TASK FORCE ON 'QUALITY OF BCS DATA' THEMATIC GROUP 2: SAMPLE SIZE

Adamos ANDREOU, Nicoletta PASHOURTIDOU

Christian SEILER and Klaus WOHLRABE

JOINT EU/OECD WORKSHOP
ON RECENT DEVELOPMENTS IN BUSINESS AND CONSUMER SURVEYS
November 2013

Research Questions

- If and how is the quality of BCS data related to the number of particpants in a survey?
- Quality: MCD, Volatility, Forecasting accuracy …
- The larger the sample size, the better the quality.
- But: There are limits in practice!
- Is the gain in quality linear or nonlinear in sample size?
- Is there an optimal sample size?

Outline

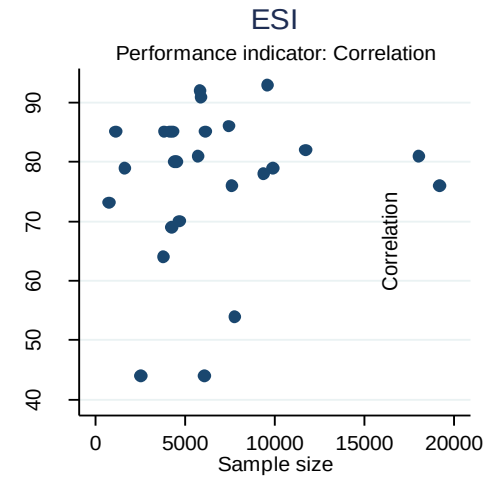
- Empirical Results using data from all EU countries
- Simulation Study

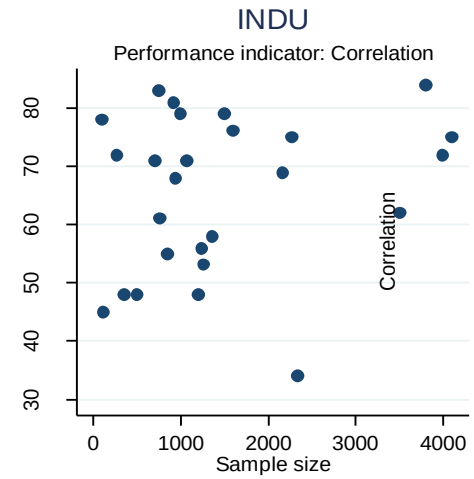
Empirical Analysis

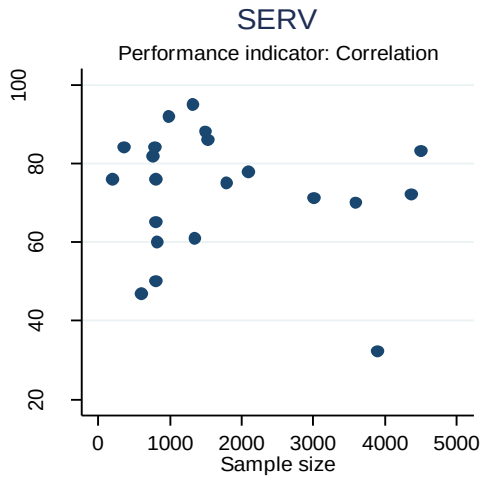
Empirical Results: Data

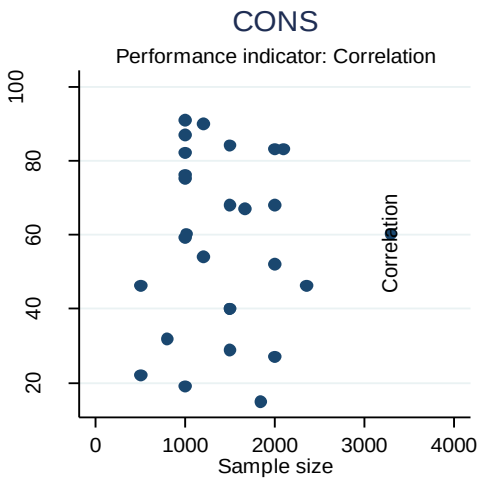
- Performance Indicators
  - Correlation
  - MCD (1,2)
  - Standard Deviation
  - Forecast Error
- Business Survey characteristics
  - (Effective) Sample size
  - Response Rate
  - Coverage
  - Survey Method
  - Weighting Scheme
  - Stratum Level
- Industry, Services, Construction, Retail

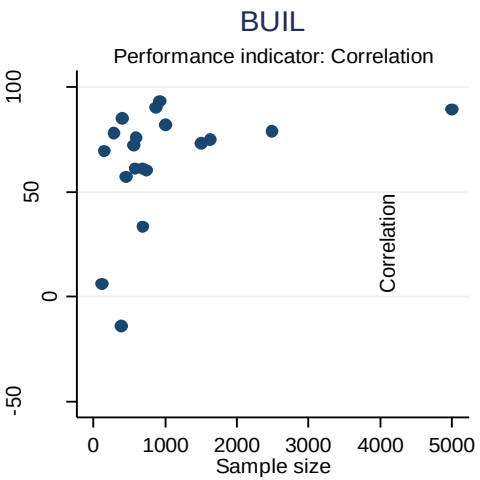
Empirical Results: Correlation Results (Tracking Performeance)

