International Total Survey Error Workshop
University of Bergamo, June 10-12, 2019

The Total Survey Error
A Time Series and System Approach

Anders Wallgren & Britt Wallgren

ba.statistik@telia.com

Formerly of Statistics Sweden – Department for research and development

Chart 4a. *Is the series going up or down?*

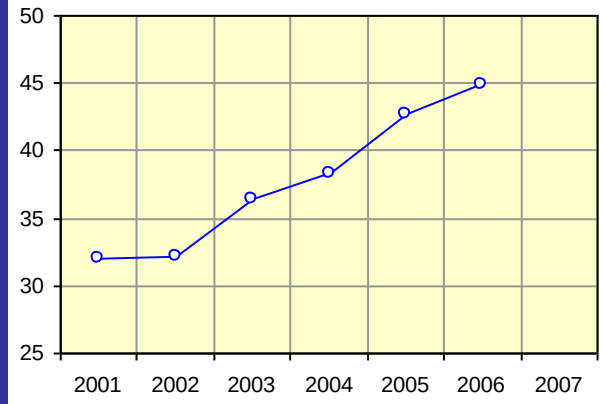


Chart 4b. *It is going down!*

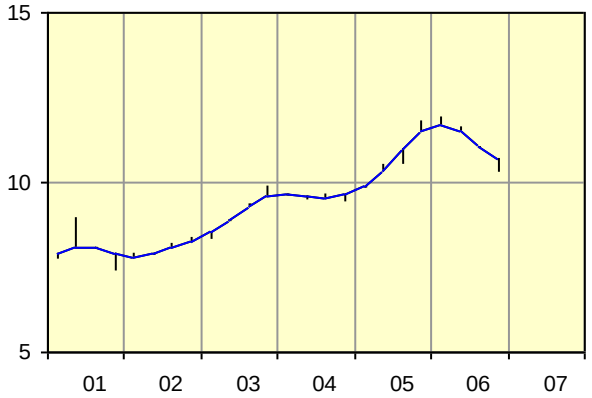


Chart 5a. *Is the series going up or down?*

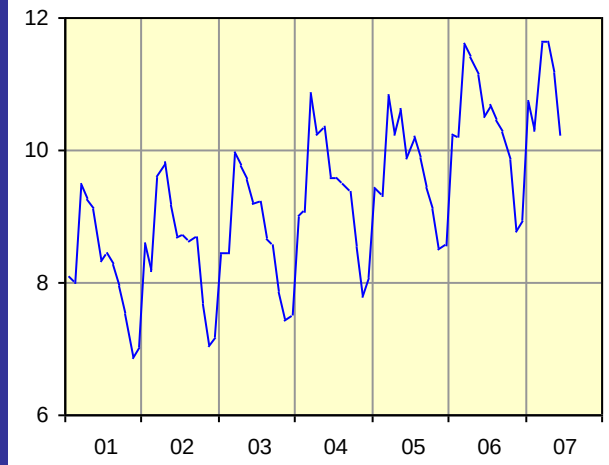


Chart 5b. *It is going down!*

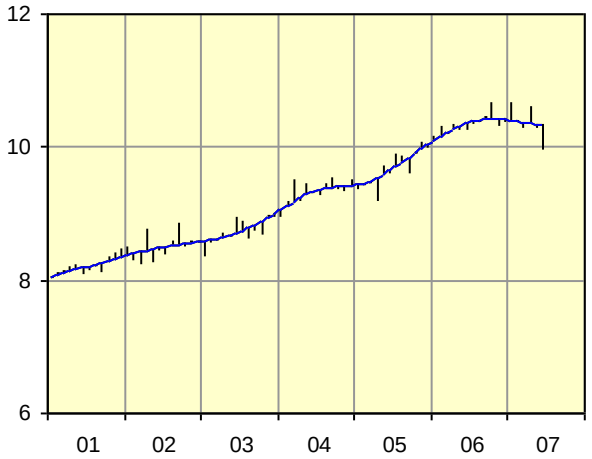


Chart 6a. *Is the series going up or down?*

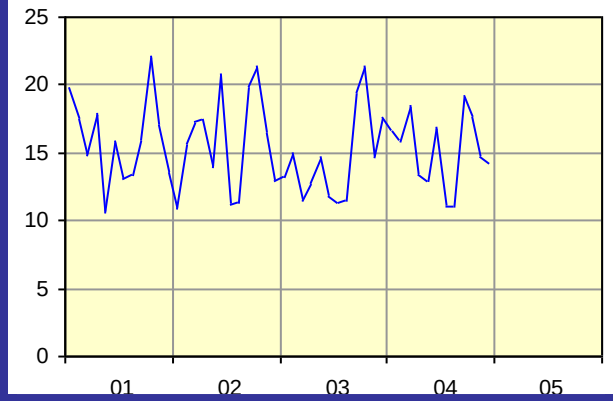


Chart 6b. *It goes down for the moment, but is uncertain*

- All sample surveys for official statistics generate time series
- The users need times series
- We must estimate time series patterns: $\hat{T}_t + \hat{S}_t + \hat{E}_t$

First we do a sample survey:

$\hat{Y} = \Sigma microdata \cdot w_i \quad \leftarrow$ *errors?* *Standard error = s($\hat{Y}$)*

Then we analyze time series based on that survey:

$\hat{Y}_t = y_t = \hat{T}_t + \hat{S}_t + \hat{E}_t \quad \leftarrow$ *errors?* *Standard deviation: s($\hat{E}$)*

The variable t is added.

