Quality of BCS data
Results of Task Force 1 - Sample Frames

Jonathan Wood

Survey Management Group

Brussels 14th – 15th November 2013

Objectives today

- Brief introduction about the CBI
- Taskforce purpose/structure/terms of reference
- Classification of institutes/frames
- General frame analyses
- Specific analyses
  - Cross checks between frame characteristics
  - Analysis of MCD {volatility} across surveys
  - Analysis of correlation [tracking performance} across surveys
- Key conclusions

Purpose of Taskforce 1 – sampling frames

- Quality of BCS data terms of reference: section 1V Taskforce on ‘quality of BCS data’
  - Analysis of ***sampling frames*** across institutes: appropriateness and comprehensiveness of sampling frames, theoretical considerations, empirical evidence on links with data volatility and bias;

Active Taskforce Members

- Jonathan Wood – CBI Head of Survey Management
- Christopher Taylor – CBI Technical Survey Development Executive
- Alan Joy – Technical and statistical expert for the CBI
- Daniel Lee – CBI Senior Economist
  - Jelena Jakic {Ipsos ME}, Penna Urrila {EK Fi}

Terms of reference

- Analysis of how each institute applies sampling frames
  - Firstly, using the sample frame column on the metadata industry/services/retail/construction supplied by the European Commission
  - Secondly, referring back to institutes **where necessary** to capture further detail on their frame practices

Terms of reference 2

- Analysis of common links / factors between sampling frames:
- Developing a metric to illustrate the comparisons and contrasts of practice – what are the common and uncommon factors? This matrices workbook is available for sharing at:
- Structural differences in sampling frame practice by institutes
- Identification and analysis of any tangible link between sampling practice and volatility and correlation.

Classification of Institutes

- DG Ecfin applied the following classification for institutes:
  - Statistical institutes
  - Business associations
  - Private bodies
  - Other public bodies
  - Academic

Classification of frames

- Bought list
- Internally compiled list
- National register
- Private register
- Combination
  - Total

Sample frames – type of institute conducting each business survey

Type of institute conducting each business survey

	Survey			
Type of institute	**INDU**	**SERV**	**RETA**	**BUIL**
Academic	3	3	3	3
Business Assosciation	4	4	4	3
Other Public bodies	3	2	2	3
Private Bodies	2	3	3	4
Statistical Institute	14	14	14	13
Total	26	26	26	26

Sample frames – type of frame used for each business survey

Type of frame used for each business survey

	Survey					
Type of institute	**INDU**	**SERV**	**RETA**	**BUIL**	**All**	**%**
Bought List	0	1	1	1	3	3%
Internally Compiled list	3	3	3	2	11	11%
National Register	16	13	13	13	55	53%
Private Register	3	3	4	4	14	13%
Combination	4	6	5	6	21	20%
Total	26	26	26	26	104	100%

Sample frames – size of frame as a percentage of the population for each business survey

Size of frame as a percentage of the population for each business survey

	Survey					
Frame size as % of population - band	INDU	SERV	RETA	BUIL	All	%
<5%	2	5	3	2	12	14%
5% to <20%	8	8	8	9	33	39%
20% to <50%	6	3	3	4	16	19%
50% to <100%	2	2	2	3	9	11%
100%	4	4	3	4	15	18%
Total	22	22	19	22	85	100%

Sample frames – frequency of updating for each business survey

Frequency of updating for each business survey

	Survey					
Updating frequency - band	**INDU**	**SERV**	**RETA**	**BUIL**	**All**	**%**
Monthly/ Continuously	4	4	3	5	16	16%
Yearly	14	14	13	13	54	55%
Interval over one year	7	7	7	7	28	29%
Total	25	25	23	25	98	100%

