GOSPODARKA NARODOWA

4(316)2023, 17–29
DOI: 10.33119/GN/174277
gnpje.sgh.waw.pl
SGH

The Polish Journal of Economics

Istnieje od / Published since 1931

Emilia Tomczyk
SGH Warsaw School of Economics, Poland
Barbara Kowalczyk
SGH Warsaw School of Economics, Poland

Consensus in Business Tendency Surveys: Comparison of Alternative Measures

Porównanie metod pomiaru konsensusu w testach koniunktury

Abstract

In this article, we aim to compare various methods of evaluating consensus in qualitative business surveys in which respondents express expectations on the ordered scale. A reliable method of measuring degree of consensus would provide researchers with valuable information, offering a leading indicator of respondent sentiment. However, there is no single generally accepted mathematical measure applicable to evaluating agreement among respondents. Several approaches are mentioned in previous studies, including indicators based on statistical dispersion, Shannon entropy, and multi-dimensional simplex. We present measures of consensus defined in literature and discuss their advantages and limitations. We then employ these indicators to expectations expressed in Polish business tendency survey in manufacturing, and compare results across various economic variables. In several cases, we find patterns in the behavior of measures of consensus: expected prices are characterized by the highest degree of consensus among respondents, and expected production and orders – by the lowest degree of consensus. We also find linkages between the degree of consensus and degree of optimism among respondents as measured by the balance statistic for prices, employment, and general business conditions.

Keywords:
manufacturing, consensus, expectations, qualitative data, business tendency surveys

JEL classification codes:
C18, C83, D84, L60

Article history:
submitted: August 1, 2023
revised: September 28, 2023
accepted: October 18, 2023

Streszczenie

W niniejszym artykule porównujemy różne metody oceny konsensusu w testach koniunktury, w których respondenci wyrażają oczekiwania na skali uporządkowanej. Wiarygodna metoda pomiaru siły konsensusu w oczekiwaniach respondentów dostarczyłaby ekonomistom cennych informacji, stanowiąc wiodący wskaźnik nastrojów podmiotów gospodarczych. Nie ma jednak jednej ogólnie przyjętej miary matematycznej służącej do oceny zgodności między wyrażanymi przez respondentów opiniami. W literaturze wymienianych jest kilka miar, w tym wskaźniki oparte na miarach dyspersji, entropii i wielowymiarowym simpleksie. W artykule przedstawiamy zdefiniowane w literaturze miary konsensusu oraz omawiamy ich zalety i ograniczenia. Następnie wykorzystujemy te wskaźniki do analizy oczekiwań wyrażonych w teście koniunktury w przetwórstwie przemysłowym w Polsce i porównujemy wyniki dla różnych zmiennych ekonomicznych. W kilku przypadkach znajdujemy powtarzalne schematy w zachowaniu miar konsensusu: oczekiwania cenowe charakteryzują się najwyższym stopniem konsensusu, a oczekiwania

Słowa kluczowe:
przetwórstwo przemysłowe, konsensus, oczekiwania, dane jakościowe, testy koniunktury

Kody klasyfikacji JEL:
C18, C83, D84, L60

Historia artykułu:
nadesłany: 1 sierpnia 2023 r.
poprawiony: 28 września 2023 r.
zaakceptowany: 18 października 2023 r.

na temat produkcji i zamówień – najniższym. Wskazujemy również powiązania między stopniem konsensusu a stopniem optymizmu wśród respondentów mierzonym statystykami bilansowymi w przypadku cen, zatrudnienia i ogólnej sytuacji gospodarczej.

Introduction

In this article, we aim to compare various methods for evaluating consensus and disagreement in qualitative surveys in which respondents express expectations on the ordered scale. Consensus among business tendency survey respondents – and its dynamics – is not only informative in itself but can also be employed in macroeconomic analyses as an approximation of the degree of uncertainty. As noted by **Conflitti** [**2011**: 16], "Disagreement among forecasters has been widely used as a proxy for uncertainty, albeit it is difficult to find a unique empirical evidence to support this relation." (We return to the "difficulty" mentioned above in the next section). Generally, a reliable method of measuring degree of consensus would provide economists and decision makers with valuable information, particularly as far as expectations are concerned, because expected consensus could serve as a leading indicator of respondent sentiment. To the best of our knowledge, all available measures of consensus among survey respondents have been neither collated nor empirically compared for business survey data so far. Researchers have discussed the theoretical properties of a newly proposed or revised measure, focused on a single indicator applied to several variables, or compared two or three measures for a limited number of variables. We describe all the consensus measures previously defined in literature for the purpose of evaluating qualitative survey data, compare their theoretical properties, and calculate their values for a wide range of variables originating from a business tendency survey in manufacturing.

In section 2, we define consensus and stress the importance of measures of disagreement in studies of economic phenomena. In section 3, we describe various measures of consensus adequate for qualitative data collected through business tendency surveys, along with their limitations. We then employ several methods of assessing consensus among respondents to a Polish business tendency survey in manufacturing, and discuss their advantages and disadvantages in the context of drawing economic conclusions from survey expectations in section 4. Section 5 concludes.