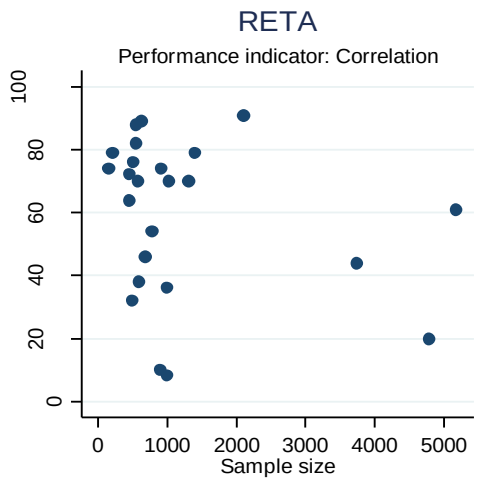




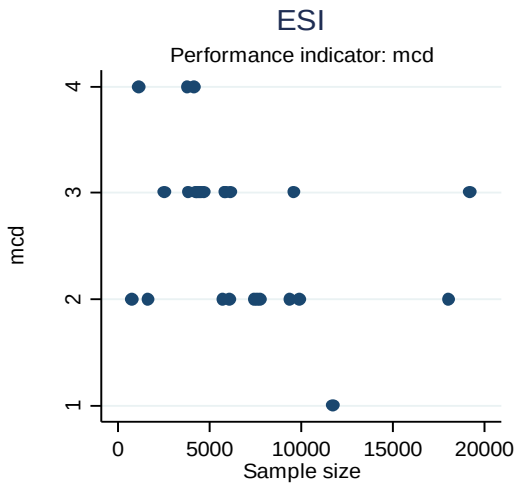


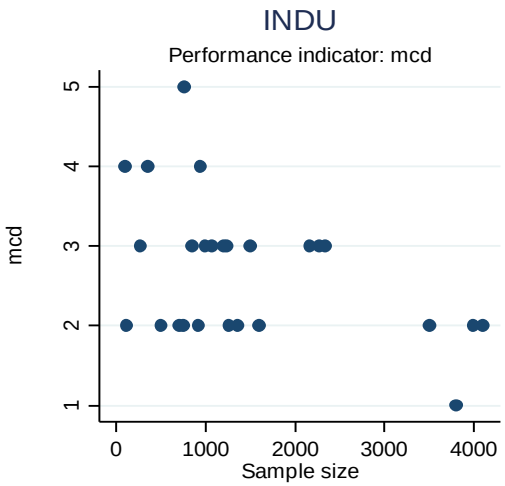


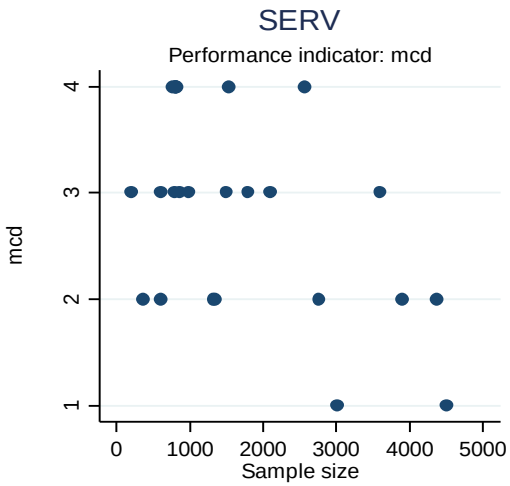


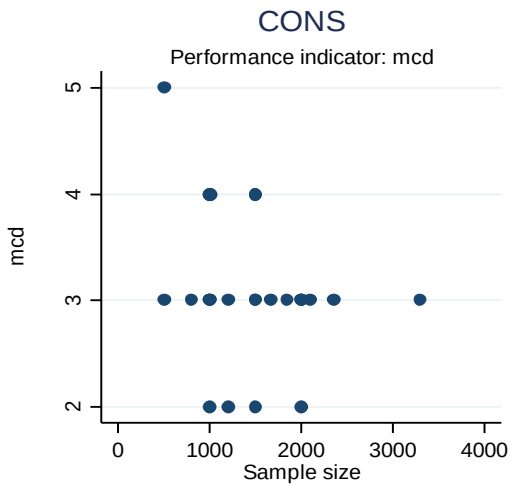


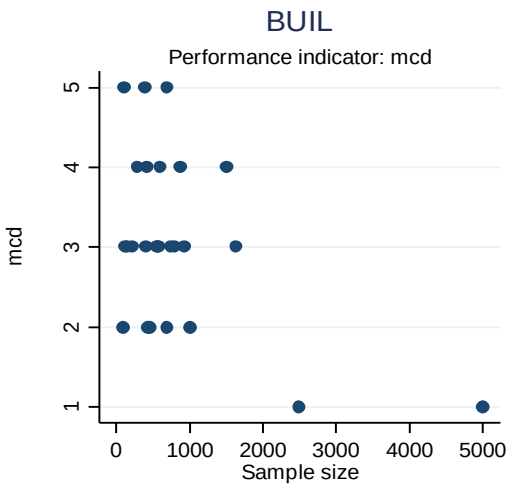
Empirical Results: Correlation Results (MCD)