Sampling and time series: Two paradigms, two theoretical areas

Errors/mistakes from the first sample survey phase creates errors during the second phase

Errors found during the second phase can reveal errors done during the first phase

$\Rightarrow$ *Sample design and time series analysis must be integrated!*

Interest and understanding of time series methods is very low

No integration of sample design and time series methods

Mistakes and misunderstandings regarding time series are common

Sampling errors dominate thinking

Lack of TSE-thinking, other errors are not observed/understood

Three examples:

1. *“Sample size is to small, can’t publish monthly data”*
2. *“Panel design is good”*
3. *“Errors in the time variable, what is that?”*

1. "Sample size is to small, can't publish monthly data"

"Instead we publish aggregate estimates for rolling three-month periods"

Not established times series analysis method
Sampling errors dominate thinking

Oct-Dec 2017
Nov-Jan 2018
Dec-Feb 2018
Jan-Mar 2018
Feb-Apr 2018
Mar-May 2018
Apr-Jun 2018
May-Jul 2018
Jun-Aug 2018
Jul-Sep 2018
Aug-Oct 2018
Sep-Nov 2018
Oct-Dec 2018

Chart 3. Standard errors according to sampling theory for some estimates in the Swedish LFS.

Both sexes, 15-74 years	Point estimate December 2018	Standard error Monthly estimate	Standard error Quarterly estimate	Standard error Yearly estimate
Employed, 1000s	5098.6	19.4	11.4	9.1
Unemployed, 1000s	326.9	11.5	6.3	4.0
Not in labour force, 1000s	2059.5	19.4	11.1	8.9
Hours worked, millions/week	148.1	0.95	0.6	0.5

TSE (monthly estimates) > TSE (quarterly estimates) > TSE (yearly estimates)

We think that this ranking reflects the general attitude among sample survey statisticians. According to the time series paradigm this ranking is wrong

1. “Sample size is to small, can’t publish monthly data”

- This old and deeply rooted practice indicates cross-section thinking:

  Standard errors of point estimates for *one* quarterly survey
  are compared with the
  standard errors of point estimates for *one* monthly survey.

  Compare instead *one* quarterly estimate with *three* monthly!

- But we don’t need point estimates, we need estimates of patterns

  The quality of the estimated time series patterns are at least as good with monthly data as with quarterly.

  120 monthly values, better than 40 quarterly values, better than 10 yearly values

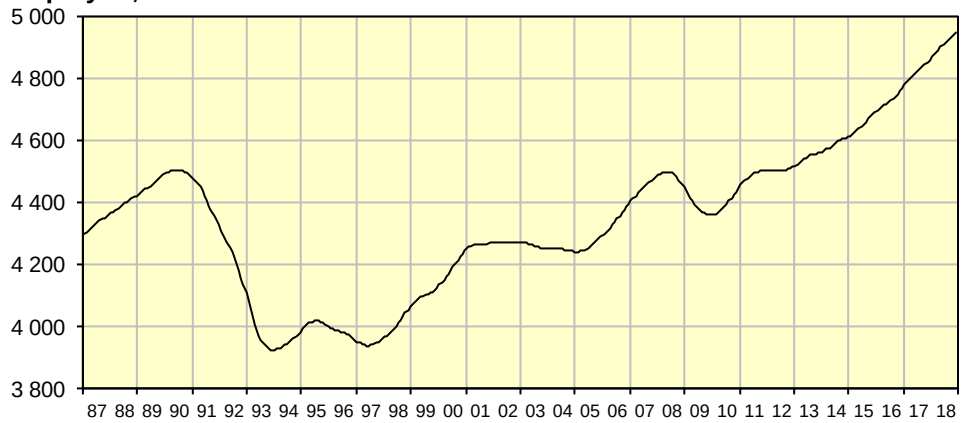
- The rolling three-months estimate, seasonally adjusted, is a 3-point moving average, a rough and primitive estimate of the trend component T_t

  In e.g. X13-ARIMA there are better trend estimators.

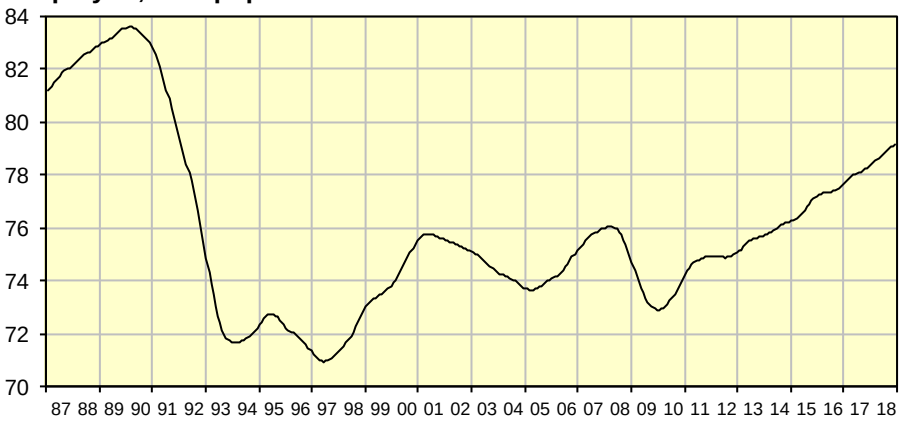
The Swedish Labor Force Survey, estimated trends, monthly data

