do not differ in approach to expectations horizon, allows to analyze them on the aggregated level without the loss of important information on forecast horizon patterns.

5. Firm size effects in measures of entropy

Tables 7 – 10 provide summary statistics for measures of entropy across four analyzed size categories (see description of RIED data in section 3).

Table 7. Summary statistics for entropy measures of production across size categories

	expectations				realizations			
class	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
min	1.2571	1.1944	1.1442	1.0724	1.2691	1.3567	0.9656	1.1137
max	1.5840	1.5634	1.5654	1.5835	1.5811	1.5755	1.5823	1.5753
avg	1.4874	1.4645	1.4315	1.4132	1.5035	1.5011	1.4596	1.4638
median	1.4965	1.4766	1.4437	1.4219	1.5144	1.5082	1.4851	1.4841
std dev	0.0582	0.0684	0.0825	0.0991	0.0539	0.0442	0.0997	0.0954

Source: own calculations on the basis of RIED data

Table 8. Summary statistics for entropy measures of prices across size categories

	expectations				realizations			
class	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
min	0.5810	0.6662	0.4889	0.3871	0.7240	0.7192	0.5230	0.4626
max	1.2759	1.3140	1.3208	1.3794	1.3007	1.3288	1.4189	1.3830
avg	0.9773	0.9966	0.9631	0.9597	1.0081	1.0215	0.9751	1.0262
median	0.9722	1.0078	0.9670	0.9593	1.0151	1.0254	0.9638	1.0154
std dev	0.1455	0.1266	0.1430	0.1647	0.1193	0.1227	0.1643	0.1643

Source: own calculations on the basis of RIED data

Table 9. Summary statistics for entropy measures of employment across size categories

	expectations				realizations			
class	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
min	0.7089	0.9159	0.7598	0.6912	0.8223	1.0069	0.5934	0.6558
max	1.3881	1.3565	1.4455	1.5163	1.4203	1.4334	1.4964	1.5821
avg	1.1027	1.1546	1.1733	1.1669	1.1731	1.2447	1.2182	1.2318
median	1.1106	1.1576	1.1841	1.1650	1.1829	1.2503	1.2138	1.2399
std dev	0.1179	0.0797	0.1352	0.1424	0.1063	0.0762	0.1319	0.1280

Source: own calculations on the basis of RIED data

Table 10. Summary statistics for entropy measures of general business conditions across size categories

	expectations				realizations			
class	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
min	1.0602	0.7156	0.8801	0.7055	0.7113	0.6725	0.5645	0.5492
max	1.5083	1.4961	1.5191	1.5127	1.4620	1.4191	1.5548	1.4008
avg	1.3368	1.3054	1.2685	1.1913	1.2068	1.2037	1.1737	1.0713
median	1.3546	1.3179	1.2846	1.2212	1.2182	1.2281	1.1803	1.0890
std dev	0.0893	0.1023	0.1173	0.1577	0.1257	0.1318	0.1418	0.1410

Source: own calculations on the basis of RIED data

Values of entropy for production and general business conditions variables, presented in Tables 7 and 10, exhibit a certain pattern: average and median values decrease and standard deviations increase with size of an enterprise. This effect is clear for expectations and less so for realized changes. There is no pattern discernible for entropy values calculated for prices and employment variables (Tables 8 and 9).

6. Concluding comments

Empirical analysis of business tendency survey data by the means of entropy and dissimilarity measures leads to the following conclusions:

1. Enterprise size does not seem to influence the degree to which either expectations or observed changes are relatively informative in light of *a posteriori* distributions observed 3–4 months later.

2. Results obtained separately for public and private firms are parallel to aggregated outcomes; both types of enterprises are characterized by similar interpretation of expectations horizon and therefore can be studied on the aggregated level without the loss of important information on forecast horizon patterns.
3. Certain firm size effects (averages decreasing and variation increasing with size) are observed for production and general business conditions variables; for prices and employment variables, there is no apparent pattern in respondents' answers.

On the basis of longer sample, the following results obtained in E. Tomczyk (2011) are also confirmed:

1. Entropy of production expectations and observed changes is generally high, that is, distribution of respondent answers is almost uniform; on the other hand, entropy of prices expectations and observed changes is relatively low. Fractions of survey answers to price question are centered on one of the three options (in practice, "no change" category).
2. Public enterprises tend to exhibit lower average entropy and higher variability of entropy than private enterprises.
3. Dissimilarity measures do not provide unequivocal answer as to which forecast horizon (three- or four-month) is supported by information content of survey data.

