**Application of entropy measures in the analysis of survey results
An example of a survey of inflation expectations in Poland**

dr Jarosław Janecki

Scientific conference
MET2023 Metodology of Statistical Research
July 3-5, 2023, Warsaw

Objective

- Extending the possibility of analyzing survey results

- The use of entropy measures in the context of survey analytics

- Evaluation of the survey results from the point of view of the uncertainty of the respondents

Uncertainty proxies

- Indices based on the volatility of valuations of financial instruments
- Indices based on the statistics of words appearing in the media (text based indexes)
- Indices based on forecast errors
- Indices based on forecast disagreement

SGH

Methodology: empirical measure of entropy

- Following Shannon's (1948) proposal, we can define an empirical measure of entropy that will act as a proxy for inflation uncertainty for our research.

- Let S^n denote the structure index or index of the sequence of n components from α_1 to α_{1n}. According to the definition proposed by Shannon, for the structure S^n the real entropy $H(S^n)$ can be written as follows:

$$(S^n) = -\sum_{i=1}^{n} \alpha_i \log_2 \alpha_i = \sum_{i=1}^{n} \alpha_i \log_2 \frac{1}{\alpha_i}$$

where:
$H(S^n)$ - entropy value for the structure S^n,
α_i - structure index, where $0 \leq \alpha_i \leq 1$, $i = 1, \dots, n$ and $\sum_{i=1}^{n} \alpha_i = 1$

SGH

Methodology: empirical measure of entropy

- Entropy is value is a measure of the concentration of elements α_i of the S^n structure.

- The maximum entropy value $H(S^n)$ is obtained when all elements of the structure are equal. In this case we have the highest uncertainty.

- The lower the entropy value, the lower the uncertainty.

SGH

Methodology: *a priori* and *a posteriori* structure

- In order to measure changes between assumed *a priori* structure (the structure of the size of predicted changes) and observed *a posteriori* structure (the structure of the size of observed changes), the expected amount of information is calculated as follows (Theil, 1967):

$$I\left(S_q^n : S_p^n\right) = \sum_{i=1}^{n} q_i \log_2 \frac{q_i}{p_i}$$

where:

$I\left(S_q^n : S_p^n\right)$ - expected amount of information,

S_p^n - assumed *a priori* structure,

S_q^n - observed *a posteriori* structure,

q_i - *a posteriori* structure indicators, $i = 1, \dots, n$,

p_i - *a priori* structure indicators, $i = 1, \dots, n$

Methodology: degree of similarity

- Calculating the value $I\left(S_q^n : S_p^n\right)$ we receive information about the degree of similarity (or dissimilarity) between the structure *a priori* and *a posteriori*, structures S_p^n i S_q^n.

- The measure $I\left(S_q^n : S_p^n\right)$ has a minimum value of zero when both structures are identical. With increasing differences between structures, it can grow to infinity.

- The value of $I\left(S_q^n : S_p^n\right)$ is a measure of the change between the predicted (*a priori*) and the observed (*a posteriori*) structure, the greater its value, the less similar the two structures are.

SGH

Methodology: VAR model

- In the next step we will check what role entropy plays in the inflation forecast. For this purpose, the VAR model will be used to estimates how variables are related to lagged values of other variables over time.
- The basic VAR model is presented in the form of the following equation:

$$X_t = B_0 + B_1 X_{t-1} + \ldots + B_m X_{t-m} + U_t$$

where:

X_t - a list of macroeconomic variables,

U_t – error terms, independent of the X_t,

$B_t, \ldots, B_m$ – set of coefficients

- For the VAR model, we chose an approach in which the following variables were taken into account:

  – inflation,

  – inflation expectations,

  – entropy of inflation expectations.