Disagreement and consensus in business tendency surveys

Expectations concerning key economic variables constitute a major field of interest for applied economists, as expectations undoubtedly influence decisions undertaken by market participants. The role of expectations in studying economic phenomena in general, and decisions of economic agents in particular, has been widely acknowledged by researchers, and resulted in an extensive body of literature on the subject.

While the importance of specifying terms "disagreement" and "consensus" before attempting any analysis of economic expectations seems quite obvious, precise definitions are difficult to pinpoint. The current literature tends to define disagreement and consensus in terms of the variance observed in survey forecasts [**Krüger, Nolte, 2016**]. Perfect consensus is usually understood as an absolute agreement among survey respondents, and is very rare in empirical settings. To account for the typical lack of perfect agreement in economic surveys, in this paper – as in many others – consensus is going to be measured in degrees: the higher the concentration of survey responses, the stronger the consensus.

Most previous analyses of disagreement among survey respondents were based on either point or density forecasts, the latter defined as predictions of distribution of the future values of the variable of interest. Qualitative surveys which require respondents to identify only the direction of forecasted change have rarely been the subject of in-depth analyses of consensus and disagreement. However, in comparison to quantitative or density forecasts – which require substantial economic expertise from the respondents – qualitative surveys

which ask only for the direction of change are generally judged more reliable, particularly when addressed not to professional forecasters but to households or companies. The methods available for measuring consensus in qualitative survey data are discussed in the next section.

Review of measures of consensus in business tendency surveys

In this section, we describe methods used to evaluate the degree of consensus in business tendency surveys based on qualitative assessments and expectations.

A typical qualitative tendency survey provides the respondents with three variants of answers: improvement (increase), no change, and decline (decrease). Some surveys are designed to collect data both on observed changes and on expectations of respondents. In that case, expectations horizon h has to be specified. Since we are particularly interested in expectations as the respondents' assessment of future behavior of economic variables, we focus on predictions expressed in business tendency surveys.

We employ the following notation:

P^{inc} – percentage of respondents expecting increase (improvement) within the prediction horizon specified in the survey,

P^{const} – percentage of respondents expecting no change within the prediction horizon specified in the survey,

P^{dec} – percentage of respondents expecting decrease (deterioration) within the prediction horizon specified in the survey.

As an aggregate measure of respondents' expectations, balance statistic has been traditionally used. It is calculated by subtracting the share of respondents who expect decline ("pessimists") from the share of respondents who expect improvement ("optimists"):

$$BAL = P^{inc} - P^{dec}. \tag{1}$$

High absolute values of balance statistic are interpreted as evidence of the degree of optimism / pessimism among survey respondents: high positive values indicate optimism with respect to the future, and high negative values – pessimism. Balance statistic provides easily accessible information on respondents' sentiment but cannot be interpreted as an indicator of consensus as it is a measure of a central tendency (location) and not of dispersion. Let us note that balance statistic can be regarded as a sample mean, provided we assign +1 to improvement, 0 to no change, and –1 to decrease:

$$\overline{x} = 1 \cdot P^{inc} + 0 \cdot P^{const} - 1 \cdot P^{dec} = P^{inc} - P^{dec} = BAL \tag{2}$$

The simplest measure of dispersion applicable to business tendency surveys in order to evaluate consensus among respondents is the range, that is, difference between the maximum and the minimum shares of respondents:

$$RNG = max\left\{P^{inc}, P^{const}, P^{dec}\right\} - min\left\{P^{inc}, P^{const}, P^{dec}\right\}. \tag{3}$$

However, since by construction the range is always positive, information on the direction of increase / decrease percentages of answers is lost. There is no way to distinguish the "positive" and "negative" kinds of unanimity among respondents, that is, an "optimistic consensus" from a "pessimistic consensus".

Limitations of the range as a measure of consensus among survey responders caused researchers to search for other methods. One of the approaches focused on the entropy, originally introduced in the context of thermodynamics and information theory. The entropy of a structure (in this case, the shares of respondents expecting improvement, no change, or decline) can be defined as

$$H = P^{inc} \cdot \log_2 \frac{1}{P^{inc}} + P^{const} \cdot \log_2 \frac{1}{P^{const}} + P^{dec} \cdot \log_2 \frac{1}{P^{dec}}. \tag{4}$$

Although H is not formally defined if any element of structure $\{P^{inc}, P^{const}, P^{dec}\}$ is equal to 0, there is an implicit assumption that its corresponding component $P\log_2\frac{1}{P}$ is also equal to 0, which follows from the mathematical theorem $\lim_{P\to 0^+} P\log_2\frac{1}{P}=0$.

The unit of measurement for the entropy defined for base-2 logarithm is called a shannon (Sh). The information content of a structure is defined in relation to the probability that a given structure is recorded from the set of all possible structures: the less probable the structure, the more information it carries. Empirical entropy (4) is therefore interpreted in the economic context as a measure of lack of consensus: the closer empirical entropy of a structure to its maximum value (for a three-element structure, $H_{max}=\log_2 n=\log_2 3=1.5850$), the more uniform the structure is, and therefore less concentrated on any particular option, suggesting higher disagreement and lower consensus.