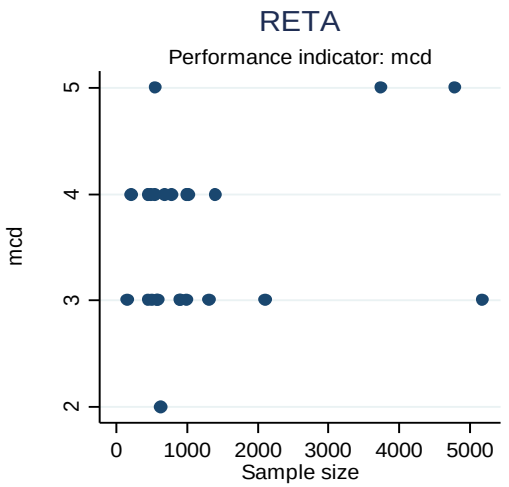




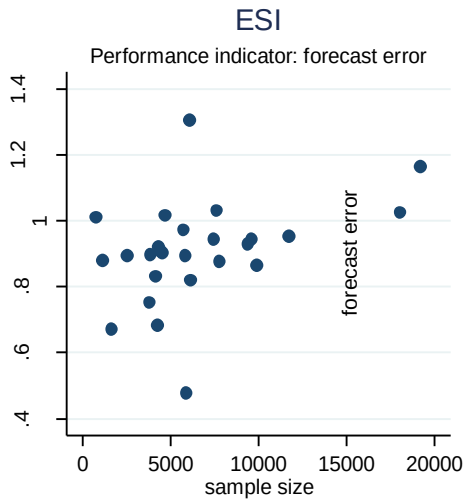


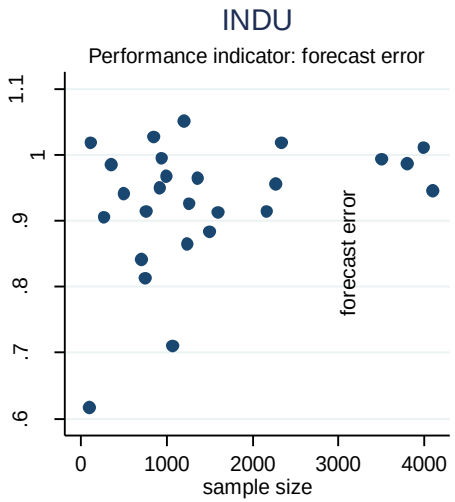


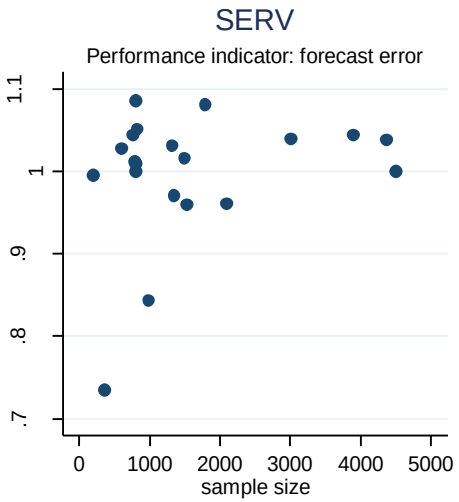


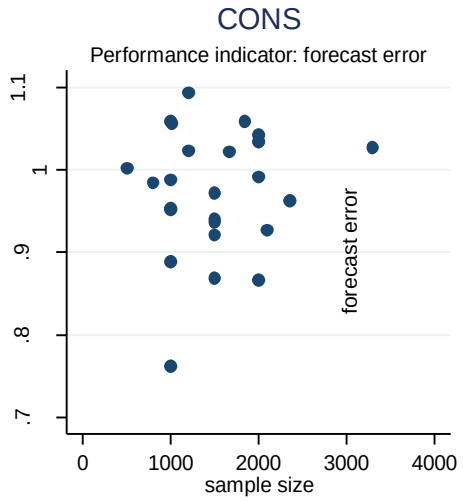


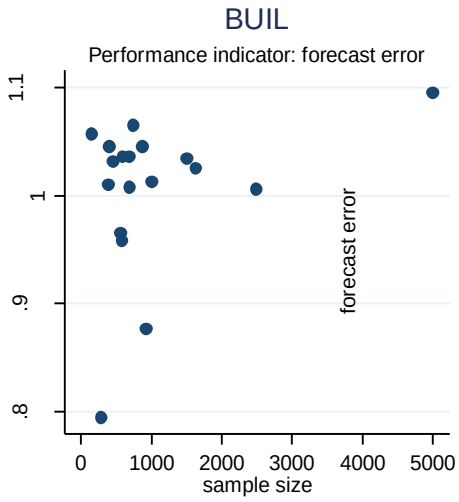
Empirical Results: Correlation Results (Forecasting Accuracy)

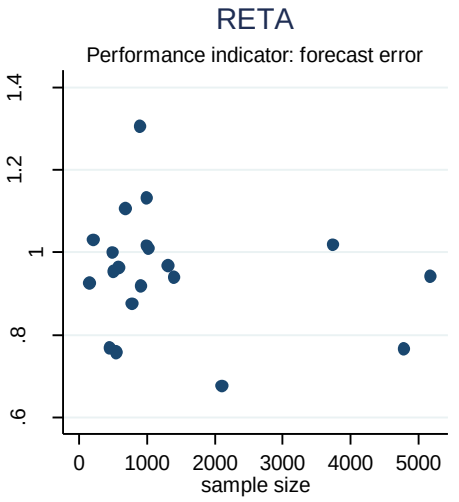












Empirical Results: Correlation Results (*Sample Size*)

	ESI	INDU	SERV	BUIL	RETA	CONS
Correlation	0.1282	0.2127	-0.1564	0.3674	-0.2472	0.0223
	[0.5325]	[0.2969]	[0.4983]	[0.1218]	[0.2443]	[0.9119]
MCD	**-0.3585**	**-0.4426**	**-0.5048**	**-0.4851**	0.2732	-0.2978
	[0.0721]	[0.0236]	[0.0101]	[0.0120]	[0.1964]	[0.1313]
Standard deviation	0.0290	-0.2987	**-0.4486**	-0.2144	**-0.4760**	-0.2138
	[0.8883]	[0.1383]	[0.0245]	[0.2929]	[0.0187]	[0.2944]
Forecast error	**0.3872**	**0.3313**	0.2475	0.3328	-0.1623	0.1005
	[0.0506]	[0.0982]	[0.2947]	[0.1772]	[0.4705]	[0.6326]

Empirical Results: Correlation Results (*Effective Sample Size*)

	ESI	INDU	SERV	BUIL	RETA	CONS
Correlation	0.0555	0.0789	-0.1848	0.3191	-0.0767	-0.1204
	[0.7876]	[0.7016]	[0.4225]	[0.1829]	[0.7215]	[0.5497]
MCD	-0.2385	**-0.4496**	**-0.4572**	**-0.4901**	0.0324	**-0.4077**
	[0.2406]	[0.0212]	[0.0216]	[0.0110]	[0.8806]	[0.0348]
Standard deviation	-0.0339	**-0.3647**	**-0.4737**	-0.1908	**-0.4123**	-0.0621
	[0.8696]	[0.0670]	[0.0167]	[0.3506]	[0.0453]	[0.7630]
Forecast error	**0.4330**	**0.3337**	0.1606	0.3643	0.0034	**0.3566**
	[0.0271]	[0.0957]	[0.4984]	[0.1372]	[0.9880]	[0.0801]

Empirical Results: Correlation Results

- The correlation of the BCS indicators with the reference series is not significantly correlated with either the sample size or the effective sample size.
- MCD is negatively and significantly correlated with the sample size in the case of ESI, INDU, SERV and BUIL.
- MCD is negatively and significantly correlated with the effective sample size in the case of all business survey indicators except for RETA.
- The standard deviation of RETA is negatively correlated with both the sample size and the effective sample size. A negative correlation between standard deviation and effective sample is also found in the case of SERV and INDU.
- MCD computed from CONS are negatively correlated with the effective sample size, but none of the volatility measures is correlated with the sample size.

Empirical Results: Regression Analysis MCD

Table (Specification)	SAMPLE / EFFECTIVE SAMPLE					RESPONSE RATE	COVERAGE						WEIGHTING SCHEME											SAMPLING FRAME	SURVEY METHOD				
							Employment	Turnover					Employment	Output	Firm				Stratum				Age/gender	Region		Online /email	Fax	Post	Other	
Survey	INDU	SERV	BUIL	RETA	CONS	ALL	BS	INDU	SERV	BUIL	RETA	CONS	BS	BS	INDU	SERV	BUIL	RETA	INDU	SERV	BUIL	RETA	CONS	CONS	ALL	ALL	BS	BS	BS	CONS
6(1)	-	-	-	+	-	-																								
	1	1	1	0	1	1																								
6(2)	-	-	-	+	-	-	-	+	+	-	-	+																		
	1	1	1	0	1	0	0	1	0	0	0	1																		
6(3)	-	-	-	+	-	-							+	+									+	-						
	1	1	1	0	1	0							0	1									0	1						
6(4)	-	-	-	+	-	-	+	+	+	-	-	+	+	+									+	-						
	1	1	1	1	1	0	0	0	0	0	1	1	0	1									0	1						
6(5)	-	-	-	+	-	-	+	-	-	-	-	+	+	+									+	-	+	-	+	+	+	-
	1	1	1	0	1	0	0	0	0	0	0	1	0	1									0	1	1	1	0	0	1	1
7(1)	-	-	-	+	-	-									-	+	-	+	+	-	+	-	+	-						
	1	1	1	0	1	0									1	1	0	1	1	0	0	0	0	1						
7(2)	-	-	-	+	-	-	-	-	-	-	-	+			-	+	-	+	+	-	+	-	+	-						
	1	1	1	0	1	0	0	0	0	0	0	1			1	1	0	1	1	0	0	0	0	1						
7(3)	-	-	-	+	-	-	+	+	+	+	+	+			-	+	-	+	+	-	+	-	+	-	+	-	+	+	+	-
	1	1	1	0	1	0	0	0	0	0	0	1			0	1	0	1	0	1	0	0	1	1	1	1	0	0	1	0
8(1)	-	-	-	+	-																									
	1	1	1	0	1																									
8(2)	-	-	-	+	-		-	+	+	-	-	+																		
	1	1	1	0	1		0	1	0	0	0	1																		
8(3)	-	-	-	+	-								+	+									+	-						
	1	1	1	0	1								0	1									0	1						
8(4)	-	-	-	+	-		+	+	+	-	-	+	+	+									+	-						
	1	1	1	0	1		0	0	0	0	0	1	0	1									0	1						
8(5)	-	-	-	+	-		+	-	-	-	-	+	+	+									-	-	+	-	+	+	+	-
	1	1	1	0	1		0	0	0	0	0	1	0	1									1	1	1	1	0	0	1	1
9(1)	-	-	-	+	-										-	+	-	+	+	-	+	-	+	-						
	1	1	1	0	1										1	1	0	1	1	0	0	0	0	0						
9(2)	-	-	-	-	-		+	-	-	-	-	+			-	+	-	+	+	-	+	-	+	-						
	1	1	1	0	1		0	0	0	0	0	1			1	1	0	1	1	0	0	0	0	1						
9(3)	-	-	-	+	-		+	+	+	+	+	+			-	+	-	+	+	-	+	-	-	-	+	-	+	+	+	-
	1	1	1	0	1		0	0	0	0	0	1			0	1	0	1	0	1	0	0	1	1	1	1	0	0	1	1