Both sexes, 16-64 years

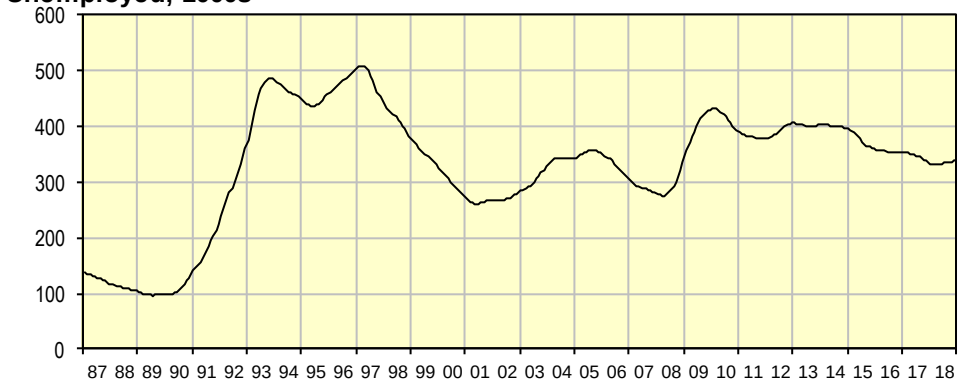
Employed, 1000s



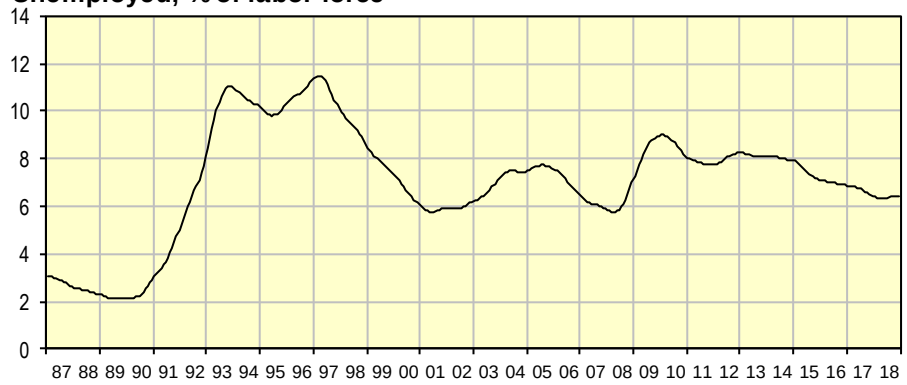
Employed, % of population



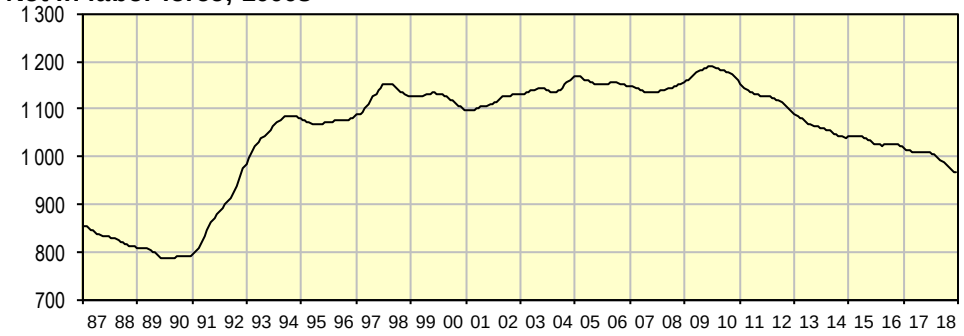
Unemployed, 1000s



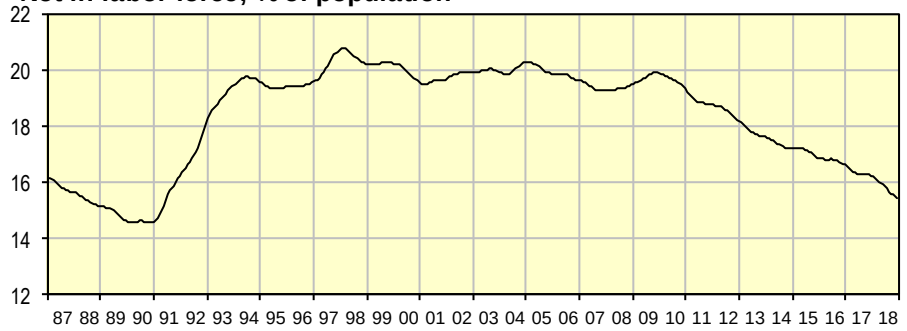
Unemployed, % of labor force



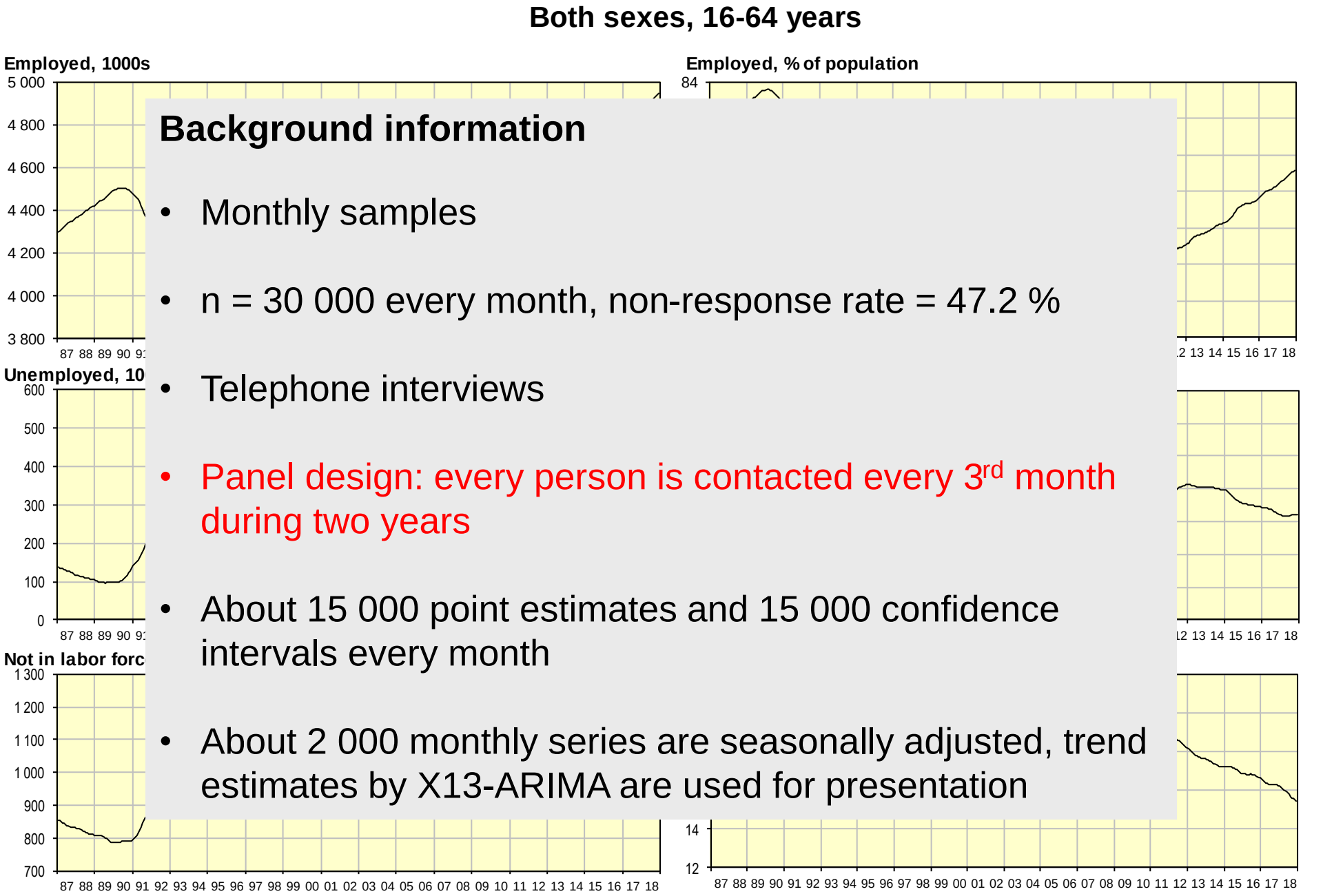
Not in labor force, 1000s



Not in labor force, % of population



The Swedish Labor Force Survey, estimated trends, monthly data



Background information

- Monthly samples
- n = 30 000 every month, non-response rate = 47.2 %
- Telephone interviews
- Panel design: every person is contacted every 3rd month during two years
- About 15 000 point estimates and 15 000 confidence intervals every month
- About 2 000 monthly series are seasonally adjusted, trend estimates by X13-ARIMA are used for presentation

2. “Panel design is good”

Chart 8. Standard errors for some estimates in the Swedish LFS.

Both sexes, 15-74 years	Employed 1000s	Unemployed 1000s	Not in labour force 1000s	Hours worked millions/week
