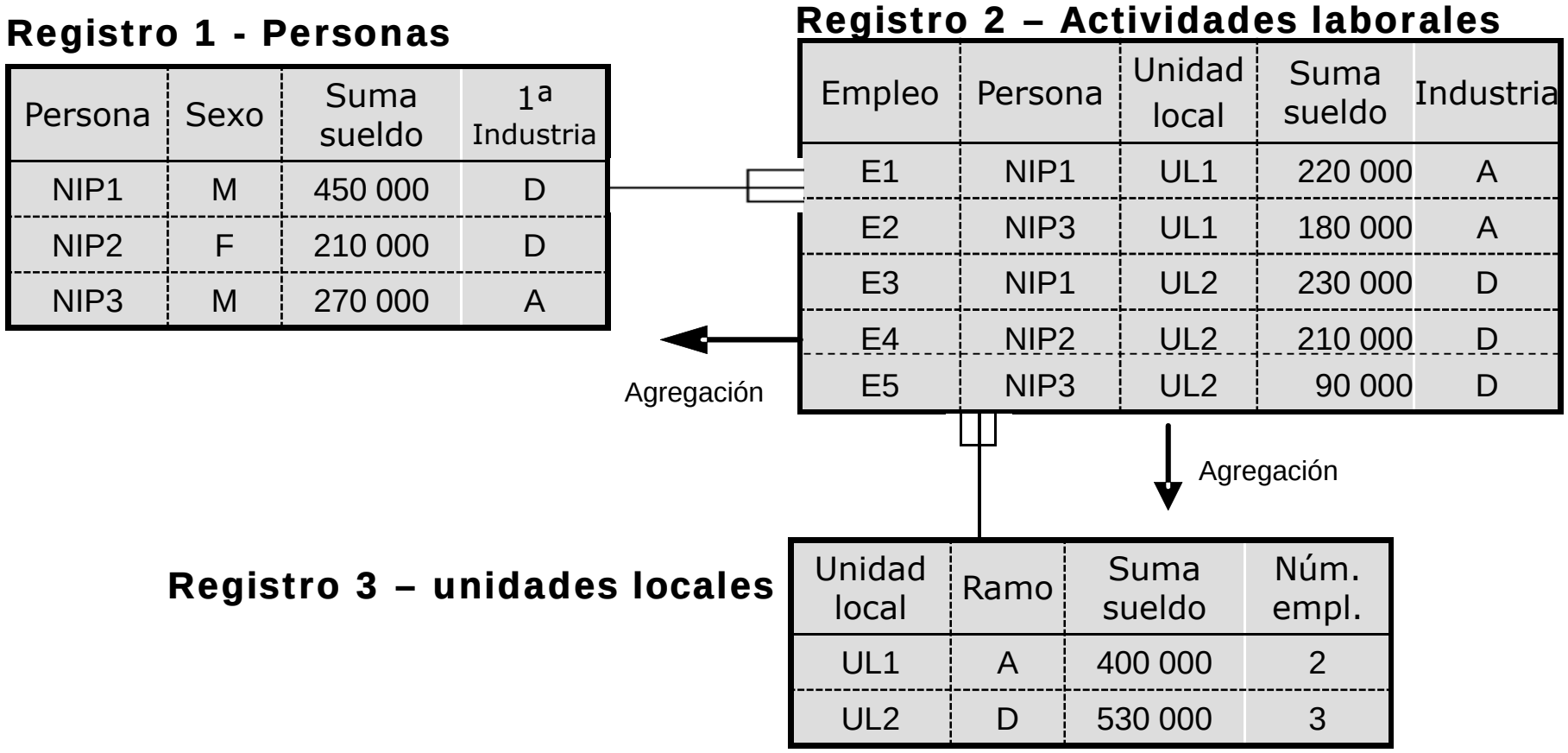
Anders & Britt Wallgren
Statistics Sweden and Örebro University
ba.statistik@telia.com

9: Variables de valores múltiples en el sistema de registros

1. Variables de valores múltiples
2. Estimación con combinación de objetos

Fig. 9.1 Número de empleados y suma de sueldos y salarios en distintos registros



Esto ilustra un problema general: Cuando integramos distintos registros y creamos nuevas variables agregando variables cualitativas, surgen inconsistencias.