Empirical Results: Regression Analysis
Forecasting Results

Table (Specification)	SAMPLE / EFFECTIVE SAMPLE					RESPONSE RATE	COVERAGE							WEIGHTING SCHEME											SAMPLING FRAME	SURVEY METHOD				
							Employment	Turnover						Employment	Output	Firm				Stratum				Age/gender	Region		Online/email	Fax	Post	Other	
Survey	INDU	SERV	BUIL	RETA	CONS	ALL	BS	INDU	SERV	BUIL	RETA	CONS	BS	BS	INDU	SERV	BUIL	RETA	INDU	SERV	BUIL	RETA	CONS	CONS	ALL	ALL	BS	BS	BS	CONS
11(1)	+	+	+	+	+	+	+	+																						
	1	0	1	0	0	0	0	0																						
11(2)	+	+	+	+	-	-	+	+	+	-	-	+	+	-	+															
	1	0	1	1	1	1	0	1	1	0	0	1	1	0	1															
11(3)	+	+	+	+				+								+		+	+											
	1	0	1	1	0	0	0	1								1	0	1	0							0	0			
11(4)	+	+	+	-	-	-	-	-	+	-	-	+	+	+	+	+	+	+	+							-	+			
	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0							1	1			
11(5)	+		+						+	+	+	+	+	+		+	+	+	+								+			
	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							1	0	1	1	0
12(1)	+	+	+	+	+	-	+	+												+	+	+	-	-	-	-	-			
	1	0	1	1	1	0	1	1												0	0	0	0	0	0	0	0			
12(2)	+	-	+	-	-			-	+	-	+	+	+	-	-					+	+	+	-	-	-	-	+			
	1	0	1	1	0	0	0	0	1	0	1	1	1	0	0					1	0	1	0	0	0	1	1			
12(3)	+	-	+	-	-	-	-	-	+	-	+	+	+	-	-					+	+	+	+	+	+	-	+	-	+	-
	1	0	1	1	1	1	1	1	1	0	1	0	0	0	1					1	0	1	0	0	0	1	1	1	1	0
13(1)	+	+	+	+																										
	1	0	1	1																										
13(2)	+	+	+	+					+	+	+	+	+	+	+															
	1	1	1	1					0	0	0	1	1	1	0															
13(3)	+	-	+	+												+	-	+	+							-	-			
	1	0	1	1												1	0	1	0							0	0			
13(4)	+	-	+	+					+	+	+	+	+	+	+	+	+	+	+							-	+			
	1	0	1	1					0	0	0	0	0	0	0	0	0	0	0							0	0			
13(5)	+	-	+	-					+	+	+	+	+	+	-	+	+	+	+							-	+		+	-
	1	0	1	1					0	0	0	0	0	0	0	0	0	0	0							1	0	1	0	0
14(1)	+	+	+	+																+	+	+	-	-	-	-	-			
	1	0	1	1																0	0	0	0	0	0	0	0			
14(2)	+		+	+					+	+	+	+	+	+	+	+	+	+	+	+	+	+					+			
	1	0	1	1					1	0	1	1	1	0	0					1	0	1	0	0	0	0	0			
14(3)	+	-	+	-					+	+	+	+	+	+	-					+	+	+	+	+	+	-	+	-	+	-
	1	0	1	1					0	0	0	0	0	0	0					0	0	0	0	0	0	1	0	1	1	1

Empirical Results: Regression Results (MCD)

- The sample size influences negatively and significantly MCD in all surveys except RETA; nevertheless, the marginal effect (about -0.0004) is rather small e.g. an increase in the sample size by 200 units is estimated to decrease MCD by about 0.08 months.
- The effective sample size has a negative and significant impact on MCD in all surveys except RETA. The marginal effect of the effective sample size on MCD is similar to that of the sample size and varies between -0.0004 and -0.0005.
- There is evidence of a negative but statistically insignificant effect of the response rate on MCD.
- Coverage in terms of employment does not affect significantly MCD. No persistently significant relation is found between coverage in terms of turnover and MCD. Coverage in consumer survey is positively associated with MCD.

Empirical Results: Regression Results (Forecasting)

- Larger sample size or effective sample size is found to increase forecast error in the case of INDU and BUIL. The sample size and effective sample size does not influence the forecast error in SERV. The results for CONS are rather mixed.
- The inclusion of the response rate does not result in a significant effect on the forecast error with a consistent direction.
- The evidence regarding the relation of coverage and forecast errors is rather weak.
- Remember: Simple forecasting model
- Future research necessary

Simulation Study

Structure

- Survey participation model
- Surveyed business cycles („Target variables“)
- Simulation set up
- Results
- Real data simulation based on Ifo micro data

Participation Model

For our simulation study, we assume that every respondent i is affected in his or her opinion formation by the business cycle C_t and an individual error term $I_{i,t}$. Then,

$$y_{i,t}^{*} = C_t + I_{i,t}$$

with $I_{i,t} \sim N(0, \sigma^2)$.