Cross-checks between frame characteristics

Sample frames – link between type of institute and type of frame

Link between type of institute and type of frame

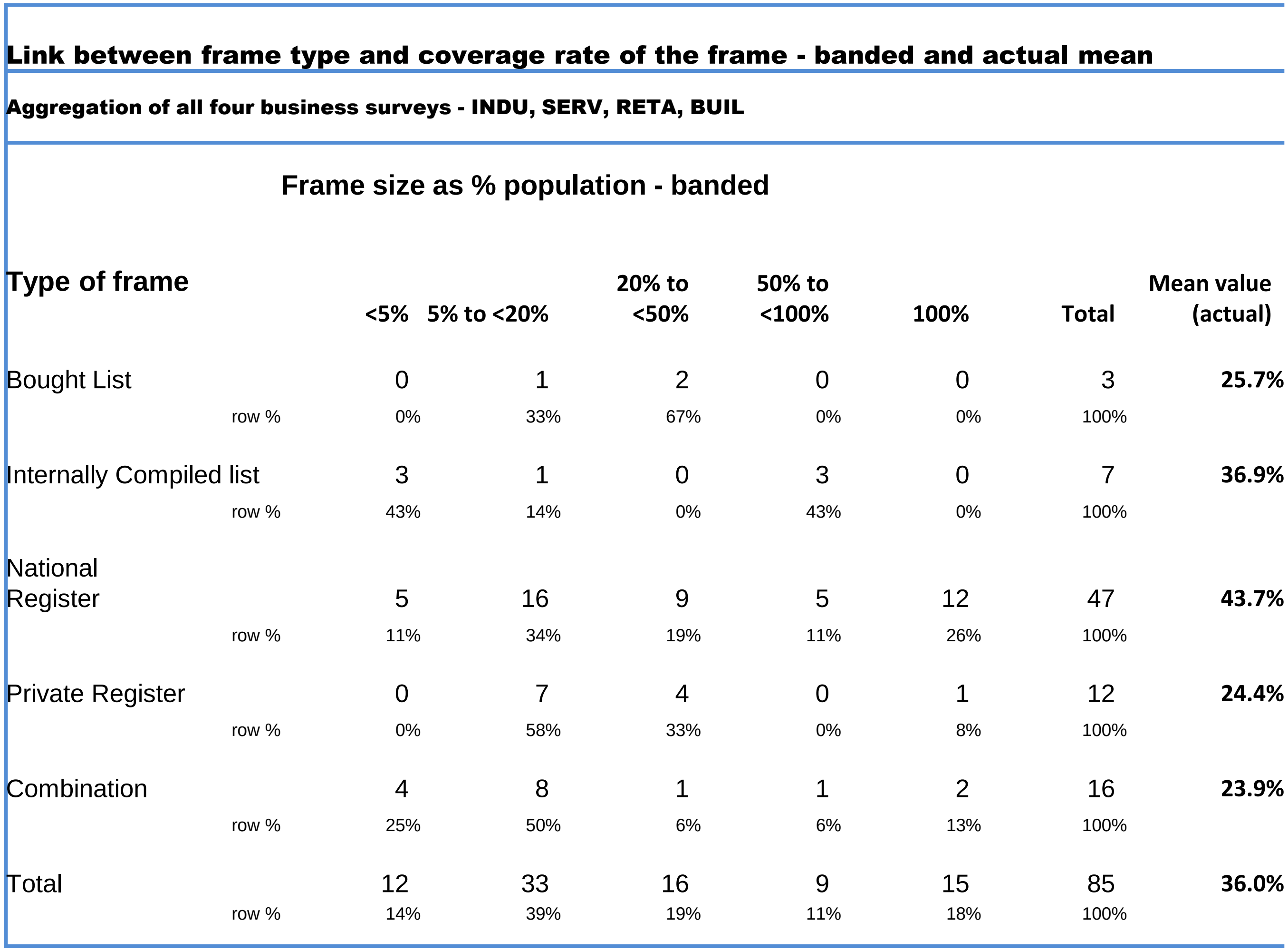
Aggregation of all four business surveys - INDU, SERV, RETA, BUIL

