
European Commission

Directorate General Economic and Financial Affairs

Task-force on Quality of EU BCS data, Sub-group 6:

Volatility Analysis and Tracking Performance of Individual Questions

Andreas Reuter

*Business and consumer surveys and
short-term forecast (ECFIN A4.2)*

Outline

1. Introduction: Why focussing on the quality of individual survey questions?
 - distinction between quality differences of balance-series on horizontal / vertical / micro level
2. Quality differences of balance-series from vertical perspective
 - analysis conducted by TURKSTAT / DG ECFIN
3. Quality differences of balance-series from micro-perspective
 - analysis conducted by DG ECFIN

1. Why focussing on the quality of individual survey questions?

Industry Survey: MCDs by Country and by Question									
	Confidence Indicator	Q1	Q2	Q3	Q4	Q5	Q6	Q7	
		prod t-3	orders	exp. ord.	stocks	prod t+3	prices t+3	empl t+3	
BE	2	3	2	2	4	3	2	3	2.7
BG	3	3	2	3	4	4	3	3	3.1
CZ	3	5	3	2	5	5	4	2	3.7
DK	2	3	2	2	3	4	4	3	3.0
DE	1	3	1	1	2	2	2	2	1.9
EE	3	3	2	3	4	4	3	2	3.0
IE									
EL	3	6	3	4	4	4	4	4	4.1
ES	3	3	3	3	3	4	3	3	3.1
FR	2	2	2	2	3	3	3	2	2.4
IT	2	2	2	2	4	3	2	3	2.6
CY	4	5	4	8	5	6	6	9	6.1
LV	4	3	3	4	4	4	3	3	3.4
LT	3	9	3	3	4	4	4	3	4.3
LU	2	4	2	2	4	4	3	4	3.3
HU	3	3	3	3	5	4	3	3	3.4
MT	4	4	4	4	6	3	5	8	4.9
NL	2	5	2	2	3	3	2	2	2.7
AT	2	2	2	2	3	3	2	2	2.3
PL	2	5	2	2	4	4	4	2	3.3
PT	3	5	3	3	5	4	4	3	3.9
RO	3	4	3	3	3	4	4	3	3.4
SI	2	5	2	2	4	4	2	2	3.0
SK	5	10	4	5	6	10	6	3	6.3
FI	2	3	2	2	3	4	3	2	2.7
SE	2	2	2	2	4	3	3	2	2.6
UK	3	3	3	3	4	3	3	4	3.3
	2.7	4.0	2.5	2.8	4.0	4.0	3.3	3.2	

Individual questions in individual countries with extremely low quality balance series

quality differences at MICRO level

treated in this presentation
big differences between lowest and highest average MCDs

HORIZONTAL quality differences

Possible explanations:

- different sample sizes
- different sampling methods
- etc.

treated in sub-groups 1-5

Q4 more volatile than Q2

VERTICAL quality differences

Possible explanations:

- Is question forward or backward looking?
- etc.

treated in this presentation



2. Quality differences of balance-series (vertical perspective)

Which quality measures to examine?

- volatility (measured as MCD of a given question in a given country)
- correlation (of a given question in a given country with its "hard-data" reference series)

Which question-specific characteristics could impact the quality indicators?

- as opposed to characteristics explained by:
- the survey (retail trade survey more volatile than industry survey)
 - the survey design (different sample sizes, etc.)

characteristic	impact on MCD	impact on correlation
time-horizon	future questions: + (future involves un-stable gut-feeling)	past questions: + (should correctly reflect what happened)
change vs. level	change questions: + (change is 1 st derivative of level)	???
judgmental vs. factual	judgemental questions: + (judgemental question: "Do you consider order books to be more than sufficient for the next 3 months?")	factual questions: + (factual question: "How has your production developed over the past 3 months?")
3 vs. >3 answers (cons. survey)	(if 3 answers, 2 categories pos. and for neg. in line with ref. series)	(if >3 answers, 2 categories pos. and for neg. in line with ref. series)
direct vs. indirect link of question with ref. series	direct questions: + (ref. series are usually rather volatile)	indirect questions: + (question is closely related to ref. series)

i) explaining differences in MCD / correlation with ref. series

- data-set:**
- cross-sectional
 - one observation (e.g. MCD) refers to:
 - one survey-question (e.g. Indu Q1, Serv Q2)
 - in one country
- } n = 945
(27 MS * 35 questions)

regression: **MCD** or **corr. coefficient** =

c	backward question	current question	change question	question with 5 answ. categories	judg-mental question	question with direct link to ref. series
----------	--------------------------	-------------------------	------------------------	---	-----------------------------	---