It is worth noting that the (elusive) perfect consensus in tendency surveys would be associated with the structure in which one of the elements is equal to 1, and the other two are equal to 0. The need for measures allowing for various levels of consensus arises from the fact that such an ideal agreement between respondents is hardly ever observed in economic practice.

Recently [**Claveria, 2019**; **Claveria et al., 2019**], a geometrical measure based on a simplex has been introduced. For a business tendency survey with three reply options, the simplex consensus measure is defined as

$$Cns\_SMPL=\frac{\sqrt{(P^{inc}-0.33)^2+(P^{const}-0.33)^2+(P^{dec}-0.33)^2}}{\sqrt{2/3}} \tag{5}$$

This formula measures the degree of agreement as the ratio between the distance of aggregated responses to the barycenter and from the barycenter to the nearest vertex: the higher (closer to 1) the value of *Cns_SMPL*, the higher the concentration of answers in one of the categories. Value of 0 is attained for equal distribution of aggregated responses between the categories.

The main disadvantage of all the measures of consensus presented above is that they do not take the distances between the categories of answers into account. Intuitively, the gap separating respondents expecting improvement and decrease (the extremes of the spectrum) is wider, in an economic sense, than the gap separating respondents expecting improvement and no change, or decrease and no change. Consensus measures should not assign the same values to these, clearly distinct, share structures (by which we mean percentages of responses in each of the three categories: increase, no change, decrease). Generally, none of the methods described above are free from weaknesses as measures of survey consensus. Drawbacks and counterexamples – that is, examples of survey response shares for which these measures do not perform as expected – are summarized in Table 1.

Table 1. Measures used to evaluate degree of agreement in tendency surveys

Measure	Drawbacks and counterexamples
Range *RNG*	• distances between answers are not taken into account • importance of "no change" neglected • information about positive / negative sentiment is lost: (0.10, 0.20, 0.70) → 0.60; also (0.70, 0.20, 0.10) → 0.60 • two different expectations structures may lead to identical ranges: (0.50, 0.35, 0.15) → 0.35; also (0.25, 0.55, 0.20) → 35 • situation in which some consensus is visible: (0.50, 0.50, 0) → 50 is indistinguishable from situation in which there is no consensus at all: (0.50, 0, 0.50) → 50
Entropy *H*	• distances between answers are not taken into account • information about positive / negative sentiment is lost: (0.10, 0.20, 0.70) → 1.16; also (0.70, 0.20, 0.10) → 1.16 • situation in which some consensus is visible: (0.50, 0.50, 0) → 1 is indistinguishable from situation in which there is no consensus at all: (0.50, 0, 0.50) → 1 • entropy for a full consensus (1, 0, 0) is 0, for no consensus at all (0.50, 0, 0.50) is 1; while for (0.20, 0.60, 0.20) is 1.37, that is, consensus for (0.50, 0, 0.50) is higher than for (0.20, 0.60, 0.20), which is counterintuitive

Measure	Drawbacks and counterexamples
Simplex *Cns_SMPL*	• distances between answers are not taken into account • information about positive / negative sentiment is lost: (0.10, 0.20, 0.70) → 0.5568; also (0.70, 0.20, 0.10) → 0.5568 • situation in which some consensus is visible: (0.50, 0.50, 0) → 0.50 is indistinguishable from situation in which there is no consensus at all: (0.50, 0, 0.50) → 0.50 • simplex for no consensus at all: (0.50, 0, 0.50) is 0.50; while for (0.20, 0.60, 0.20) it is 0.40, that is, consensus for (0.50, 0, 0.50) is higher than for (0.20, 60, 20), which is counterintuitive

Source: Authors' own elaboration.

A novel approach to measuring consensus in qualitative survey data has been presented in 2006 but largely overlooked since. **Tastle and Wierman [2006]** introduce a measure (henceforth *Cns_TW*) which offers significant improvement over alternative methods. For a business tendency survey with three reply options +1 (increase), 0 (no change), and –1 (decrease), their consensus measure is defined as

$$Cns\_TW = 1 + P^{inc}\log_2\left(1-\frac{|1-BAL|}{2}\right) + P^{const}\log_2\left(1-\frac{|0-BAL|}{2}\right) + P^{dec}\log_2\left(1-\frac{|-1-BAL|}{2}\right). \tag{6}$$

Therefore *Cns_TW* satisfies the condition:

$$0 \le Cns\_TW \le 1 .$$

The main advantage of this measure is that it takes into account the distances between the answers, i.e. the distance between improvement and no change is two times smaller than the distance between improvement and decline. Therefore, the minimum of consensus is obtained for (0.5, 0, 0.5), and the share structure (0.5, 0.5, 0) is viewed to be the higher consensus than (0.5, 0, 0.5), which was not true either for simplex, entropy, or range.