		Type of frame					
Type of institute		Bought List	Internally Compiled list	National Register	Private Register	Combination	**Total**
Academic		0	0	4	2	6	12
	row %	0%	0%	33%	17%	50%	100%
Business Association		0	3	1	8	3	15
	row %	0%	20%	7%	53%	20%	100%
Other Public bodies		1	0	9	0	0	10
	row %	10%	0%	90%	0%	0%	100%
Private Bodies		2	1	5	0	4	12
	row %	17%	8%	42%	0%	33%	100%
Statistical Institute		0	7	36	4	8	55
	row %	0%	13%	65%	7%	15%	100%
Total		3	11	55	14	21	104
	row %	3%	11%	53%	13%	20%	100%

There are strong links between institutes and frames

- For example, only one ‘business association’ survey uses a national register, but 90% of ‘other public bodies’ surveys do so
- Difficult to identify the independent impacts (if any) of institute types and frame types
- Note limited sample sizes – only 3 academic institutes for example (producing 12 surveys).

Sample frames – link between frame type and coverage rates of the frame



Link between frame type and coverage rate of the frame - banded and actual mean

Aggregation of all four business surveys - INDU, SERV, RETA, BUIL

Frame size as % population - banded

Type of frame		**<5%**	**5% to <20%**	**20% to <50%**	**50% to <100%**	**100%**	**Total**	**Mean value (actual)**
Bought List		0	1	2	0	0	3	**25.7%**
	row %	0%	33%	67%	0%	0%	100%	
Internally Compiled list		3	1	0	3	0	7	**36.9%**
	row %	43%	14%	0%	43%	0%	100%	
National Register		5	16	9	5	12	47	**43.7%**
	row %	11%	34%	19%	11%	26%	100%	
Private Register		0	7	4	0	1	12	**24.4%**
	row %	0%	58%	33%	0%	8%	100%	
Combination		4	8	1	1	2	16	**23.9%**
	row %	25%	50%	6%	6%	13%	100%	
Total		12	33	16	9	15	85	**36.0%**
	row %	14%	39%	19%	11%	18%	100%	

Sample frames - links between frame type and frequency of updating

Aggregation of all four business surveys - INDU, SERV, RETA, BUIL

		Updating frequency			
Type of frame		**Monthly/ Continuously**	**Yearly**	**Interval over one year**	**Total**
Bought List		0	2	1	3
	row %	0%	67%	33%	100%
Internally Compiled list		1	8	2	11
	row %	9%	73%	18%	100%
National Register		5	27	18	50
	row %	10%	54%	36%	100%
Private Register		0	9	4	13
	row %	0%	69%	31%	100%
Combination		9	8	4	21
	row %	43%	38%	19%	100%
Total		15	54	29	98
	row %	15%	55%	30%	100%

There are links between frame coverage and updating frequency and institute/frame types: 1

- Surveys using national registers have the highest frame coverage, on average, followed by internally compiled lists

- Consequently, statistical institutes have a high average frame coverage of 43%

- Again - limited number of surveys and institutes mean caution is required.