1. Standard error Monthly estimate	19.4	11.5	19.4	0.95
2. Standard error Monthly estimate – change previous month	27.5	16.2	27.4	1.35
3. Standard error Quarterly estimate	11.4	6.3	11.1	0.55
4. Standard error Quarterly estimate – change previous quarter	**9.4**	**7.2**	**9.0**	**0.45**

This is why people want to use a panel design

LFS has been optimized to have small standard errors for changes with previous quarter

But what is this? How can $s(\hat{E})$ be *smaller* than the sampling standard error?

Chart 7. Standard errors for some estimates in the Swedish LFS compared with the residual component *Et*

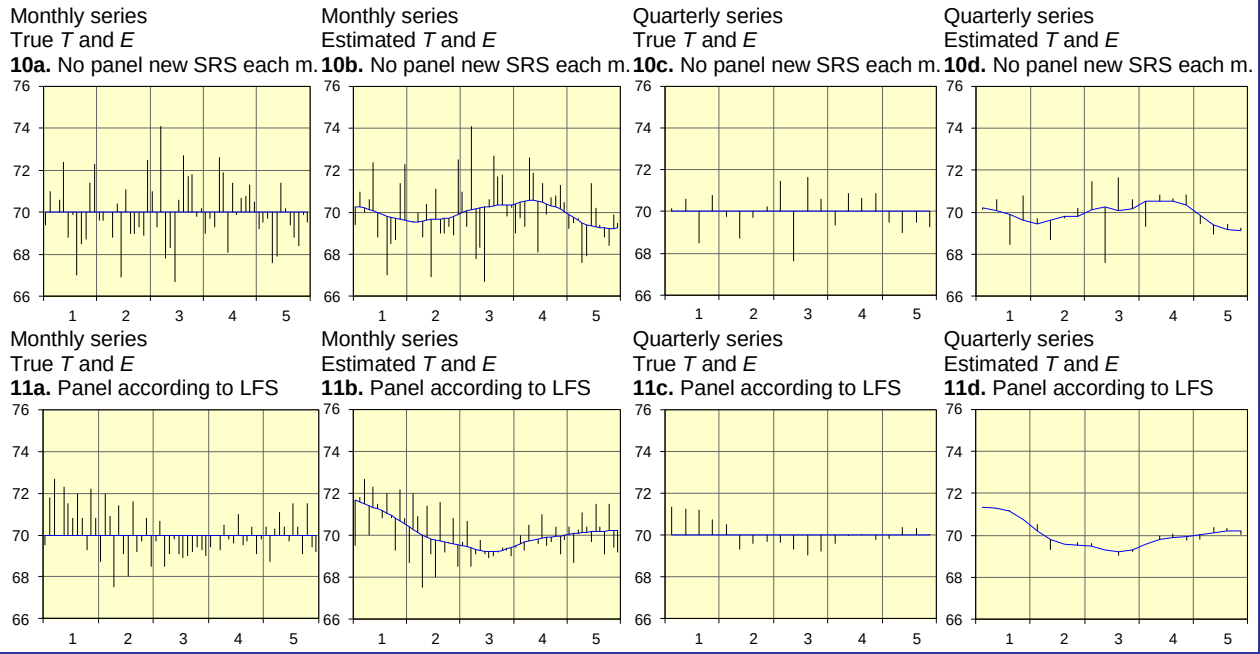
		Monthly estimates		Quarterly estimates	
Both sexes, 15-74 years	Point estimate Dec 2018	Standard error due to sampling	Standard deviation (E_t)	Standard error due to sampling	Standard deviation (E_t)
Employed, 1000s	5098.6	19.4	14.6	**11.4**	**6.1**
Unemployed, 1000s	326.9	11.5	9.9	**6.3**	**4.0**
Not in labour force, 1000s	2059.5	19.4	13.8	**11.1**	**6.3**
Hours worked, millions/week	148.1	**0.95**	**3.97**	**0.6**	**1.9**