As each survey participant is restricted to give answers on a 3-level scale, we observe

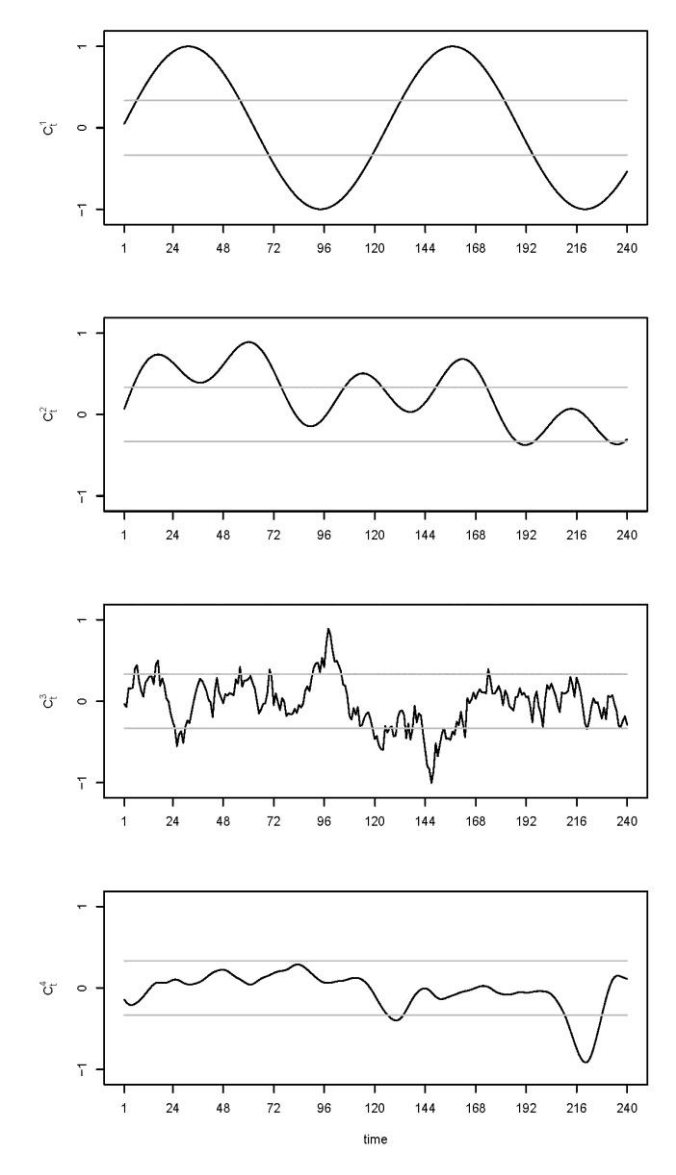
$$y_{i,t}^{*} = \begin{cases} + & if \quad y_{i,t}^{*} > \tau^{+} \\ = & if \quad \tau^{+} \geq y_{i,t}^{*} > \tau^{-} \\ - & if \quad \tau^{-} \geq y_{i,t}^{*} \end{cases}$$

Potential caveat: no firm decison as in the business surveys

Simulation Set Up

- Four different types of business cycle functions
- All functions are scaled to range between -1 and 1
- Thresholds are defined by $\tau^{+} = \frac{1}{3}$ and $\tau^{-} = -\frac{1}{3}$
- For the irregular component, we assume $I_t \sim N(0, \sigma^2)$.