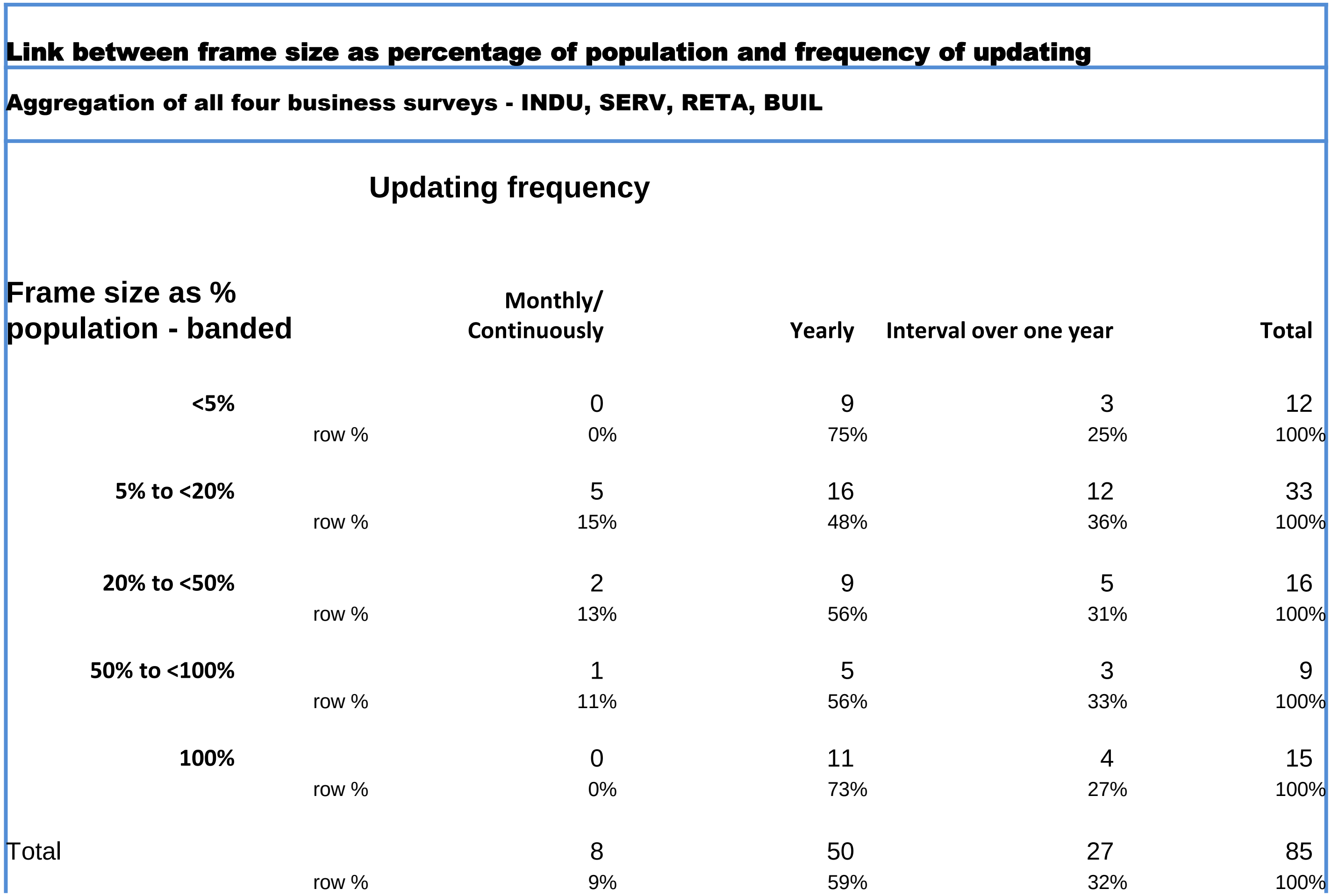
There are links between frame coverage and updating frequency and institute/frame types: 2

- ‘Combination’ frame types are most frequently updated, followed by national registers and internally compiled lists

- Relatedly, academic institutes and business associations have higher-than-average update frequencies

- Key finding: it is difficult to dis-entangle the effects of institute/frame type and frame coverage/frequency of updating

Sample frames – the link between frame size as a percentage of population and frequency of updating



Link between frame size as percentage of population and frequency of updating

Aggregation of all four business surveys - INDU, SERV, RETA, BUIL

Updating frequency

Frame size as % population - banded		Monthly/ Continuously	Yearly	Interval over one year	Total
<5%		0	9	3	12
	row %	0%	75%	25%	100%
5% to <20%		5	16	12	33
	row %	15%	48%	36%	100%
20% to <50%		2	9	5	16
	row %	13%	56%	31%	100%
50% to <100%		1	5	3	9
	row %	11%	56%	33%	100%
100%		0	11	4	15
	row %	0%	73%	27%	100%
Total		8	50	27	85
	row %	9%	59%	32%	100%

The links between frame coverage and frequency of updating are less marked

- There is no particularly strong link between the frame coverage and the frequency of updating across surveys. Those with a small (<5%) or maximum(100%) frame coverage are less likely to be updated continuously or monthly.