The final measure of consensus considered in this paper also takes into account the distances between the answers. On the basis of Bachmann *et al*. (2013) we define the variance-based measure as:

$$Var = P^{inc} + P^{dec} - (BAL)^2 = 1 - P^{const} - (BAL)^2. \tag{7}$$

Let us note that formula (7) can be regarded as sample variance for business tendency surveys. Indeed, if three reply options: increase, no change, and decrease are quantified as +1, 0, and –1 respectively, we have:

$$Var = \frac{1}{n}\sum_{i=1}^{3} x_i^2 - \overline{x}^2 = 1\cdot(P^{inc} + P^{dec}) + 0\cdot P^{const} - (BAL)^2. \tag{8}$$

The variance for business tendency surveys has a very convenient feature: it satisfies the condition $0 \le Var \le 1$. The additional advantage of the proposed indicator is that it is a much simpler measure than *Cns_TW*, easy to calculate and implement in business tendency surveys.

Due to the fact that high values of entropy and variance indicate lack of consensus (high dispersion), in order to meaningfully compare them with other measures, it is necessary to rescale them appropriately. We decide on a generally accepted unit interval as the range for all attainable values of consensus. Therefore let us define the entropy-based consensus measure as

$$Cns\_ENT = (H_{max} - H) / H_{max}, \tag{9}$$

and the variance-based consensus measure as

$$Cns\_VAR = 1 - Var = P^{const} + (BAL)^2. \tag{10}$$

In Table 2, we list the five measures of consensus in business tendency surveys: range (*RNG*), simplex (*Cns_SMPL*), entropy-based (*Cns_ENT*), Tastle and Wierman (*Cns_TW*), and variance-based (*Cns_VAR*), and compare

their values for selected shares of survey answers. The shares are selected to illustrate perfect consensus (last two rows), definite lack of consensus (first row), and various share structures in between those two extremes.

Table 2. Measures of survey consensus for selected shares of survey responses

p^{inc}	p^{const}	p^{dec}	*RNG*	*Cns_SMPL*	*Cns_ENT*	*Cns_TW*	*Cns_VAR*
0.50	0.00	0.50	0.50	0.50	0.37	0.00	0.00
0.33	0.33	0.33	0.00	0.00	0.00	0.33	0.33
0.85	0.00	0.15	0.85	0.79	0.62	0.39	0.49
0.25	0.50	0.25	0.25	0.25	0.05	0.50	0.50
0.20	0.60	0.20	0.40	0.40	0.43	0.60	0.60
0.50	0.50	0.00	0.50	0.50	0.37	0.58	0.75
0.71	0.28	0.01	0.70	0.61	0.74	0.63	0.77
0.01	0.98	0.01	0.97	0.97	0.90	0.98	0.98
1.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00

Source: Authors' own elaboration.

All consensus measures reach their maximum value of 1 when all respondents' forecasts belong to the same category. They differ, however, in terms of conditions necessary to attain the minimum value of 0. Let us note that three measures (*RNG*, *Cns_SMPL*, *Cns_ENT*) link the lowest consensus to share structure $P^{inc} = 0.33,\ P^{const} = 0.33,\ P^{dec} = 0.33$, when respondents evenly distribute their forecasts among all three categories (improvement, no change, decrease). Both *Cns_TW* and *Cns_VAR* link the lowest consensus to $P^{inc} = 0.50,\ P^{const} = 0,\ P^{dec} = 0.50$, when respondents are divided into two opposing groups, 50% each, declaring improvement and decrease. Three measures: *RNG*, *Cns_SMPL*, *Cns_ENT* regard the share structure $P^{inc} = 0.50,\ P^{const} = 0,\ P^{dec} = 0.50$, in which respondents are divided into absolute optimists and absolute pessimists, as identical to $P^{inc} = 0.50,\ P^{const} = 0.50,\ P^{dec} = 0$, in which respondents are divided into absolute optimists and uncommitted. Intuitively, in the latter case consensus should be much stronger. Both *Cns_TW* and *Cns_VAR* distinguish these two different cases and assign a higher value to the latter one.

To summarize, the consensus measures based on the Tastle-Wierman formula (*Cns_TW*) and variance (*Cns_VAR*) are the most advantageous as measures of the degree of agreement between qualitative survey respondents. However, they are burdened with the disadvantage of disregarding positive or negative nature of consensus.

Although we have tried to cover all measures of consensus used in economic studies, especially in business tendency surveys, we would like to emphasize that other measures of consensus can be found in the literature. For example, in the case of the Delphi method based on experts' opinions, an index of dispersion of the relative classification is used [**Wójciak, 2015**]. Additionally, there are several diversity indices described in ecological studies (e.g., **Daly et al. [2018]**). But all the mentioned indicators do not take into account the distance between answers, i.e. structures {50,50,0} and {50,0,50} are treated equally, while respondents who expect increase and no change are more similar in their responses than those expecting increase and decrease. The Delphi and ecological indicators share this disadvantage with the consensus measures already present in economic literature (range, simplex, and entropy measures), and therefore we do not include them in our analysis.

Consensus measures for the RIED tendency survey

Data on expectations and changes observed in the economy have been collected since 1986 by the Research Institute for Economic Development (RIED) at the Warsaw School of Economics through business tendency surveys. Launched for the manufacturing industry, currently they also cover households, the farming sector, the construction industry, the service industry, and the financial sector. The empirical part of this paper is based on the monthly survey addressed to enterprises in manufacturing, selected on the strength of the sample size:

data are available since 1997 with monthly frequency. The quality of the RIED survey data for the purposes of macroeconomic inference has been repeatedly confirmed; see **Adamowicz, Walczyk** [**2017**] and, with respect to the credibility of the RIED data with non-response, **Kowalczyk, Tomczyk** [**2008**, **2010**].