SGH

Data

- Data covers the period from May 2001 to January 2023. The study used the results of surveys on inflation expectations for Poland, conducted for the European Commission, these are questions 5 and 6:

Question 5. In your opinion, how have consumer prices (consumer goods and services) changed in the last 12 months?
A.   increased significantly (weight: 1)
B.   increased moderately (weight: 0,5)
C.   increased slightly (weight: 0)
D.   remained more or less the same (weight: -0,5)
E.   Dropped (weight: -1,0)
F.   I don't know (weight: 0)

Question 6. Compared to the last 12 months, what changes do you anticipate in consumer prices (consumer goods and services) over the next 12 months?
A.   there will be faster growth (weight: 1)
B.   it will grow at a similar rate (weight: 0,5)
C.   will grow slower (weight: 0)
D.   will remain more or less at the same level (weight: -0,5)
E.   there will be a decline (weight: -1,0)
F.   I don't know (weight: 0)

SGH

Data

- An example of an *a posteriori* structure may be the answers to the question on consumer assessment in relation to consumer prices in the last 12 months (question 5).

- An example of an *a priori* structure may be the answers obtained as part of the answer to the question about consumer prices expected by consumers in the next 12 months (question 6).

- Total *a posteriori* and *a priori* inflation assessments for euro area countries were taken into account, broken down by:

  – Education: ED1 - primary, ED2 – secondary, ED3 - further education;
  – Age of respondents: AG1 – 16-29, AG2 - 30-49 , AG3 - 50-64, AG4 - +65;
  – Gender: M - male, F – female.

SGH

Table 1. *A posteriori* entropy values.

	H_TOT.5	H_ED1.5	H_ED2.5	H_ED3.5	H_AG1.5	H_AG2.5	H_AG3.5	H_AG4.5	H_M.5	H_F.5
May 2001 - Jan.2023										
min	1,6649	1,5010	1,6165	1,3915	1,5714	1,5687	1,3837	1,2822	1,6500	1,4797
max	2,1515	2,2537	2,1478	2,2107	2,2078	2,1411	2,2481	2,2667	2,1632	2,1403
avg.	1,9062	1,9571	1,8832	1,8168	1,9435	1,8516	1,8483	1,8814	1,9245	1,8738
median	1,9183	1,9657	1,8902	1,8296	1,9474	1,8425	1,8417	1,8903	1,9299	1,8800
std dev.	0,1160	0,1440	0,1119	0,1538	0,1167	0,1310	0,1578	0,1850	0,1203	0,1285
March. 2020 - May.2022										
min	1,6753	1,7227	1,6165	1,4320	1,5825	1,6184	1,4501	1,3207	1,7531	1,5548
max	2,0892	2,2294	2,0841	2,0005	2,2078	2,1411	2,0891	2,0785	2,1251	2,0398
avg.	1,9246	2,0337	1,9134	1,8025	1,9868	1,9478	1,8245	1,7380	1,9783	1,8478
median	1,9648	2,0512	1,9601	1,8330	2,0006	1,9965	1,8376	1,7694	1,9888	1,9023
std dev.	0,1219	0,1367	0,1305	0,1602	0,1253	0,1512	0,1724	0,2177	0,1122	0,1486

Notes:

Education: ED1 - primary, ED2 – secondary, ED3 - further education;

Age of respondents: AG1 – 16-29, AG2 - 30-49 , AG3 - 50-64, AG4 - +65;

Gender: M - male, F – female.

SGH

Table 2. *A priori* entropy values.