Cuatro encuestas con la misma variable: ¡Las inconsistencias son claras!

Cuentas Nacionales recibe información de muchos encuestas con variables diferentes. Hay inconsistencias, ¿pero dónde?

Fig. 4. Número de empleados en nov. 2004 según cuatro diferentes encuestas

Empleados noviembre 2004, miles

Actividad económica NACE 2002	Registro empresarial Empresas (1)	Establecimientos (2)	(1)-(2)	Registro de empleo Empleo principal (3)	(2)-(3)	EFL primero NACE (4)	*Margen de error*	(3)-(4)
A Agricultura y silvicultura	34	36	*-1*	37	*-1*	26	*5*	*10*
B Pesca	1	1	*0*	0	*0*	0	*1*	*0*
C Minería y explotación de cantera	9	8	*1*	7	*0*	5	*2*	*3*
D Manufacturera	679	629	*51*	710	*-81*	635	*23*	*75*
E Electricidad, gas y agua	21	22	*-1*	28	*-6*	29	*5*	*-1*
F Construcción	197	209	*-12*	215	*-6*	199	*14*	*17*
G Ventas a mayoreo y menudeo	456	453	*3*	484	*-31*	456	*20*	*28*
H Hoteles y restaurantes	89	93	*-4*	99	*-5*	106	*10*	*-8*
I Transporte, comunicación	240	242	*-2*	243	*-1*	236	*15*	*7*
J Intermediación financiera	83	77	*6*	85	*-8*	78	*9*	*7*
K Bienes raíces, actividad empresarial	457	524	*-67*	457	*67*	470	*20*	*-13*
L Gobierno	139	215	*-77*	239	*-24*	230	*15*	*9*
M Educación	382	408	*-27*	431	*-23*	462	*20*	*-30*
N Salud y trabajo social	836	684	*152*	675	*8*	675	*24*	*0*
O Otras actividades de prestación de servicios	142	163	*-21*	175	*-12*	168	*13*	*8*
Actividad desconocida	0	0	*0*	38	*-37*	4		*34*
Total	**3 763**	**3 763**	***0***	**3 924**	***-160***	**3 778**		***145***

Una empresa puede tener uno o más establecimientos (1) 1er ramo por empresas (N=900 151)
(2) er ramo por establecimiento (N=974 576)

Diagrama 4. Número de empleados en nov. 2004 según cuatro estudios distintos

Empleados noviembre 2004, miles

Encuesta de Fuerza Laboral

(1) 1er Empresas en la industria

(2) 1er Estableci-mientos en la industria

(3) 1er Industria por el esta-blecimiento del 1er trabajo de un empleado

(4) Igual que (3)

	Registro empresarial Empresas (1)	Establecimientos (2)	(1)-(2)	Registro de empleo Empleo principal (3)	(2)-(3)	EFL primero NACE (4)	*Margen de error*	(3)-(4)
A	34	36	*-1*	37	*-1*	26	*5*	*10*
B	1	1	*0*	0	*0*	0	*1*	*0*
C	9	8	*1*	7	*0*	5	*2*	*3*
D	679	629	*51*	710	*-81*	635	*23*	*75*
E	21	22	*-1*	28	*-6*	29	*5*	*-1*
F	197	209	*-12*	215	*-6*	199	*14*	*17*
G	456	453	*3*	484	*-31*	456	*20*	*28*
H	89	93	*-4*	99	*-5*	106	*10*	*-8*
I T	240	242	*-2*	243	*-1*	236	*15*	*7*
J	83	77	*6*	85	*-8*	78	*9*	*7*
K	457	524	*-67*	457	*67*	470	*20*	*-13*
L	139	215	*-77*	239	*-24*	230	*15*	*9*
M	382	408	*-27*	431	*-23*	462	*20*	*-30*
N	836	684	*152*	675	*8*	675	*24*	*0*
O servicios	142	163	*-21*	175	*-12*	168	*13*	*8*
U Actividad desconocida	0	0	*0*	38	*-37*	4		*34*
Total	**3 763**	**3 763**	***0***	**3 924**	***-160***	**3 778**		***145***

¡Esto representa un grave problema en las estadísticas económicas en todos los países!