In each survey question, respondents are asked to evaluate both the 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 (see Appendix 1). Our sample consists of eight questions from monthly tendency surveys in industrial production ($k = 1, \ldots, 8$) for the period of April 1997 – April 2023 (that is, 313 observations). As we are interested in evaluating the level of consensus on economic variables included in the tendency survey with respect to the future, we take only expectations responses into account. On the basis of previous analyses of the RIED data, we assume expectation horizon $h = 3$.

In order to accurately define the variables in the RIED business tendency surveys, let us augment notation introduced in section 3 as follows:

$P_{k,t}^{inc}$ – percentage of respondents expecting, in time t, improvement three months in the future with respect to survey question k,

$P_{k,t}^{const}$ – percentage of respondents expecting, in time t, no change three months in the future, with respect to survey question k,

$P_{k,t}^{dec}$ – percentage of respondents expecting in time t, decline three months in the future, with respect to survey question k.

Five measures of consensus have been calculated for expectations concerning eight variables from the RIED industrial manufacturing survey: range (*RNG*), simplex (*Cns_SMPL*), entropy-based (*Cns_ENT*), Tastle and Wierman (*Cns_TW*), and variance-based (*Cns_VAR*) measures. Balance statistics have been widely used in the past to quantify and evaluate dynamics of the RIED survey data. Measures of entropy and dissimilarity of *a priori* (expectations) and *a posteriori* (observed changes) structures have also been applied to RIED data (see **Tomczyk** [**2011**, **2023**]) with a general conclusion that they provide a useful instrument of survey data analysis, among them for the purposes of evaluating the dynamics of economic variables at various stages of a business cycle. The remaining measures, to the best of our knowledge, are here applied to the Polish business survey data for the first time.

The results are summarized in Table 3 (means) and Table 4 (coefficients of variation).

Table 3. Mean values of measures of consensus

RIED survey question	*RNG*	*Cns_ENT*	*Cns_SMPL*	*Cns_VAR*	*Cns_TW*
01: production	0.314	0.078	0.284	0.522	0.476
02: orders	0.311	0.079	0.285	0.522	0.484
03: export orders	0.361	0.103	0.332	0.555	0.517
04: stocks	0.604	0.296	0.586	0.725	0.706
05: prices	0.667	0.363	0.630	0.762	0.709
06: employment	0.578	0.274	0.541	0.708	0.647
07: financial standing	0.475	0.176	0.440	0.632	0.590
08: general situation of the economy	0.486	0.196	0.443	0.635	0.557

Source: Authors' own elaboration.

Table 4. Coefficients of variation for measures of consensus

RIED survey question	*RNG*	*Cns_ENT*	*Cns_SMPL*	*Cns_VAR*	*Cns_TW*
01: production	0.204	0.408	0.203	0.079	0.091
02: orders	0.270	0.530	0.269	0.101	0.109
03: export orders	0.204	0.399	0.203	0.085	0.100
04: stocks	0.104	0.202	0.105	0.057	0.060

RIED survey question	*RNG*	*Cns_ENT*	*Cns_SMPL*	*Cns_VAR*	*Cns_TW*
05: prices	0.116	0.221	0.139	0.067	0.118
06: employment	0.110	0.182	0.130	0.052	0.114
07: financial standing	0.153	0.299	0.159	0.076	0.093
08: general situation of the economy	0.174	0.366	0.175	0.081	0.115

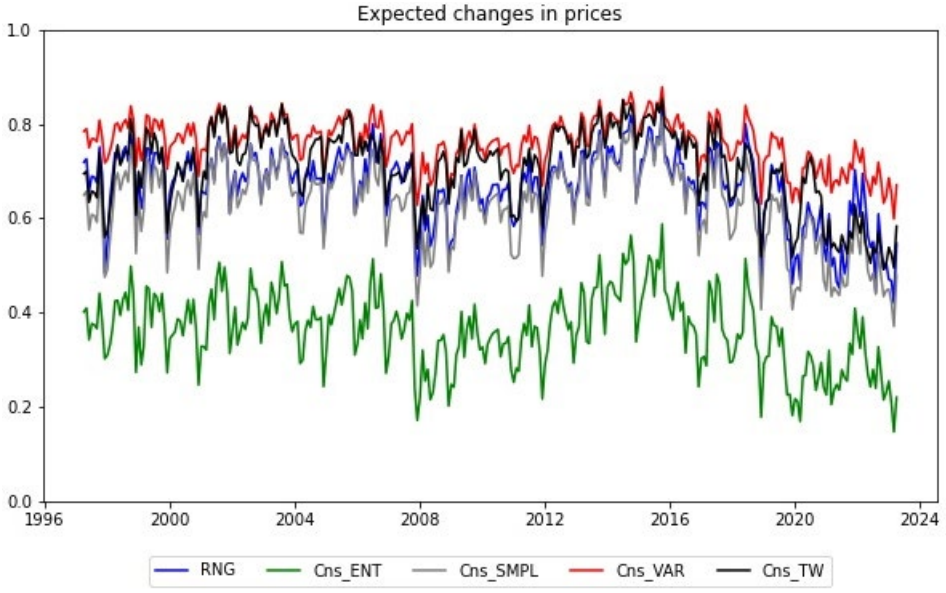
Source: Authors' own elaboration.