- The fact that 100% coverage surveys are updated less regularly suggests a slight trade-off

- ‘Multi-collinearity’ shouldn’t be a major issue when analysing frame coverage and updating frequency

Analysis of MCD (volatility) across surveys

Analysis of MCD – initial hypotheses

- Higher updating frequencies would be expected to reduce volatility (and the MCD)

- *Absolute* frame size may be more important than the frame size as a % of the total population

Average volatility by updating frequency

Mean MCD by frequency of updating

	Survey				
Updating frequency - band	**INDU**	**SERV**	**RETA**	**BUIL**	**All**
Monthly/ Continuously	1.8	2.3	3.7	2.4	**2.5**
Yearly	2.6	3.1	3.5	3.0	**3.0**
Interval over one year	3.4	2.6	3.9	3.6	**3.4**
Total	**2.7**	**2.8**	**3.6**	**3.0**	**3.0**

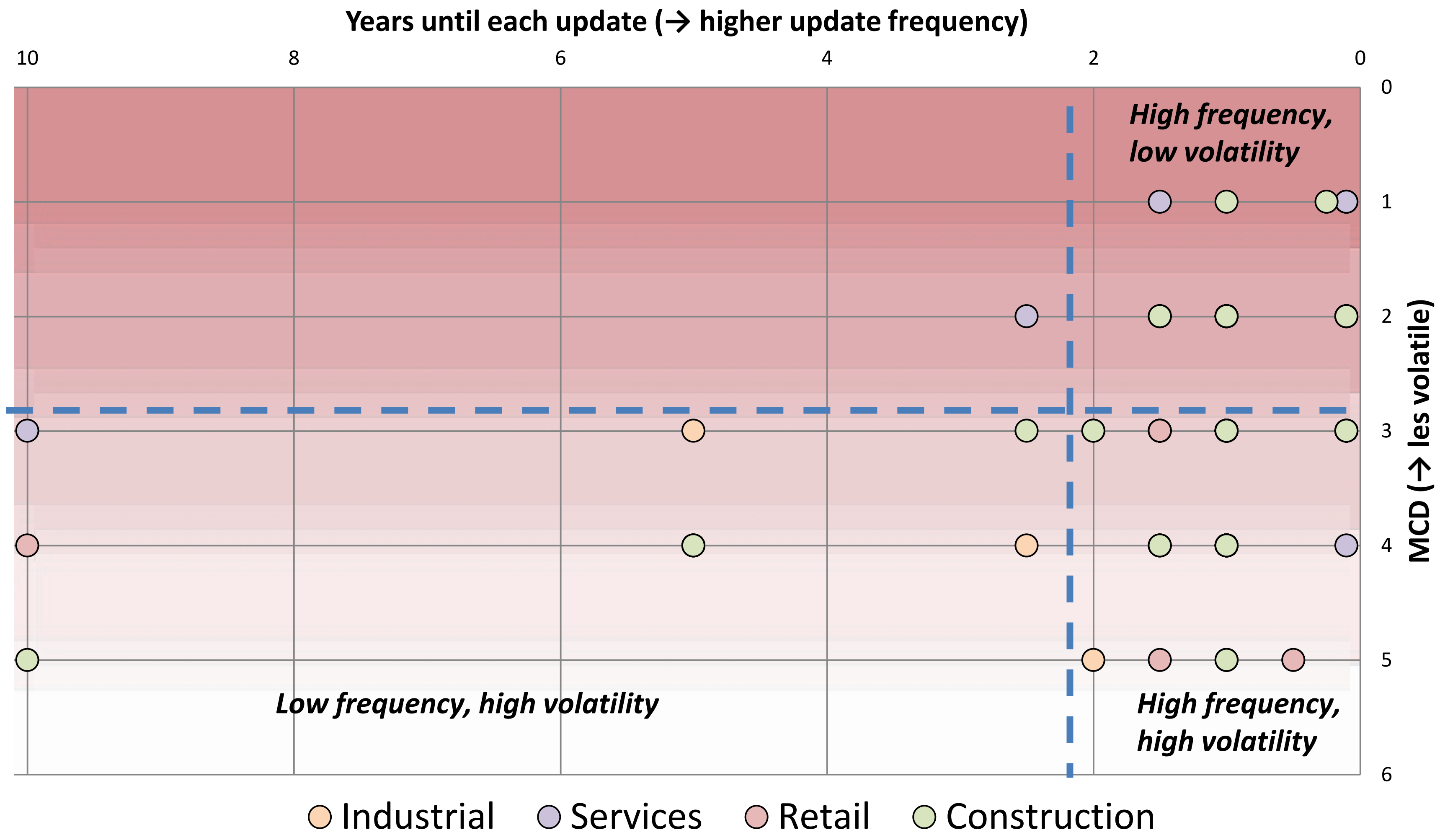
Higher updating frequency reduces volatility