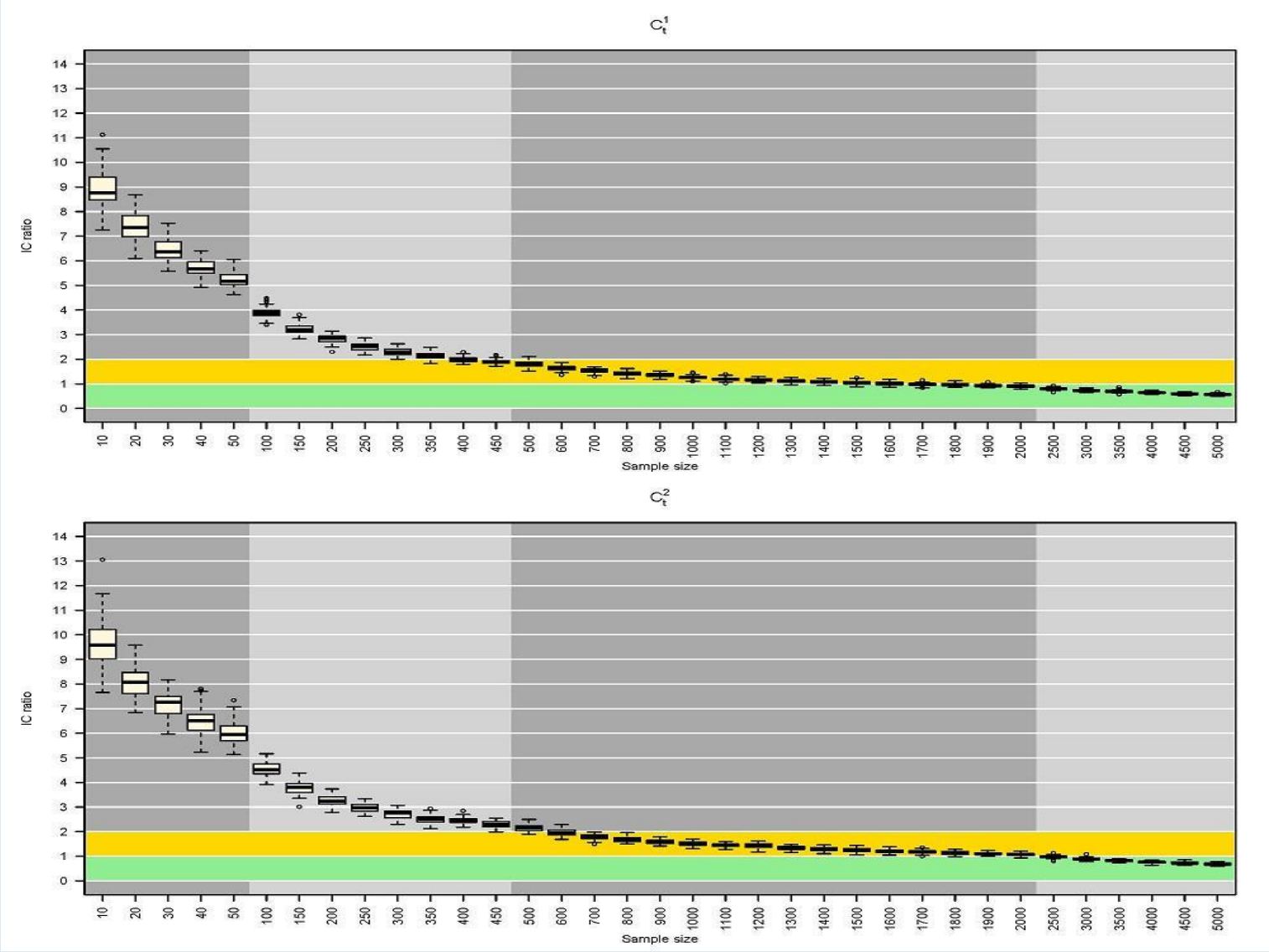
Investigated Business Cycles



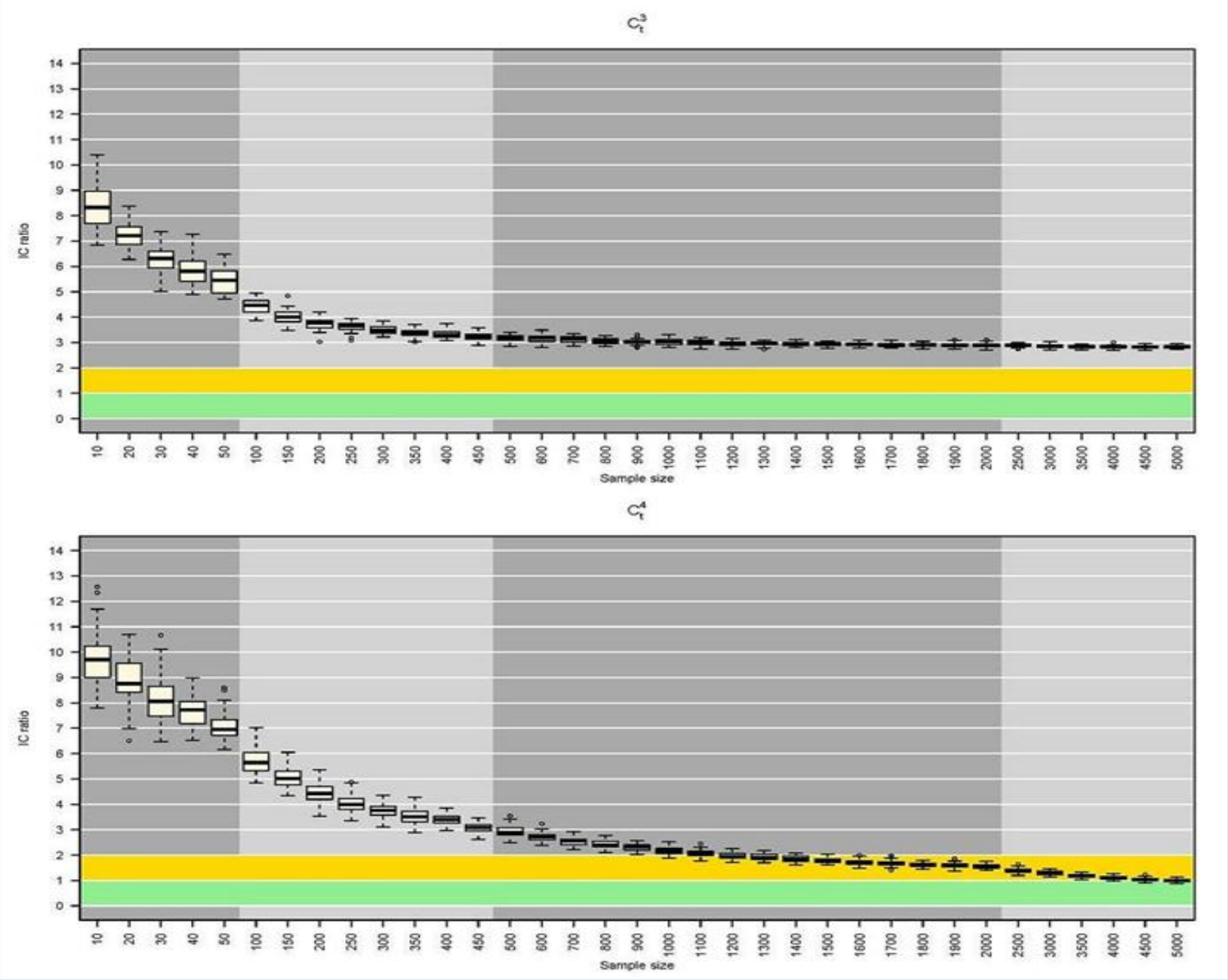
Simulation Set Up

- Step 1: Calculate response probabilities $P(y_{i,t} = '=' \mid C_t, t)$, $P(y_{i,t} = '+' \mid C_t, t)$ and $P(y_{i,t} = '-' \mid C_t, t)$ for cycle function C_t^j.
- Step 2: Draw n observations given the calculated probabilities in Step 1.
- Step 3: Build the balance statistics from the drawn responses, extract the components C_t and I_t using the Hodrick-Prescott filter and calculate the IC ratio.
- Step 4: Repeat Steps 2 and 3 $m = 50$ times.
- Step 5: Repeat Steps 1-4 for each combination of $j = 1, .., 4$, $T = 120, 240, 360$ and $n = 10, 20, \ldots, 50, 100, 150, \ldots, 450, 500, 600, \ldots, 1900, 2000, 2500, \ldots, 5000$.

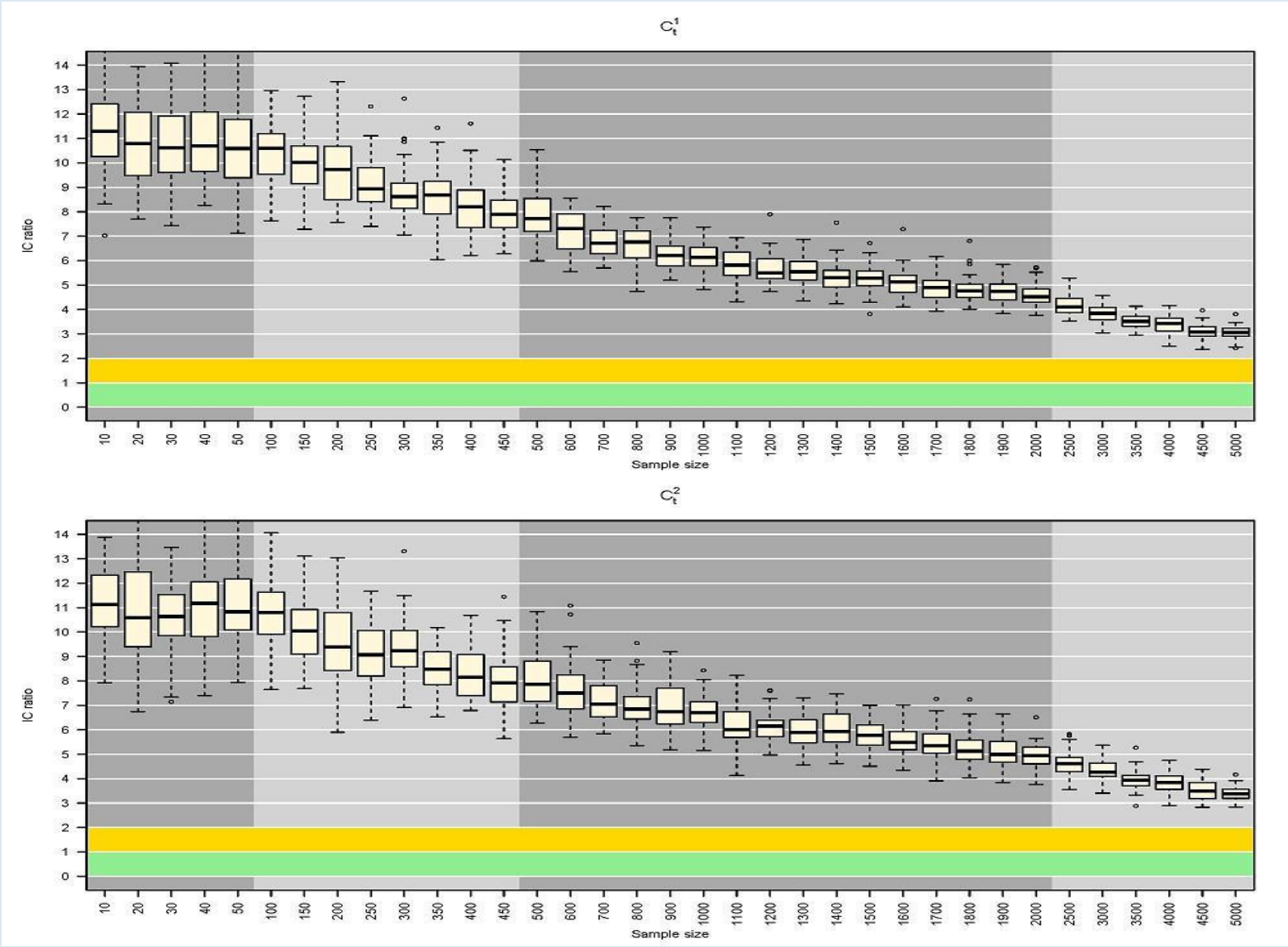
Results: ($T = 240, \sigma^2 = 1$)