Although measures of consensus presented in this article are based on diverse mathematical methods, including the basic notion of range, variance, Shannon entropy and geometrical idea of a simplex, all indices are highly consistent in terms of evaluating the degree of agreement among respondents across economic variables. On average, consensus is the highest for prices, irrespective of the measure used. This may be due to the fact that prices are often dictated by earlier orders, expenditures and interest rates, and most enterprises included in the RIED survey are price takers without the power to set market prices. The relatively high consensus on prices may also follow from the general (economy-wide) character of price movements, an indirect consequence of strong interconnections between various sectors of the manufacturing industry.

The lowest consensus is noted for production and orders – again, independently of the measure used. We attribute this finding to the heterogeneity of business sectors represented in the RIED questionnaire: expectations on changes in the size of production and orders may significantly differ among industrial sectors due to their relative size, dependence on external financing or general business conditions, range of goods produced, and other factors.

While Table 3 compares the average values of the five measures, Figures 1 and 2 present the values of the measures of consensus in time. Once again, it is clear that all the measures consistently identify price changes as characterized by much higher consensus than expected changes in production.

Figure 1. Values of the five measures of consensus for expected changes in prices

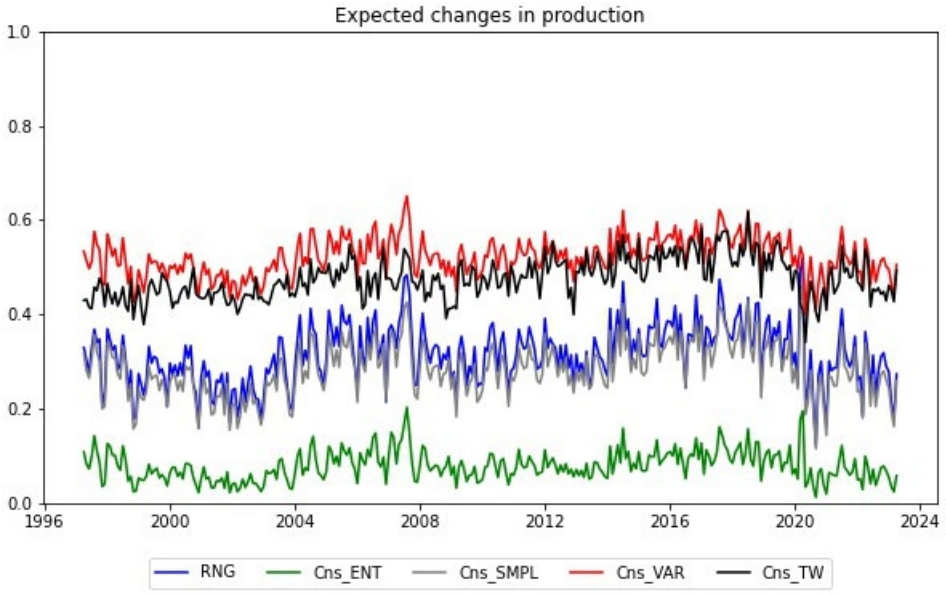


Source: Authors' own elaboration.

As far as changes in consensus over time are concerned (Table 4), the five measures are not so perfectly aligned. In particular, *Cns_TW* indicates a different set of variables with the highest and the lowest variability of consensus than all the other measures. Interestingly, the two variables for which average consensus is the lowest (that is, production and orders), are characterized by the highest coefficient of variation – again, with the exception of *Cns_TW*. This means that the degree of consensus for production and

orders is highly variable in time: there are months in which the degree of consensus differs considerably from the mean. The lowest variability of consensus is observed for stocks and employment; again, except for *Cns_ TW*. The atypical behavior or the Tastle-Wierman consensus measure *Cns_TW* makes it a promising subject for future research.

Figure 2. Values of the five measures of consensus for expected changes in production



Source: Authors' own elaboration.

It has to be emphasized that for all the economic variables and for all the measures of consensus (except for *Cns_ENT*), the coefficient of variation is lower than 0.27. This can be interpreted as, generally, low variability in changes of consensus over time, and relatively high values of coefficient of variation for *Cns_ENT* could be attributed to the fact that *Cns_ENT* takes values very close to zero.

It has been already noted in the literature that survey respondents' disagreement tends to be higher for expected unemployment than expected inflation [**Conflitti, 2011**]; we also find that, on average, consensus is higher (and therefore disagreement lower) for expected changes in prices than for changes in employment. However, these findings are not directly comparable as in the RIED survey, respondents are asked for expected changes in prices of their output instead of expected inflation, and for changes in employment instead of expected aggregated unemployment.