Results obtained so far on the basis of entropy and dissimilarity measures provide new insights into behavior of expectations and realizations expressed in business tendency surveys and offer a promising field for future research. Two issues are considered to be the next steps in the empirical analysis of RIED survey data: whether current financial standing of an enterprise (see Table 1, question number 07) influences degree of concentration of answers on a particular option; and whether entropy and dissimilarity measures are sensitive to sector in which a firm operates, as classified by the Statistical Classification of Economic Activities.

7. References

Borgesi S. (2008) *Global Sustainability: Social and Environmental Conditions*, Palgrave Macmillan, Basingstoke.

Chomątowski S., Sokołowski A. (1978) *Taksonomia struktur* [*Taxonomy of structures*], Przegląd Statystyczny 2:217–226.

ECB (2004) *Sectoral Specialisation in the EU. A Macroeconomic Perspective*, Occasional Paper Series No 19 (http://www.ecb.int).

Foster J. E., Sen A. (1997) *On Economic Inequality*, Oxford University Press.

Frenken K., Nuvolari A. (2004) *Entropy statistics as a framework to analyse technological evolution,* in: Foster J., Hölzl W. (ed.) *Applied Evolutionary Economics and Complex Systems,* Edward Elgar, Cheltenham.

Gabor D., Gabor A. (1979) *The application of entropy to the measurement of social and economic freedom*, International Journal of Social Economics 6:415–428.

Georgescu-Roegen N. (1971) *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge.

Jaynes E. T. (1991) *How should we use entropy in economics?*, St. John's College, Cambridge (http://bayes.wustl.edu/etj/articles/entropy.in.economics.pdf).

Kowalczyk H. (2010) *O eksperckich ocenach niepewności w ankietach makroekonomicznych* [*On expert assessment of uncertainty in macroeconomic surveys*], Bank i Kredyt 5:101–122.

Marcinkowska I., Ruzik A., Strawiński P., Walewski M. (2008) *Badanie struktury i zmian rozkładu wynagrodzeń w Polsce w latach 2000–2006* [*Analysis of structure and dynamics of wage distribution in Poland in 2000–2006*], Departament Analiz Ekonomicznych i Prognoz, Ministerstwo Pracy i Polityki Społecznej, Warszawa.

Nehmer R. A. (2011) *Information entropy, economic analysis and financial reporting*, Review of Business Information Systems 5:73–79.

Raine A., Foster J., Potts J. (2006) *The new entropy law and the economic process*, Ecological Complexity 3:354–360.

Rényi A. (1961) *On measures of entropy and information,* Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1:547–561.

RIED (2012) *Business Survey No 281. February 2012*, Warsaw School of Economics, Warszawa.

Shannon C. E. (1948) *A mathematical theory of communication*, The Bell System Technical Journal 27:379–423, 623–656.

Singh V. A., Pathak P., Pandey P. (2010) *Monitoring the teaching-learning process via an entropy based index*, in: B. Basu, B. K. Chakrabarti, S. R. Chakravarty, K. Gangopadhyay (eds.) *Econophysics and Economics of Games, Social Choices, and Quantitative Techniques*, Springer, Dordrecht.

Stattev S., Raleva S. (2006) *Bulgarian GDP structures – convergence with the EU*, South-Eastern Europe Journal of Economics 2:193–207.

Theil H. (1967) *Economics and Information Theory*, North-Holland Publishing Company, Amsterdam.

Theil H., Chen D. (1996) *The informational fit and fitting of cross-country demand systems*, in: H. Theil (ed.) *Studies in Global Econometrics*, Kluwer Academic Publishers, Dordrecht.

Tomczyk E. (2005) *Are expectations of Polish industrial enterprises rational? Evidence from business tendency surveys*, in: Adamowicz E., Klimkowska J. (eds.) *Economic Tendency Surveys and Cyclical Indicators. Polish contribution to the 27th CIRET Conference,* Warsaw School of Economics, Warszawa.

Tomczyk E. (2011) *Application of measures of entropy, information content and dissimilarity of structures to business tendency survey data,* Przegląd Statystyczny 58:88–101.

WCED (1987) *Report of the World Commission on Environment and Development: Our Common Future*, NGO Committee on Education (http://www.un-documents.net/wced-ocf.htm).

Wędrowska E. (2009) *Oczekiwana ilość informacji o zmianie struktur jako miara niepodobieństwa struktur* [*Expected information content of structure change as a measure of dissimilarity of structures*], paper presented at the XIth Conference „Dynamic Econometric Models", September 2009, Toruń.