2. “Panel design is good”

A sampling experiment that compares panel designs with simple random sampling

- In all simulations, we have a population where 70 % are employed each month and quarter
- In Chart 10a we have 60 monthly values with new independent simple random samples (SRS) each month. In Chart 10c we have the corresponding 20 quarterly values
- In Chart 11a we have 60 monthly values according to the panel design used in the LFS. One out of eight panels is new every month. Each person in the population has the same employment status all months.

Chart 10 and Chart 11. Effects of the panel design in the LFS.

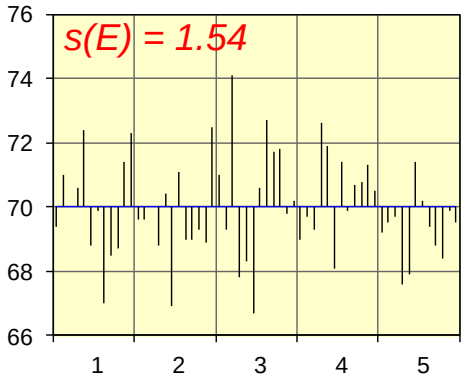


2. “Panel design is good”

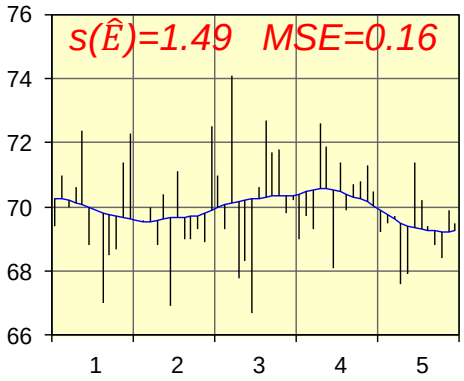
A sampling experiment that compares panel designs with simple random sampling

Monthly trend estimates ≈ quarterly trend estimates
SRS better than panel, MSE: SRS 0.16 – Panel 0.42
Quarterly panel: false accuracy, random errors are transformed into bias!
TSE thinking important

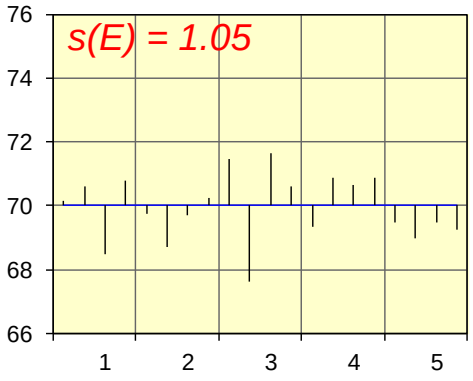
Monthly series
True *T* and *E*
10a. No panel new SRS each m.



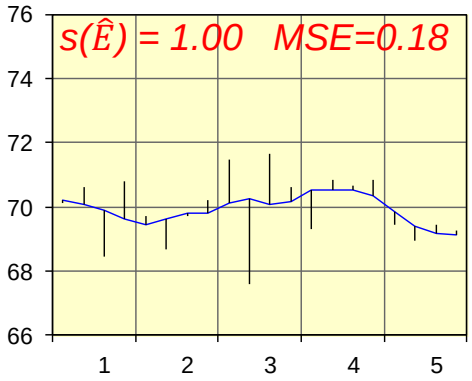
Monthly series
Estimated *T* and *E*
10b. No panel new SRS each m.



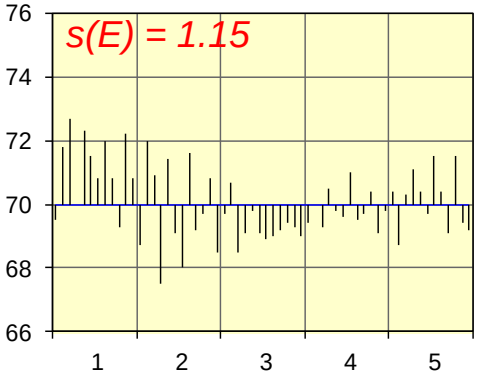
Quarterly series
True *T* and *E*
10c. No panel new SRS each m.