- Descriptive statistics suggest that volatility does indeed decline with increased updating frequency

- Volatility doesn’t decline with increased frequency for the services and retail surveys, but this could be due to the small dataset

- But update frequency explains only a small part of the variability of MCDs

All surveys: months for cyclical dominance (MCD) vs frequency of updating



But updating frequency explains only a small part of the variation in volatility

- Averages mask substantial variation in effectiveness of updating frequency in reducing volatility

- Many high-frequency surveys have high volatility. Is high frequency a ‘necessary but not sufficient’ condition for low volatility?

- Omitted variables needed to explain remaining variation

Average volatility by frame size

Mean MCD by frame size as a % of population	
Frame size as % of population - band	**MCD- All Surveys**
Sample frame as % Population - up to 20	3.2
Sample frame as % Population 21-50	3.0
Sample frame as % Population 51-99	2.8
Sample frame as % Population 100	2.9
Total	3.0

Mean MCD by frame size as a % of population	
Absolute frame size	**MCD - All Surveys**
1-999	3.0
1,000-4,999	3.1
5,000-9,999	3.1
10,000-29,999	3.0
30,000-199,000	3.5
200,000+	2.4
Total	3.0

Frame size has a limited impact on volatility

- The relationship between frame size (absolute or % coverage) and volatility is not particularly strong

- A small subset of surveys with very large frame sizes (200,000+) do have a lower-than average MCD.

Omitted variables make rigorous statistical analysis difficult

- A simple OLS regression of MCD on sample coverage and update frequency produces coefficients with the ‘right sign’ but explains little of the variation (R^2=0.08)
- Attempt to use institute/frame type as instruments was unfruitful

OLS regression on MCD

Variable	Coefficient	P-Value
Constant	2.6	0.00
Frequency	-0.4	0.11
Coverage	0.2	0.01
R-squared	**0.08**	
Observations	**91**	

Average volatility by institute and frame type

Mean MCD by institute type

Institute Type	MCD - All Surveys
Academic	3.3
Business Association	3.0
Other Public bodies	2.5
Private Bodies	3.4
Statistical Institute	3.0
Total	**3.0**

Mean MCD by frame type

Frame Type	MCD- All Surveys
Bought List	4.3
Internally Compiled list	3.1
National Register	3.0
Private Register	3.2
Combination of Registers	2.9
Total	**3.0**

Analysis of correlations (tracking performance) across surveys

Analysis of correlation – initial hypotheses

- Frame size coverage likely to be an important factor in improving tracking performance.