	H_TOT.6	H_ED1.6	H_ED2.6	H_ED3.6	H_AG1.6	H_AG2.6	H_AG3.6	H_AG4.6	H_M.6	H_F.6
May 2001 - Jan.2023										
min	1,6228	1,4075	1,4873	1,2183	1,5904	1,4873	1,4963	1,4322	1,5415	1,5470
max	2,2866	2,3450	2,2481	2,3308	2,3267	2,2481	2,3426	2,3476	2,2862	2,2769
avg.	1,9754	2,0034	1,9231	1,8836	1,9864	1,9180	1,9282	1,9826	1,9800	1,9505
median	1,9710	2,0203	1,9191	1,8873	1,9979	1,9119	1,9141	2,0068	1,9760	1,9483
std dev.	0,1316	0,1452	0,1480	0,1902	0,1393	0,1491	0,1685	0,1599	0,1394	0,1377
March. 2020 - May.2022										
min	1,7096	1,4075	1,6904	1,5490	1,6627	1,6904	1,6327	1,4322	1,7675	1,5470
max	2,1096	2,2693	2,1115	2,1506	2,2849	2,1115	2,2211	2,2231	2,1416	2,0820
avg.	1,9153	1,9388	1,9004	1,7954	1,9159	1,8947	1,8638	1,8253	1,9467	1,8463
median	1,9053	1,9576	1,9119	1,7965	1,8602	1,9059	1,8675	1,8246	1,9343	1,8516
std dev.	0,1034	0,1994	0,1067	0,1265	0,1826	0,1066	0,1315	0,2111	0,1046	0,1216

Notes:

Education: ED1 - primary, ED2 – secondary, ED3 - further education;

Age of respondents: AG1 – 16-29, AG2 - 30-49 , AG3 - 50-64, AG4 - +65;

Gender: M - male, F – female.

SGH

Table 3. Measures of similarity between *a priori* and *a posteriori* structures.

|	I_EA.ED1	I_EA.ED2	I_EA.ED3	I_EA.AG1	I_EA.AG2	I_EA.AG3	I_EA.AG4	I_EA.M	I_EA.F
May 2001 - Jan.2023									
min	0,0074	0,0087	0,0155	0,0232	0,0172	0,0124	-0,0063	0,0241	0,0243
max	1,4147	0,9307	1,5684	1,0118	0,9364	1,1103	0,9885	0,8995	1,0782
avg.	0,2079	0,1957	0,2621	0,1923	0,2118	0,2219	0,2471	0,1862	0,2063
median	0,1660	0,1525	0,2048	0,1523	0,1589	0,1708	0,2071	0,1422	0,1638
std dev.	0,1612	0,1505	0,2140	0,1552	0,1626	0,1705	0,1704	0,1415	0,1542
March. 2020 - May.2022									
min	0,0074	0,0390	0,0435	0,0232	0,0507	0,0557	-0,0063	0,0306	0,0243
max	0,3024	0,3658	0,5283	0,3741	0,3563	0,4487	0,4509	0,3776	0,3769
avg.	0,1243	0,1298	0,1503	0,1294	0,1484	0,1416	0,1718	0,1209	0,1256
median	0,0966	0,1088	0,1052	0,1047	0,1385	0,1156	0,1445	0,0892	0,1135
std dev.	0,0758	0,0677	0,1133	0,0835	0,0751	0,0975	0,1127	0,0778	0,0770

Notes:

Education: ED1 - primary, ED2 – secondary, ED3 - further education;

Age of respondents: AG1 – 16-29, AG2 - 30-49 , AG3 - 50-64, AG4 - +65;

Gender: M - male, F – female.

SGH

Response analysis (*a priori* entropy)

Response to Cholesky One S.D. (d.f. adjusted) Innovations ± 2 S.E.