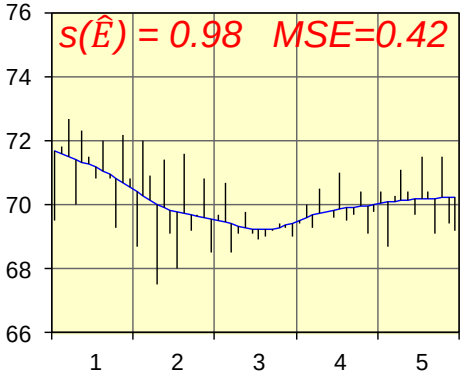
Quarterly series
Estimated *T* and *E*
10d. No panel new SRS each m.



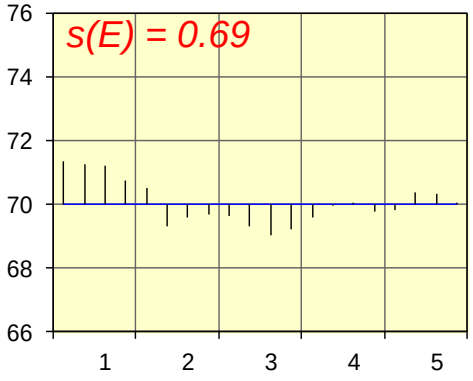
Monthly series
True *T* and *E*
11a. Panel according to LFS



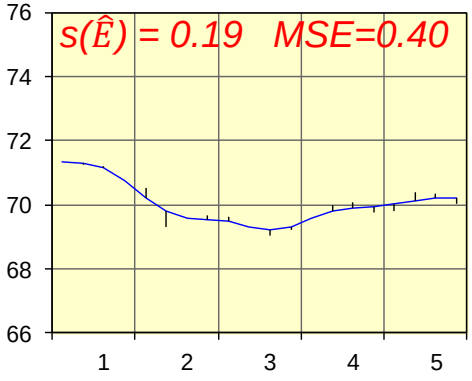
Monthly series
Estimated *T* and *E*
11b. Panel according to LFS



Quarterly series
True *T* and *E*
11c. Panel according to LFS



Quarterly series
Estimated *T* and *E*
11d. Panel according to LFS



2. “Panel design is good”

Quarterly panel: false accuracy, we publish this!

3. “Errors in the time variable, what is that?”

$Var\,(E_t) = Var\,(u_t) + Var\,(w_t)$, where

$Var\,(u_t)$ = variance of the estimator according to sampling theory, and

$Var\,(w_t)$ = variance of other short-term disturbances

In Chart 7 we have for the estimator of *Hours worked*:

$Var\,(E_t) = Var\,(u_t) + Var\,(w_t) = 3.97^2 = 0.95^2 + 14.9 =$

15.8	=	0.9	+ 14.9	as percent of $Var\,(E_t)$ we get:
100	=	6	+ 94	which means that other short-term disturbances than sampling errors is the main quality problem

We have earlier (1994) found similar problems with monthly business surveys:

$Var\,(E_t) = Var\,(u_t) + Var\,(w_t)$

100	=	4	+ 96	Hours worked, steel industry
100	=	1	+ 99	Hours worked, engineering industry
100	=	10	+ 90	Hourly pay, steel industry
100	=	17	+ 83	Hourly pay, engineering industry

Here, “months” are not calendar months. This creates errors in the time variable, a more serious problem than the sampling error of the estimates

3. “Errors in the time variable, what is that?”

Chart 17. The measurement periods originally used for the Swedish LFS for 2016

Year 2015	Year 2016												Year 2017
Dec	Jan	Feb	Mars	April	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan
1 T 49	1 F	1 M 5	1 T	1 F	1 S	1 O	1 F	1 M 31	1 T	1 L	1 T	1 T	1 S
2 O	2 L	2 T	2 O	2 L	2 M 18	2 T	2 L	2 T	2 F	2 S	2 O	2 F	2 M 1
3 T	3 S	3 O	3 T	3 S	3 T	3 F	3 S	3 O	3 L	3 M 40	3 T	3 L	3 T
4 F	4 M 1	4 T	4 F	4 M 14	4 O	4 L	4 M 27	4 T	4 S	4 T	4 F	4 S	4 O
5 L	5 T	5 F	5 L	5 T	5 T	5 S	5 T	5 F	5 M 36	5 O	5 L	5 M 49	5 T
6 S	6 O	6 L	6 S	6 O	6 F	6 M 23	6 O	6 L	6 T	6 T	6 S	6 T	6 F
7 M 50	7 T	7 S	7 M 10	7 T	7 L	7 T	7 T	7 S	7 O	7 F	7 M 45	7 O	7 L
8 T	8 F	8 M 6	8 T	8 F	8 S	8 O	8 F	8 M 32	8 T	8 L	8 T	8 T	8 S
9 O	9 L	9 T	9 O	9 L	9 M 19	9 T	9 L	9 T	9 F	9 S	9 O	9 F	9 M 2
10 T	10 S	10 O	10 T	10 S	10 T	10 F	10 S	10 O	10 L	10 M 41	10 T	10 L	10 T
11 F	11 M 2	11 T	11 F	11 M 15	11 O	11 L	11 M 28	11 T	11 S	11 T	11 F	11 S	11 O
12 L	12 T	12 F	12 L	12 T	12 T	12 S	12 T	12 F	12 M 37	12 O	12 L	12 M 50	12 T
13 S	13 O	13 L	13 S	13 O					13 T	13 T	13 S	13 T	13 F
14 M 51	14 T	14 S	14 M 11	14 T					14 O	14 F	14 M 46	14 O	14 L
15 T	15 F	15 M 7	15 T	15 F					15 T	15 L	15 T	15 T	15 S
16 O	16 L	16 T	16 O	16 L					16 F	16 S	16 O	16 F	16 M 3
17 T	17 S	17 O	17 T	17 S					17 L	17 M 42	17 T	17 L	17 T
18 F	18 M 3	18 T	18 F	18 M 16					18 S	18 T	18 F	18 S	18 O
19 L	19 T	19 F	19 L	19 T					19 M 38	19 O	19 L	19 M 51	19 T
20 S	20 O	20 L	20 S	20 O	20 F	20 M 25	20 O	20 L	20 T	20 T	20 S	20 T	20 F
21 M 52	21 T	21 S	21 M 12	21 T	21 L	21 T	21 T	21 S	21 O	21 F	21 M 47	21 O	21 L
22 T	22 F	22 M 8	22 T	22 F	22 S	22 O	22 F	22 M 34	22 T	22 L	22 T	22 T	22 S
23 O	23 L	23 T	23 O	23 L	23 M 21	23 T	23 L	23 T	23 F	23 S	23 O	23 F	23 M 4
24 T	24 S	24 O	24 T	24 S	24 T	24 F	24 S	24 O	24 L	24 M 43	24 T	24 L	24 T
25 F	25 M 4	25 T	25 F	25 M 17	25 O	25 L	25 M 30	25 T	25 S	25 T	25 F	25 S	25 O
26 L	26 T	26 F	26 L	26 T	26 T	26 S	26 T	26 F	26 M 39	26 O	26 L	26 M 52	26 T
27 S	27 O	27 L	27 S	27 O	27 F	27 M 26	27 O	27 L	27 T	27 T	27 S	27 T	27 F
28 M 53	28 T	28 S	28 M 13	28 T	28 L	28 T	28 T	28 S	28 O	28 F	28 M 48	28 O	28 L
29 T	29 F	29 M 9	29 T	29 F	29 S	29 O	29 F	29 M 35	29 T	29 L	29 T	29 T	29 S
30 O	30 L		30 O	30 L	30 M 22	30 T	30 L	30 T	30 F	30 S	30 O	30 F	30 M 5
31 T	31 S		31 T		31 T		31 S	31 O		31 M 44		31 L	31 T