- Absolute frame size in itself less likely to be an important factor

- Higher updating frequency likely to be positive, but importance unclear *a priori*

Average correlation by sample coverage

Mean correlation by sample frame as % of population

	Survey				
Frame size as % of population - band	**INDU**	**SERV**	**RETA**	**BUIL**	**All**
Sample frame as % Population - up to 20	68	72	60	63	**66**
Sample frame as % Population 21-50	60	61	66	62	**62**
Sample frame as % Population 51-99	72	89	67	60	**73**
Sample frame as % Population 100	70	81	71	87	**76**
Total	**66**	**74**	**64**	**66**	**68**

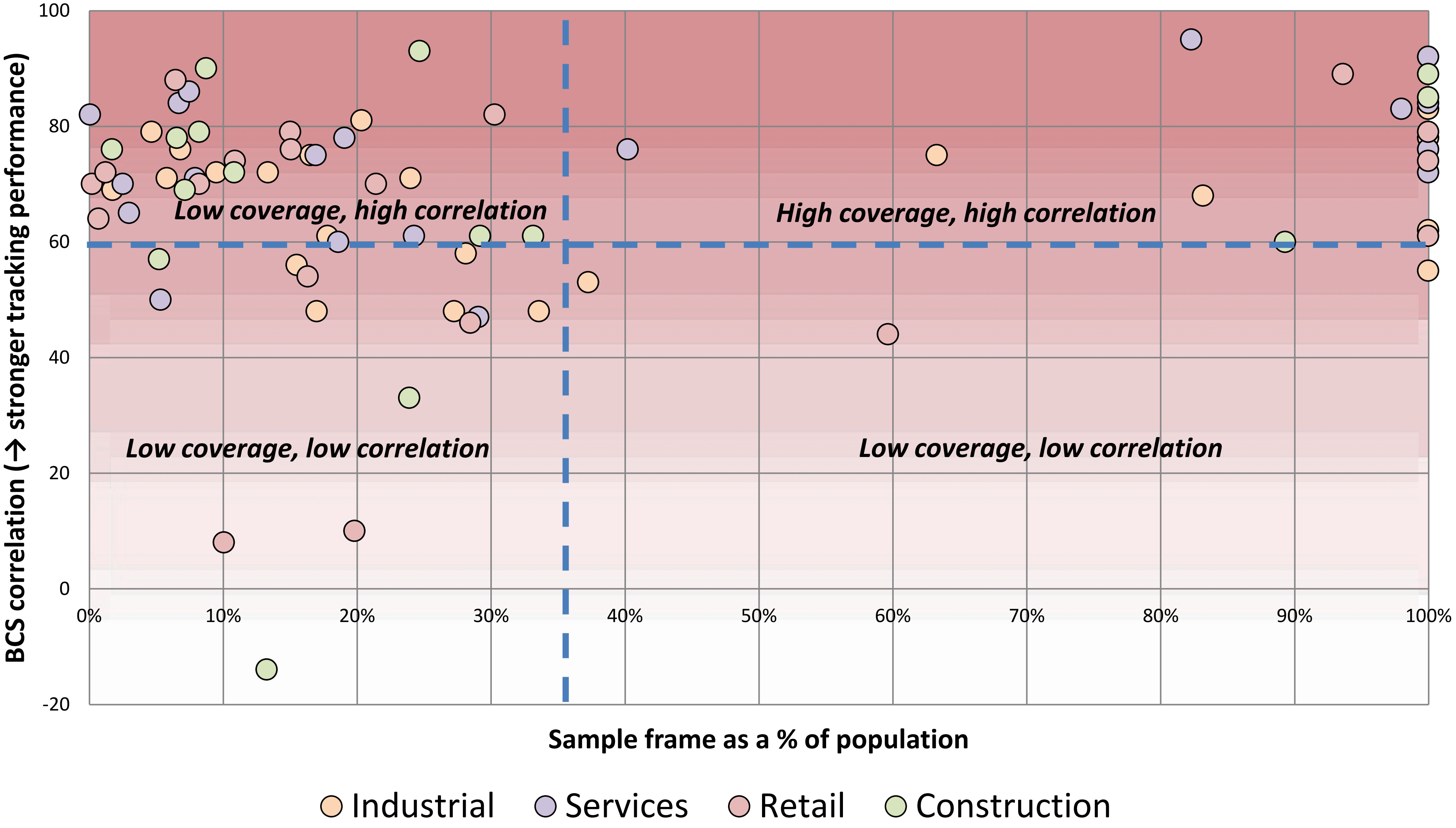
Higher frame coverage is associated with stronger tracking performance