An attempt was made to verify whether macroeconomic events with obvious positive (access to the EU) or negative (financial crisis of 2008–09, pandemics of 2019–21, war in Ukraine) predicted influence on expectations could be matched to particularly high consensus among survey respondents. Also, we examined whether months directly preceding Polish parliamentary elections (when higher uncertainty could be expected among economic agents in general) are characterized by lower consensus. We did not find consistent evidence to support these hypotheses, in contrast to (admittedly few) results reported previously in literature. **Bachmann et al.** [**2013**] conclude that "bad times breed uncertainty" (p. 28); they define "uncertainty" in a way analogous to our formula (7). Polish manufacturing industry survey data does not confirm this finding.

The lack of a visible connection between the level of consensus observed in business tendency surveys and major macroeconomic events may be caused by the microeconomic profile of the RIED business tendency survey in manufacturing: respondents are pointedly asked about expectations on the performance of their own enterprise and sector of operation.

We have also searched for patterns in high and low consensus versus positive and negative values of balance statistic to verify whether the degree of consensus tends to be related to the degree of optimism in any of the

eight economic variables included in the RIED business tendency survey. In a few individual cases, a pattern emerges. For prices, the lowest values of consensus measures, both *Cns_VAR* and *Cns_TW*, are associated with positive values of balance statistic. The opposite case (low consensus with negative balance statistic) occurs for employment. We present the lowest ten values of the *Cns_VAR* and *Cns_TW* consensus measures with corresponding values of balance statistics for prices (Table 5) and employment (Table 6). The tables illustrate that all the values of balance statistics that match low levels of consensus are positive in the case of prices and negative in the case of employment.

Table 5. Consensus measures and balance statistics for expected changes in prices

Cns_VAR	Balance statistic	Period
0.598	0.280	Mar-23
0.628	0.228	Dec-07
0.630	0.290	Dec-18
0.631	0.048	Mar-20
0.632	0.491	Nov-22
0.633	0.306	Dec-19
0.646	0.506	Dec-22
0.647	0.538	Jun-22
0.652	0.353	Feb-23
0.654	0.295	Mar-21

Cns_TW	Balance statistic	Period
0.492	0.491	Nov-22
0.495	0.280	Mar-23
0.503	0.506	Dec-22
0.505	0.538	Jun-22
0.516	0.306	Dec-19
0.519	0.290	Dec-18
0.522	0.353	Feb-23
0.525	0.509	Sep-21
0.526	0.562	May-22
0.527	0.434	Aug-22

Source: Authors' own elaboration.

Table 6. Consensus measures and balance statistics for expected changes in employment

Cns_VAR	Balance statistic	Period
0.622	–0.248	Apr-02
0.624	–0.207	Sep-97
0.626	–0.098	Feb-04
0.627	–0.286	Jun-02
0.629	–0.236	Nov-03
0.630	–0.185	Apr-97
0.631	–0.395	Mar-99
0.632	–0.068	May-08
0.633	–0.188	Jun-97
0.635	–0.201	Aug-97

Cns_TW	Balance statistic	Period
0.498	–0.395	Mar-99
0.510	–0.378	Feb-99
0.511	–0.372	Jun-99
0.516	–0.373	Aug-99
0.516	–0.371	May-99
0.517	–0.286	Jun-02
0.520	–0.419	Sep-99
0.520	–0.502	Oct-01
0.522	–0.356	Sep-00
0.523	–0.383	Jun-01

Source: Authors' own elaboration.

This result suggests that survey respondents tend to disagree on prices when they generally expect prices to rise, and to disagree on changes in employment when they generally expect employment to fall. The finding is confirmed by negative correlation coefficients for balance statistics and measures of consensus in case of prices (–0.581 for *Cns_VAR*; – 0.818 for *Cns_TW*), and positive correlation coefficients for balance statistics and measures of consensus in the case of employment (0.554 for *Cns_VAR*; 0.889 for *Cns_TW*). We can find partial support for these findings in previous analyses: there is evidence that expectations of higher inflation are associated with more disagreement on expected inflation [**Zarnowitz, Lambros, 1987**; **D'Amico, Orphanides, 2008**]. Again, it is important to note that respondents of the RIED business tendency surveys are not asked about expected inflation but about expected changes in the prices of goods produced, and the comparison can only be approximate.

The interpretation becomes even more complex in the case of the general situation of the economy because both the highest and the lowest values of measures of consensus can be matched to negative values of the bal-

ance statistic. It seems that pessimistic outlook on the economy in general can be both the subject of visible disagreement among the respondents (low values of *Cns_VAR* and *Cns_TW* in the first half of 2021) and evident agreement (high values of *Cns_VAR* for the second half of 2022 and of *Cns_TW* in the spring of 2020).

Conclusions and directions of further research

In this paper, we collected five consensus measures previously used to evaluate qualitative survey data, compared their theoretical properties, and conducted empirical analysis for eight variables included in the RIED monthly business tendency survey in manufacturing. The comparison of the measures of consensus leads to a general conclusion that the theoretical properties of two of these methods – variance-based and Tastle-Wierman – make them particularly appropriate in evaluating the degree of consensus among tendency survey respondents. However, the empirical part of our analysis shows that the Tastle-Wierman measure (*Cns_TW*) behaves differently in terms of measuring the variability of consensus in time, and should be further inspected in the future.