11 weeks out of 52 belong to two months: 21 % of the interviews have errors in time

3. “Errors in the time variable, what is that?”

Chart 15. Hours worked all employed, millions per week, seasonally adjusted values and trends

Chart 15a. Original series without correction
s(E) = 3.9

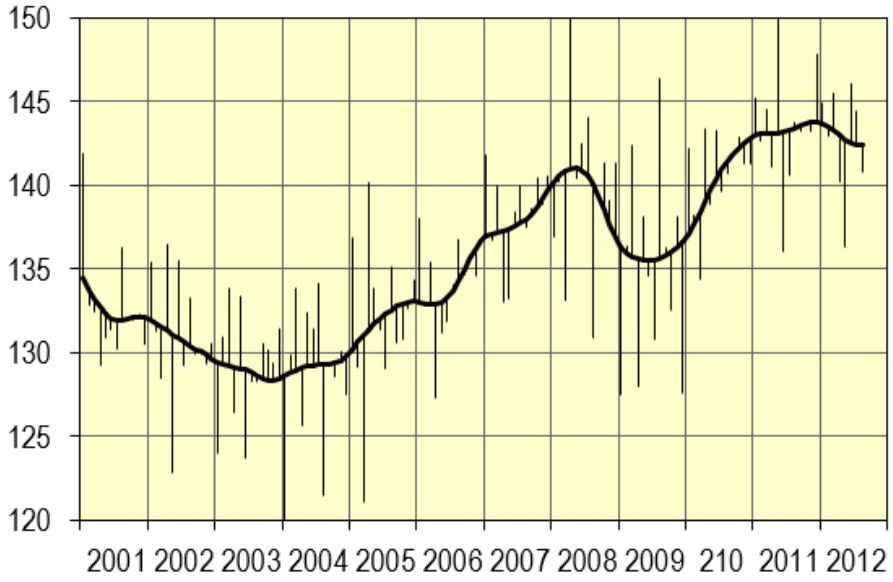
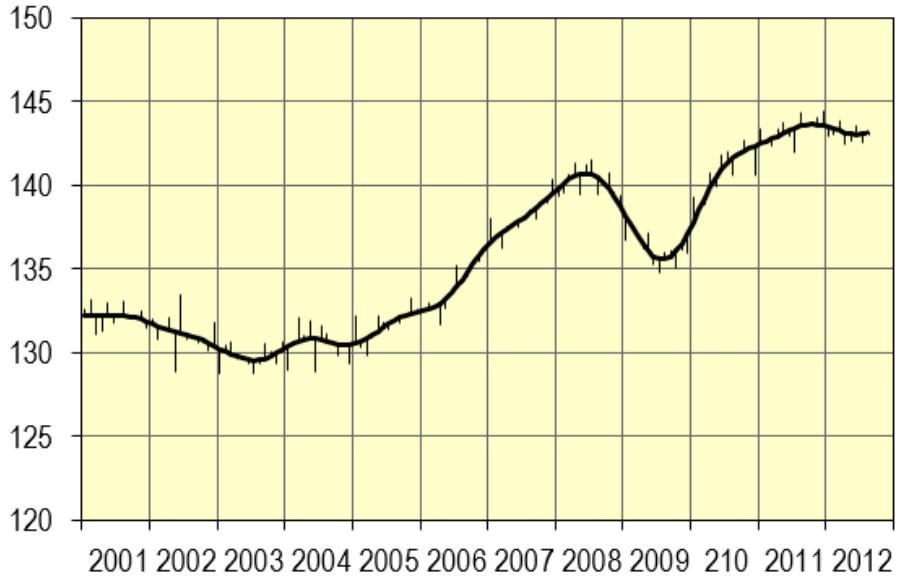


Chart 15b. Corrected for moving measurement periods and calendar variation. *s(E)* = 0.8



In the LFS, “months” are not calendar months. Instead measurement periods of 4, 4, and 5 weeks are used.