+

question-specific dummies (e.g. INDU Q1)

} no dummies for consumer questions, since otherwise (nearly) perfect collinearity

ii) results for MCD: stepwise down-testing results in the following specification:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.486048	0.301882	14.86025	0.0000
backward-question	-0.320521	0.159897	-2.004543	0.0453
change-question	-2.444327	0.235885	-10.36235	0.0000
5 answering categories	1.328767	0.224946	5.907045	0.0000
judgmental question	-1.877652	0.279223	-6.724554	0.0000
direct link to reference series	1.949384	0.220271	8.849948	0.0000
INDU4	1.353143	0.340891	3.969429	0.0001
INDU6	1.304433	0.388611	3.356657	0.0008
INDU7	1.112125	0.388611	2.861797	0.0043
SERV1	1.737133	0.386412	4.495550	0.0000
SERV4	2.496191	0.391509	6.375824	0.0000
SERV5	2.083279	0.397534	5.240502	0.0000
SERV6	2.499945	0.397534	6.288630	0.0000
BUIL1	1.209415	0.328782	3.678468	0.0002
BUIL3	0.776220	0.340891	2.277032	0.0230
BUIL4	1.842894	0.388611	4.742264	0.0000
BUIL5	1.419817	0.388611	3.653573	0.0003
RETA2	3.217692	0.356633	9.022418	0.0000
RETA3	2.262627	0.402490	5.621566	0.0000
RETA5	2.306105	0.402490	5.729590	0.0000
RETA6	2.001757	0.402490	4.973428	0.0000
Adjusted R-squared	0.226883	= un-expected		= as expected ⁶

iii) results for corr. coefficient: stepwise down-testing results in the following specification:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	48.89464	2.107619	23.19899	0.0000
backward-question	10.03601	2.856350	3.513577	0.0005
5 answering categories	-27.55193	2.670782	-10.31605	0.0000
judgmental question	-15.47063	4.770779	-3.242789	0.0012
direct link to reference series	11.96317	3.035868	3.940610	0.0001
INDU2	24.99907	8.093801	3.088669	0.0021
INDU3	23.46061	8.093801	2.898590	0.0038
INDU4	-68.80862	8.093801	-8.501398	0.0000
BUIL3	28.52336	8.967554	3.180729	0.0015
RETA1	-15.63295	7.414598	-2.108401	0.0353
RETA2	-40.46749	8.413524	-4.809814	0.0000
Adjusted R-squared	0.331077	= un-expected		= as expected

iv) "crude" sub-summary:

Questions that perform well:

- backward-looking questions >>low volatility
>>high tracking performance
- questions with direct link to reference series >>high tracking performance
(although at cost of increased volatility)

Questions that perform less well:

- questions with 5 answering categories >>high volatility
>>low tracking performance
- judgmental questions >>low tracking performance: **15pp** lower than reference group
>>lower MCD **cannot** compensate for the decreased tracking performance

Which questions cannot be judged upon?

- change questions >>associated with lower MCD – **not intuitive!**

Findings might be useful for composite indicator construction!