- Frame coverage over 50% is associated with a somewhat higher correlation across all surveys

- Caution needed: only 20 surveys with a known correlation have a sample coverage above 50%

- No clear difference between surveys with 1-20% and 20-50% coverage

All surveys: BCS correlation vs sample frame as a % of population



But frame coverage explains only a small part of the variability in correlation

- Averages mask substantial variation in the tracking performance of surveys with a relatively low frame coverage

- Many low-coverage surveys have strong correlation. Is high coverage a sufficient but not necessary condition for strong tracking performance?

- Omitted variables needed to explain remaining variation

Average correlation by updating frequency

Mean correlation by frequency of updating

	Survey				
Updating frequency - band	INDU	SERV	RETA	BUIL	All
Monthly/ Continuously	65	51	44	59	**56**
Yearly	66	78	68	65	**69**
Interval over one year	63	71	53	68	**63**
Total	**65**	**72**	**60**	**64**	**65**

Relationship between updating frequency and tracking performance looks relatively weak

- Role of updating frequency in improving correlation unclear.

- Perhaps perversely, the highest-frequency surveys have *lower* correlations on average. (Only 15 surveys are in this category.)

As before, omitted variables make rigorous statistical analysis difficult

- A simple OLS regression of BCS on sample coverage and update frequency produces coefficients with the ‘right sign’ but explains none of the variation.
- Once again, attempt to use institute/frame type as instruments was unfruitful

OLS regression on BCS correlation

Variable	Coefficient	P-Value
Constant	58.9	0.00
Frequency	-0.9	0.62
Coverage	8.9	0.30
R-squared	**0.01**	
Observations	**91**	

Average correlation by institute and frame type

Mean BCS correlation by institute type	
Institute Type	**BCS - All Surveys**
Academic	57
Business Association	66
Other Public bodies	57
Private Bodies	73
Statistical Institute	67
Total	**65**

Mean BCS correlation by frame type	
Frame Type	**BCS - All Surveys**
Bought List	79
Internally Compiled list	68
National Register	66
Private Register	54
Combination of Registers	66
Total	**65**

Key conclusions

Key conclusions

- Links between institute and frame type and frame coverage/updating frequency make it difficult to dis-entangle their independent effects on volatility and tracking performance

- Frame size and updating frequency explain only a small part of the differences in volatility and tracking performance between surveys.

Key conclusions - volatility

- Updating frequency is a key determinant of survey volatility.

- However, it explains only part of the variation in survey MCDs – updating frequency can be thought of as ‘necessary, but not sufficient’ for low volatility

- Frame size appears to have a less influence – though a small sub-set of surveys with very large absolute frame sizes do have low volatility

Key conclusions – tracking performance

- Frame coverage over about 35% is associated with a stronger tracking performance

- Below 35%, the relationship is less clear. Frame coverage can be thought of as ‘sufficient, but not necessary’ for strong correlation

- The relationship with updating frequency looks relatively weak

Thank you for listening

- Grateful thanks to Christian Gayer and the DG Ecfin team during this taskforce work for their guidance, advice and support
- Also, grateful thanks to my colleagues Alan Joy, Daniel Lee and Christopher Taylor for the significant work enacted throughout this project.

CBI
THE VOICE OF BUSINESS