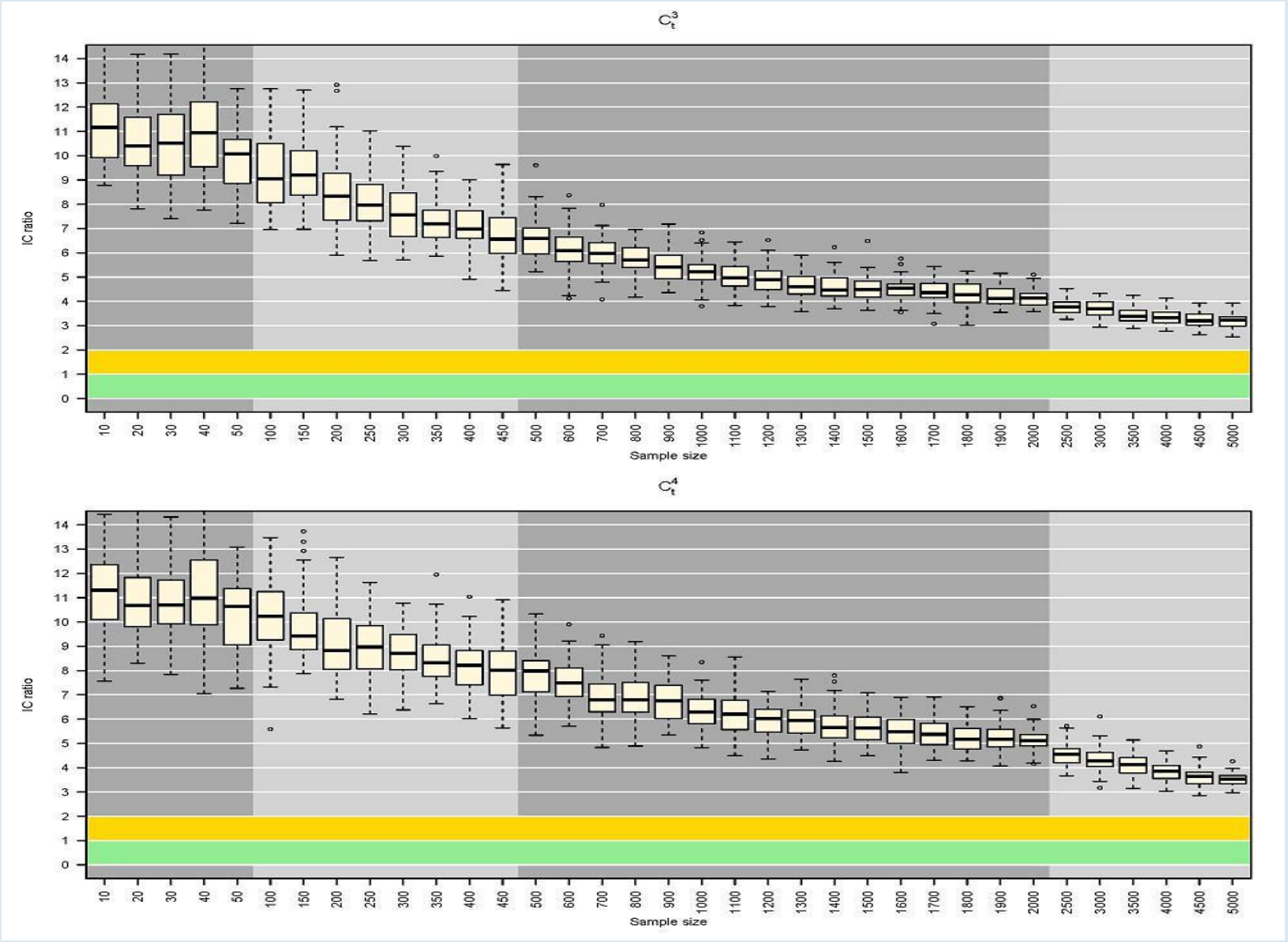
Results: ($T = 240, \sigma^2 = 1$)



Results: ($T = 120, \sigma^2 = 5$)



Results: ($T = 120, \sigma^2 = 5$)