3. Quality differences of balance-series from micro-perspective

Industry Survey: MCDs by Country and by Question									
	Confidence Indicator	Q1	Q2	Q3	Q4	Q5	Q6	Q7	average MCD: across questions, by country
		prod t-3	orders	exp. ord.	stocks	prod t+3	prices t+3	empl t+3	
BE	2	3	2	2	4	3	2	3	2.7
BG	3	3	2	3	4	4	3	3	3.1
CZ	3	5	3	2	5	5	4	2	3.7
DK	2	3	2	2	3	4	4	3	3.0
DE	1	3	1	1	2	2	2	2	1.9
EE	3	3	2	3	4	4	3	2	3.0
IE									
EL	3	6	3	4	4	4	4	4	4.1
ES	3	3	3	3	3	4	3	3	3.1
FR	2	2	2	2	3	3	3	2	2.4
IT	2	2	2	2	4	3	2	3	2.6
CY	4	5	4	8	5	6	6	9	6.1
LV	4	3	3	4	4	4	3	3	3.4
LT	3	9	3	3	4	4	4	3	4.3
LU	2	4	2	2	4	4	3	4	3.3
HU	3	3	3	3	5	4	3	3	3.4
MT	4	4	4	4	6	3	5	8	4.9
NL	2	5	2	2	3	3	2	2	2.7
AT	2	2	2	2	3	3	2	2	2.3
PL	2	5	2	2	4	4	4	2	3.3
PT	3	5	3	3	5	4	4	3	3.9
RO	3	4	3	3	3	4	4	3	3.4
SI	2	5	2	2	4	4	2	2	3.0
SK	5	10	4	5	6	10	6	3	6.3
FI	2	3	2	2	3	4	3	2	2.7
SE	2	2	2	2	4	3	3	2	2.6
UK	3	3	3	3	4	3	3	4	3.3
average MCD: by question, across countries	2.7	4.0	2.5	2.8	4.0	4.0	3.3	3.2	

Point of departure:

- "horizontal" differences have been analysed/explained by sub-groups 1-5
- "vertical" differences have been analysed/explained in first part of this presentation
- quality differences at "micro level" remain unexplained by "horizontal" and "vertical" differences

How come that individual survey questions in individual countries are of particularly low quality?

Hypothesis:

Wrong / distorting translation of original EU BCS question

Criteria for determining idiosyncratic behaviour of survey questions

**Correlation of balance-series*
and reference series**

correlation of balance series
with reference series has
counter-intuitive sign

correlation of balance series
with reference series is
particularly low

*and /
or*

MCD of balance-series*

MCD of balance series is
>=150% of average MCD of all
balance series generated by
respective national
questionnaire

Results:

- **74 national questions show idiosyncratic behaviour**
- **= 11% of the analysed questions**

translation of 74 Q from
national questionnaire to
English to check for
wrong/distorting translations

* Only balance-series of EU-27 in industry / services / retail trade / construction were analysed (to keep translation work manageable)



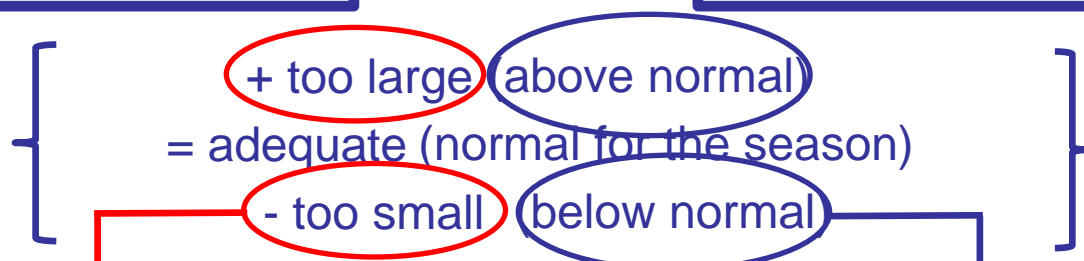
Results of the Translation Exercise

- 37 / 74 questions with idiosyncratic behaviour are based on formulations clearly deviating from EU BCS questionnaire
 - 28 / 74 questions with idiosyncratic behaviour display deviations which:
 - can be argued to explain (part of) the idiosyncratic behaviour
 - can be observed in a number of cases
-
- These 28 questions shall be discussed here.
 - They can be grouped into **4 types of translation errors / deviations**

Type 1 of Translation Error / Deviation

INDU Q4: Do you consider your current stock of finished products to be...?

RETA Q2: Do you consider the volume of stock you currently hold to be...?