SGH
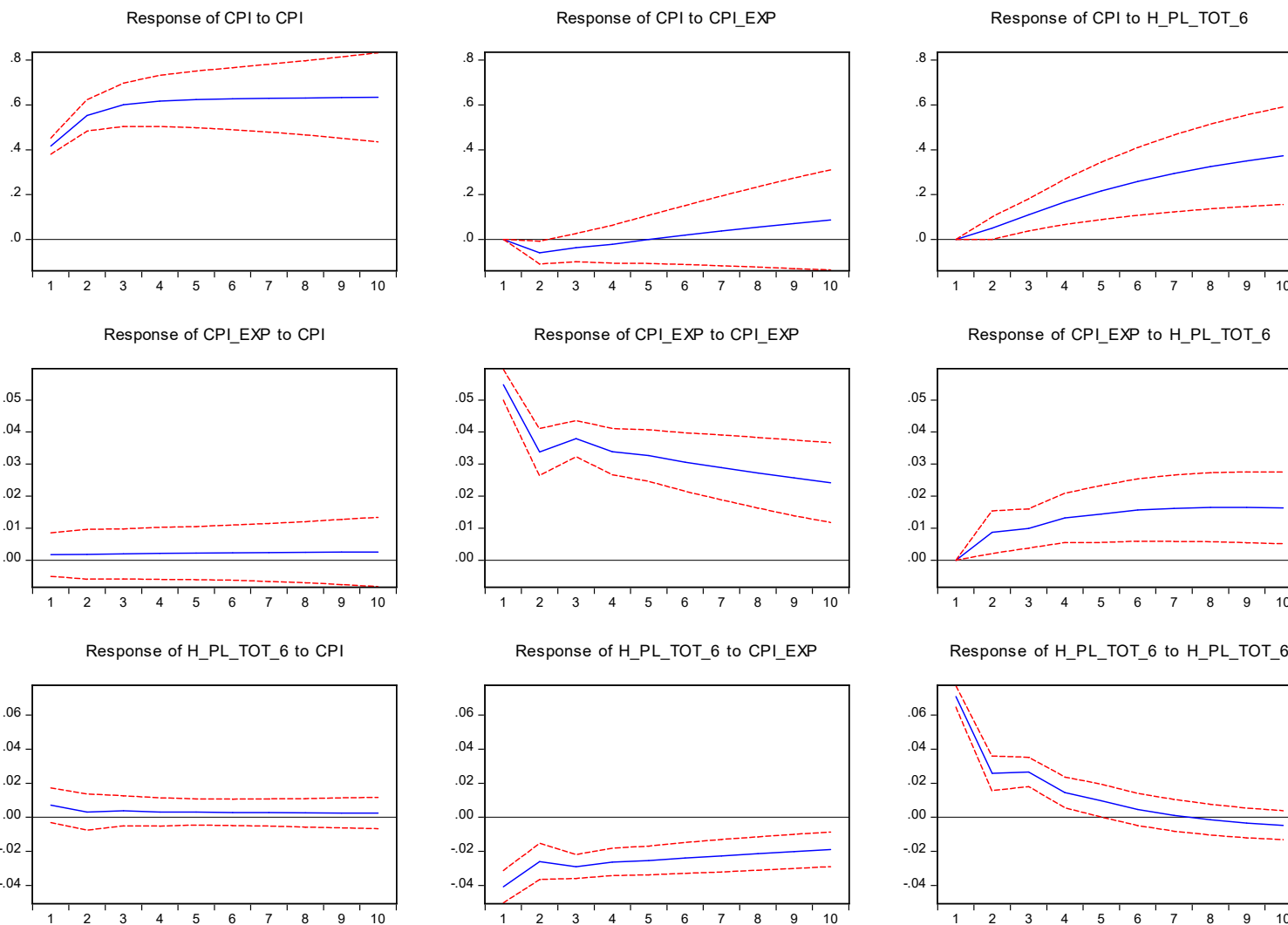
Conclusions

- The greatest uncertainty, occurs among people with the lowest education, the youngest and among men.

- The assessment of the degree of similarity between the *a priori* and *a posteriori* structure indicates a lower similarity in the group of the oldest and most educated people and among women.

- VAR models, in which entropy measures were taken into account, generated smaller errors than in the case of taking into account the standard deviation.

- Entropy measures allow to determine the uncertainty appearing among people answering survey questions. The results indicate the varied uncertainty of the respondents in relation to the answers to questions about the evolution of consumer prices.

SGH

SGH
Szkoła Główna
Handlowa
w Warszawie

Thank you