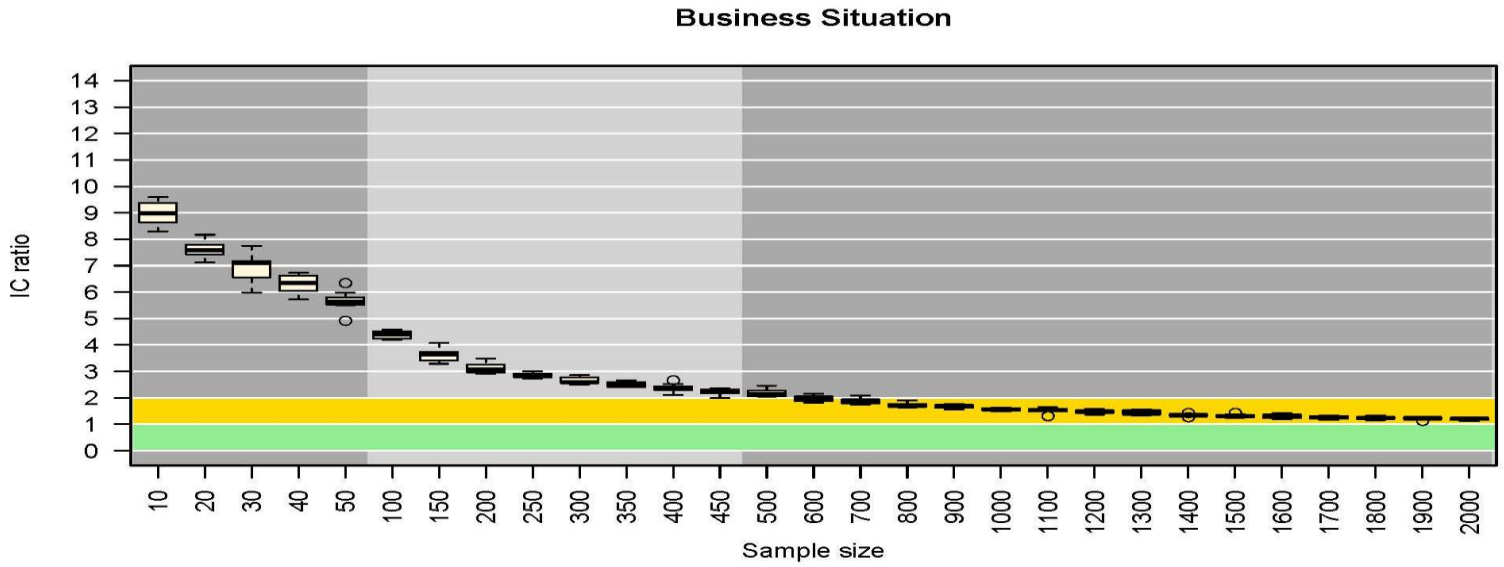
Results: Summary

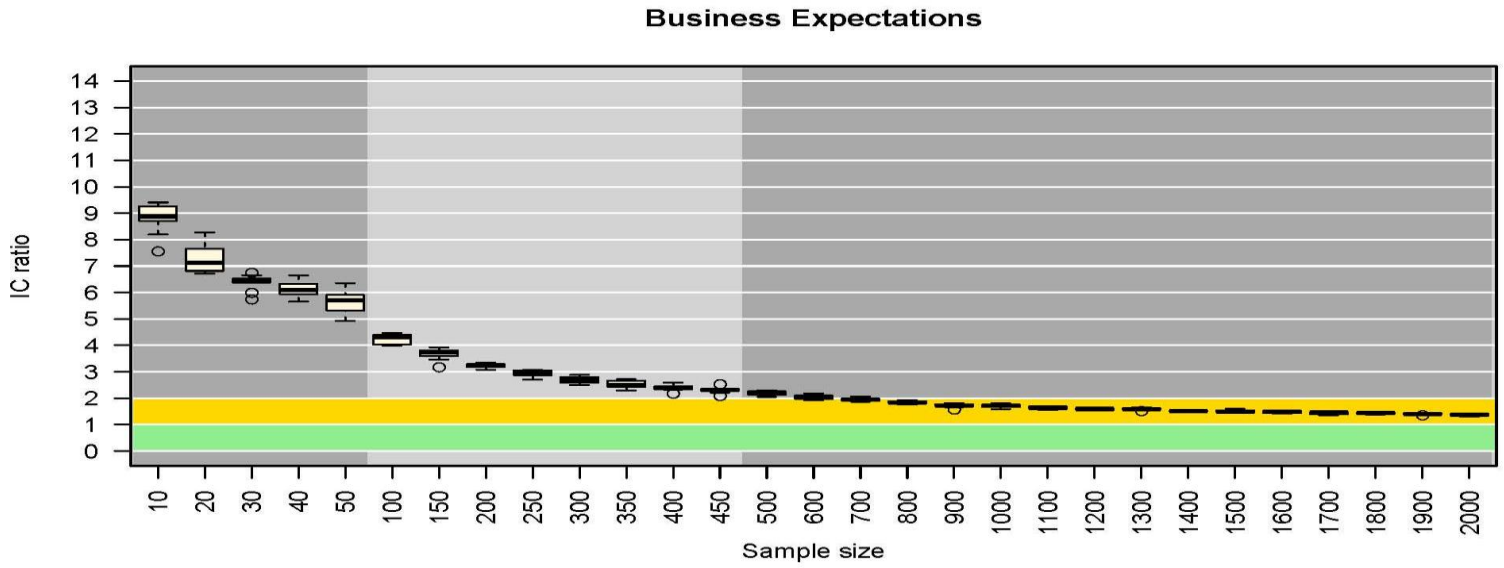
	C_t^1 and IC <2			C_t^1 and IC <1		
	T=120	**T=240**	**T=360**	**T=120**	**T=240**	**T=360**
$\sigma^2 = 0.5$	150	150	150	600	500	500
$\sigma^2 = 1$	600	600	450	2500	1900	2500
$\sigma^2 = 5$	-	-	-	-	-	-
	C_t^2 and IC <2			C_t^2 and IC <1		
	T=120	**T=240**	**T=360**	**T=120**	**T=240**	**T=360**
$\sigma^2 = 0.5$	200	200	150	250	800	600
$\sigma^2 = 1$	700	700	600	3500	3500	3000
$\sigma^2 = 5$	-	-	-	-	-	-
	C_t^3 and IC <2			C_t^3 and IC <1		
	T=120	**T=240**	**T=360**	**T=120**	**T=240**	**T=360**
$\sigma^2 = 0.5$	-	-	-	-	-	-
$\sigma^2 = 1$	-	-	-	-	-	-
$\sigma^2 = 5$	-	-	-	-	-	-
	C_t^4 and IC <2			C_t^4 and IC <1		
	T=120	**T=240**	**T=360**	**T=120**	**T=240**	**T=360**
$\sigma^2 = 0.5$	250	400	300	1000	1400	1200
$\sigma^2 = 1$	800	1500	1100	4500	5000	5000
$\sigma^2 = 5$	-	-	-	-	-	-

Results: Real Data Example

- Ifo Bussines Survey
- 1980-2012
- Draw different sample sizes and calculate the balance statistic
- Both for the situation and expectation question
- What is the effect on the IC?

Results: Real Data Example





Simulation Results: Summary

- Rather large improvements in the MCD are reached when the sample size is increased from a low level.
- The IC ratio approximately decreases by $\frac{1}{\sqrt{n}}$
- Thus enhancements die out as the sample size increases.
- The larger the uncertainty of the survey respondents about the underlying variable, the larger the sufficient sample size.
- The lesser noisy and volatile the target variable the lesser participating firms are necessary.

Summary

Summary

- Sample size matters!
- It improves the MCD and reduces volatility.
- The effect on forecasting performance is inclusive, further research is necessary.
- But the positive effect dies out with an increasing sample size
- If the target variable is “well-structured” and the uncertainty of the firms is low than 200 firms might be sufficient