The errors in the time variables make seasonal adjustments difficult.

$s(\hat{E})$ measures the disturbing noise in the seasonally adjusted series.

Up to 2009, only Chart 15a was available for the users. During 2010 a method for correcting for moving measurement periods was introduced and Chart 15b became available.

3. “Errors in the time variable, what is that?”

Chart 16. Hours worked all employed, millions per week, seasonally adjusted values and trends

Chart 16a. Original series *without* outliers during 2016

Chart 16b. Original series *with* outlier during May 2016

The time series patterns that describe what is happening now, are extra important and errors here contribute to the TSE.

Chart 16a: *“Growth has ceased”* Chart 16b: *“Growth is continuing”*

Automatic outlier identification and correction is a method that is used by many statistical institutes. Outliers are indicators of quality problems, to include an outlier does not solve the problem, it only hides the problem.

3. “Errors in the time variable, what is that?”

11 weeks out of 52 belong to two months: 21 % of the interviews have error in the time variable

Chart 18. Original microdata

PIN	Week	LFS "year"	LFS "month"	Labour status	Hours worked	w_i
19001	53	2015	12	Employed	42	442.2
19002	9	2016	3	Unemployed	null	421.0

Person 19001 was employed during December 2015 and January 2016

Person 19002 was unemployed during February 2016 and March 2016

Split such interviews into two records in Chart 19

Calibration of weights to correct for errors in the time variable, new weight v_i

Chart 19. Microdata, corrected for errors in the time variable

PIN	Week	Calendar year	Calendar month	Labour status	Hours worked	w_i	days/week	v_i
19001	53	2015	12	Employed	42	442.2	4/7	252.7
19001	53	2016	1	Employed	42	442.2	3/7	189.5
19002	9	2016	2	Unemployed	null	421.0	1/7	60.1
19002	9	2016	3	Unemployed	null	421.0	6/7	360.9

3. “Errors in the time variable, what is that?”

Chart 18. Original microdata

PIN	Week	LFS "year"	LFS "month"	Labour status	Hours worked	w_i
19001	53	2015	12	Employed	42	442.2
19002	9	2016	3	Unemployed	null	421.0
19003	13	2016	3	Not in labour force	null	423.6
19004	17	2016	4	Employed	16	415.1
19005	22	2016	6	Employed	29	465.3
19006	26	2016	6	Not in labour force	null	434.0
19007	35	2016	9	Employed	35	429.5
19008	39	2016	9	Not in labour force	null	423.6
19009	44	2016	11	Not in labour force	null	415.9
19010	48	2016	12	Employed	6	433.8
19011	52	2016	12	Not in labour force	null	411.8

Chart 20. Sampling weights by year and month

Based on original …

Month"	"2015"	"2016"
3	*	844.6
4	*	415.1
6	*	899.3
9	*	853.1
11	*	415.9
12	442.2	845.6

and corrected microdata

Month	2015	2016	2017
1	*	189.5	58.8
2	*	60.1	*
3	*	602.9	*
4	*	537.3	*
5	*	192.2	*
6	*	580.4	*
7	*	186.0	*
8	*	184.1	*
9	*	548.0	*
10	*	180.4	*
11	*	542.4	*
12	252.7	600.9	*

PIN	Week	Calendar year	Calendar month	w_i	days/week	v_i
19001	53	2015	12	442.2	4/7	252.7
19001	53	2016	1	442.2	3/7	189.5
19002	9	2016	2	421.0	1/7	60.1
19002	9	2016	3	421.0	6/7	360.9
19003	13	2016	3	423.6	4/7	242.1
19003	13	2016	4	423.6	3/7	181.5
19004	17	2016	4	415.1	6/7	355.8
19004	17	2016	5	415.1	1/7	59.3
19005	22	2016	5	465.3	2/7	132.9
19005	22	2016	6	465.3	5/7	332.4
19006	26	2016	6	434.0	4/7	248.0
19006	26	2016	7	434.0	3/7	186.0
19007	35	2016	8	429.5	3/7	184.1
19007	35	2016	9	429.5	4/7	245.4
19008	39	2016	9	423.6	5/7	302.6
19008	39	2016	10	423.6	2/7	121.0
19009	44	2016	10	415.9	1/7	59.4
19009	44	2016	11	415.9	6/7	356.5
19010	48	2016	11	433.8	3/7	185.9
19010	48	2016	12	433.8	4/7	247.9
19011	52	2016	12	411.8	6/7	353.0
19011	52	2017	1	411.8	1/7	58.8

Persons by year and month:

Errors are corrected for 21 %

of the interviews

3. “Errors in the time variable, what is that?”

Chart 21. Hours worked all employed, millions per week, seasonally adjusted values and trends

Chart 21a. Errors in the time variable not corrected

Chart 21b. Time variable corrected with the new method

Chart 21a.

Errors in the time variable $\Rightarrow$ Seasonal adjustment failed $\Rightarrow$ Errors in $\hat{T}$

Chart 21b.

Corrected time variable $\Rightarrow$ All outliers vanished

Estimate $\hat{T}$ has high quality

Questions to the audience: What is the situation at your NSI?

Do people agree:

1. “Sample size is to small, can’t publish monthly data” Yes/No
2. “Panel design is good” Yes/No
3. “Errors in the time variable, do we have that?” Yes/No