- stocks are assessed in relation to (expected) demand*
- firm replying "too large" is likely to drive down production / stop purchasing further stocks*
- stocks are assessed compared to "normal times" (in terms of season)*
- firm replying "above normal" might be:*
 - i) unhappy with high stocks*
 - >> drive down production / purchases*
 - ii) happy with high stocks (expecting high demand)*
 - >> **NOT** drive down production / purchases*

answering categories "too large" etc. capture cyclical signals best, categories "above normal", etc. will do worst

Results:

- 3 / 6 INDU Q4 balance series with counter-intuitive / low correlation
- 2 / 7 RETA Q2 balance series with counter-intuitive / low correlation

are based on questionnaires with "above normal" etc. as **only** answering categories

Recommendations:

- Ideally, EU BCS questionnaire would be changed to keep "too large", etc. as **only** answ. categories
- At least, national questionnaires should **not only** feature the "above / below normal" option

Type 2 of Translation Error / Deviation: Wrong Time-horizon

Several EU BCS questions ask for developments "**over the past 3 months**"

Several EU BCS questions ask for developments "**over the next 3 months**"

Some national questionnaires inquire instead "developments... :

- ...in the quarter that just ended, compared to the previous one"
- ...over the preceding 2-3 months"
- ...in the last 3 months, compared to the previous period"
- ...in the previous month (*singular!*)"
- ...in this month, compared to the average of the last 3 months"

Some national questionnaires inquire instead "developments... :

- ...during the past (sic!) three months"
- ...in this month, compared to the same month a year ago" (only for question referring to levels)
- ...over the next 2-3 months"

- using different time-horizons – **esp. shorter ones** - might be associated with higher volatility (= higher MCD)
example: DE Building survey uses correct time-horizons, except for Q1 (developments "in last month" instead of "over the past 3 months"). **Result:** MCD of Q1: 9; average of all other questions of the survey: 2.3.

Results:

- 9 of the "problematic" balance series are based on questions inquiring wrong time-horizon
- 7 of the 9 questions have MCD \geq 150% of average MCD of all balance series generated by respective national questionnaire

Recommendations:

Time-horizons inquired in national questionnaires should be brought in line with EU BCS questionnaire.

Type 3 of Translation Error / Deviation: Wrong concept inquired

Examples:

~~**INDU Q4:** Do you consider your current stock of finished products to be... too large (above normal), etc.?~~

~~**INDU Q1:** How has your production developed over the past 3 months? It has... increased, etc.?~~

inquires "production growth"

DK INDU Q4: The stock of finished products is... smaller, approximately unchanged, larger, (no stocks of finished products)?

DK INDU Q4 has correlation of +0.06 with IPI

SI INDU Q1: Our production dynamics in the last month (sic!) compared to the previous month (sic!) was... buoyant, without change, decreasing.

"production dynamics" = change in "production growth"
= 2nd derivative of "production level"
>> higher volatility than "production growth" (which is 1st derivative of "production level")
>> SI INDU Q1: higher MCD than the other questions

Recommendations:

Bring national questionnaires back in line with EU BCS questionnaire as regards concepts inquired.

Type 4 of Translation Error / Deviation: Adding questions to questionnaire

Adding questions inquiring concepts which could have been understood as being sub-categories of the initial EU BCS question

Example:

- EL INDU questionnaire contains original EU BCS question on "past production" (INDU Q1)
 - a question is added, inquiring the level of past "total sales".
-
- respondent will reply to INDU Q1, relying on data **other than sales**
 - assuming that "sales" serve as proxy used by respondents when answering INDU Q1, asking an explicit question on "sales" means that respondents will feel the need to assess production differently
 - given that production is hard to measure if sales are not considered, respondents' answers to INDU Q1 in Greece are likely to be more volatile

Indeed, EL INDU Q1 has MCD of 6, compared to average MCD of 4 regarding all other questions of the questionnaire.

Recommendations:

Avoid adding questions inquiring sub-concepts of the concepts inquired by EU BCS questions.

Or: put the additional questions at the end of the questionnaire, separated from "similar" EU BCS questions.

sub-summary: Overview of Recommendations

Recommendation 1:

- ideally, EU BCS questionnaire would be changed to keep "too large", etc. as **only** answ. categories
- at least, national questionnaires should **not only** feature the "above / below" normal option

Recommendation 2:

- time-horizons inquired in national questionnaires should be brought in line with EU BCS questionnaire.

Recommendation 3:

- bring national questionnaires back in line with EU BCS questionnaire as regards concepts inquired.

Recommendation 4:

- avoid adding questions inquiring sub-concepts of the concepts inquired by EU BCS questions.
- alternative: put the additional questions at the end of the questionnaire, separated from "similar" EU BCS questions.

Furthermore:

- institutes should check their questionnaires in the light of the findings regarding translations
- DG ECFIN will come back to individual institutes in case of severely deviating translations