It has to be emphasized that all the measures of consensus do not distinguish between structures {1,0,0} and {0,0,1}, which are fundamentally different: the first reflects perfect optimism, and the second – perfect pessimism. Only by combining two indicators – a measure of consensus and a measure of the degree of optimism, e. g. the balance statistic *BAL* – we can distinguish between the two cases. A measure which describes both the strength and the direction (optimistic or pessimistic) of a consensus would constitute a very useful variable in econometric modelling and in the construction of business tendency indicators.

Also, the empirical part of this paper identifies economic variables on the extremes of the consensus spectrum: expected prices with the highest degree of consensus among survey respondents, and expected production and orders with the lowest degree of consensus. In a few cases (prices, employment, and general business conditions), we also found linkages between the degree of consensus and the degree of optimism among respondents as measured by the balance statistic.

An obvious expansion of this analysis would involve international comparisons. The possibilities for cross-country analyses are unfortunately limited due to data constraints: few European agencies publish sufficiently detailed data on expectations in the manufacturing industry to enable calculations analogous to these presented above (namely, percentages of respondents expecting improvement and deterioration for several economic variables), and if they do, the wording of the questionnaires does not make them directly comparable to the RIED surveys.

The degree of consensus might be further used as an explanatory variable in econometric models to stand for uncertainty among economic agents, or employed as a leading indicator of respondent sentiment. However, in econometric modelling it would be important to distinguish between "positive" consensus (agreement that things will go well) and "negative" consensus (agreement that things will go badly), and not only agreement or lack of it. Therefore, the consensus measures described in this paper are not directly usable for this purpose. Mixed results noted so far in the literature (for example, positive in **Krüger and Nolte [2016]**, mixed to negative in **D'Amico and Orphanides [2008]**) have been based on quantitative and density forecasts and therefore are not directly comparable with measures derived from qualitative expectations.

As the final direction of further study we would like to point to the importance of analysis of the "no change" category, which may include both honest "no change" responses and secondary reactions such as "I have no idea" or even "why should I bother". This is a very important point in all analyses of qualitative survey data, sadly neglected because of the problems involved in gathering appropriate individual-level data.

Literature

Adamowicz E., Walczyk K. [2017], Zaburzenia cykliczności aktywności gospodarczej w Polsce w świetle wyników badania koniunktury gospodarczej IRG SGH, *Prace i Materiały Instytutu Rozwoju Gospodarczego SGH,* 101.

Bachmann R., Elstner S., Sims E. R. [2013], Uncertainty and Economic Activity: Evidence from Business Survey Data, *American Economic Journal: Macroeconomics, American Economic Association*, 5 (2): 217–249.

Claveria O. [2019], Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations, *Journal for Labour Market Research*, 53 (3): 1–10.

Claveria O., Monte E., Torra S. [2019], Economic Uncertainty: A Geometric Indicator of Discrepancy Among Experts' Expectations, *Social Indicators Research*, 143: 95–114.

Conflitti C. [2011], *Measuring Uncertainty and Disagreement in the European Survey of Professional Forecasters*, ECARES (Université libre de Bruxelles) Working Paper.

Daly A. J., Baetens J. M., De Baets B. [2018], Ecological Diversity: Measuring the Unmeasurable, *Mathematics*, 6/7 (119), https://doi.org/10.3390/math6070119.

D'Amico S., Orphanides A. [2008], *Uncertainty and Disagreement in Economic Forecasting*, Finance and Economics Discussion Series No. 2008–56, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D. C.

Kowalczyk B., Tomczyk E. [2008], Rationality of Expectations of Industrial Enterprises – Analysis Based on Business Tendency Surveys with Item Non-response, *Bank i Kredyt*, 8: 3–11.

Kowalczyk B., Tomczyk E. [2010], Influence of non-response in business tendency surveys on properties of expectations, *Statistics in Transition (New Series)*, 11 (2): 403–422.

Krüger F., Nolte I. [2016], Disagreement versus uncertainty: Evidence from distribution forecasts, *Journal of Banking and Finance*, 72: S172 – S186.

Tastle W. J., Wierman M. J. [2006], Consensus and dissention: A measure of ordinal dispersion, *International Journal of Approximate Reasoning*, 45: 531–545.

Tomczyk E. [2011], Application of measures of entropy, information content and dissimilarity of structures to business tendency survey data, *Przegląd Statystyczny*, 58 (1–2): 88–101.

Tomczyk E. [2023], Dynamics of survey expectations and assessments before and during the pandemic: entropy and dissimilarity measures applied to business tendency survey data, *Statistic in Transition (New Series)*, 24 (2): 185–199.

Wójciak M. [2015], Metody oceny zgodności opinii ekspertów na potrzeby badania foresight, *Studia Ekonomiczne*, 220: 58–77.

Zarnowitz V., Lambros L. A. [1987], Consensus and Uncertainty in Economic Prediction, *Journal of Political Economy*, 95 (3): 591–621.

Appendix 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: RIED database.