En un INE hay tres categorías:

1. Especialistas por tema:
   Contenido, definiciones "1° NACE"
2. Metodólogos: matemáticas – probabilidad – inferencia – muestreo, paradigma de "una encuesta a la vez".
3. Especialistas en TI: sistemas, variables de valores múltiples

¡El problema no es necesario!

¡El problema no es necesario!

El problema surgió cuando (Pág. 148):

- Ocurrencia repetida de objetos - genera variables de valores múltiples (Registros por año calendario).
- Relaciones de muchos a uno – genera variables de valores múltiples.
- Las propias variables de valores múltiples (NACE, Occ., Educación …) , Por ej. gasolinera (45201, 45320, 47112, 47300, 56100, 77110).
- Combinaciones de las anteriores (NACE, Empleo -> Persona).

Registro por año calendario de población de personas 2005

Per-sona	Dirección	Muni-cipio	De fecha aaaammdd	A fecha aaaammdd	Ponderador= tiempo en el domicilio, años
NIP 1	Dirección1	1	20050101	20050517	136/365 = 0.37
NIP 1	Dirección2	2	20050518	20051231	229/365 = 0.63
NIP 2	Dirección3	1	20050101	20051231	365/365 = 1.00
NIP 3	Dirección4	2	20050101	20050314	73/365 = 0.20
NIP 3	Dirección5	1	20050315	20050925	194/365 = 0.53
NIP 3	Dirección6	2	20050926	20051231	98/365 = 0.27
NIP 4	Dirección7	2	20050101	20050605	156/365 = 0.43
NIP 4	Dirección8	1	20050606	20051231	209/365 = 0.57

Población promedio 2005

Municipio	Frecuencia absoluta	Frecuencia relativa
1	0.37 + 1.00 + 0.53 + 0.57 = 2.47	62%
2	0.63 + 0.20 + 0.27 ´0.43 = 1.53	38%
Total	4.00	100%

Relaciones de muchos a uno – genera variables de valores múltiples. Una persona puede tener muchos trabajos

Número de empleados y suma de sueldos en distintos registros

Persona	Unidad de establecimientos	Sueldo	Industria	Ponderador
PIN1	UL1	220 000	A	22/45 = 0.49
PIN1	UL2	230 000	D	23/45 = 0.51
PIN 2	UL2	210 000	D	21/21 = 1.00
PIN3	UL1	180 000	A	18/27 = 0.67
PIN3	UL2	90 000	D	9/27 = 0.33

Empleados por industria

Industria	Número de empleados
A	0.49 + 0.67 = 1.16
D	0.51 + 1.00 + 0.33 = 1.84
Total	3.00

Registro de personas con datos sobre ocupación

Persona	Ocupación	Extensión	Ponderador alt. 1	Ponderador alt. 2	Ponderador alt. 3
NIP1	Estadístico	100	1	0.80	1.00
NIP1	Granjero	15	0	0.12	0.15
NIP1	Político	10	0	0.08	0.10
NIP2	Camillero hosp.	30	1	0.60	0.30
NIP2	Limpiador	20	0	0.40	0.20
NIP3	Vendedor	10	1	1.00	0.10
NIP4	Vendedor	50	1	1.00	0.50
NIP5	Vendedor	20	1	1.00	0.20
NIP6	Vendedor	100	1	1.00	1.00
Total			**6**	**6.00**	**3.55**

Personas y empleados de tiempo completo por ocupación; 3 alternativas

	Alternativa 1		Alternativa 2		Alternativa 3	
Ocupación	**Núm.**	**%**	**Núm.**	**%**	**Núm.**	**%**
Político	0.00	0.0	0.08	1.3	0.10	2.8
Estadístico	1.00	16.7	0.80	13.3	1.00	28.2
Camillero hospital	1.00	16.7	0.60	10.0	0.30	8.5
Dependiente tienda	4.00	66.7	4.00	66.7	1.80	50.7
Granjero	0.00	0.0	0.12	2.0	0.15	4.2
Personal de limpieza	0.00	0.0	0.40	6.7	0.20	5.6
Total	**6.00**	**100.0**	**6.00**	**100.0**	**3.55**	**100.0**

Estimación de resultados con una industria utilizando dos métodos

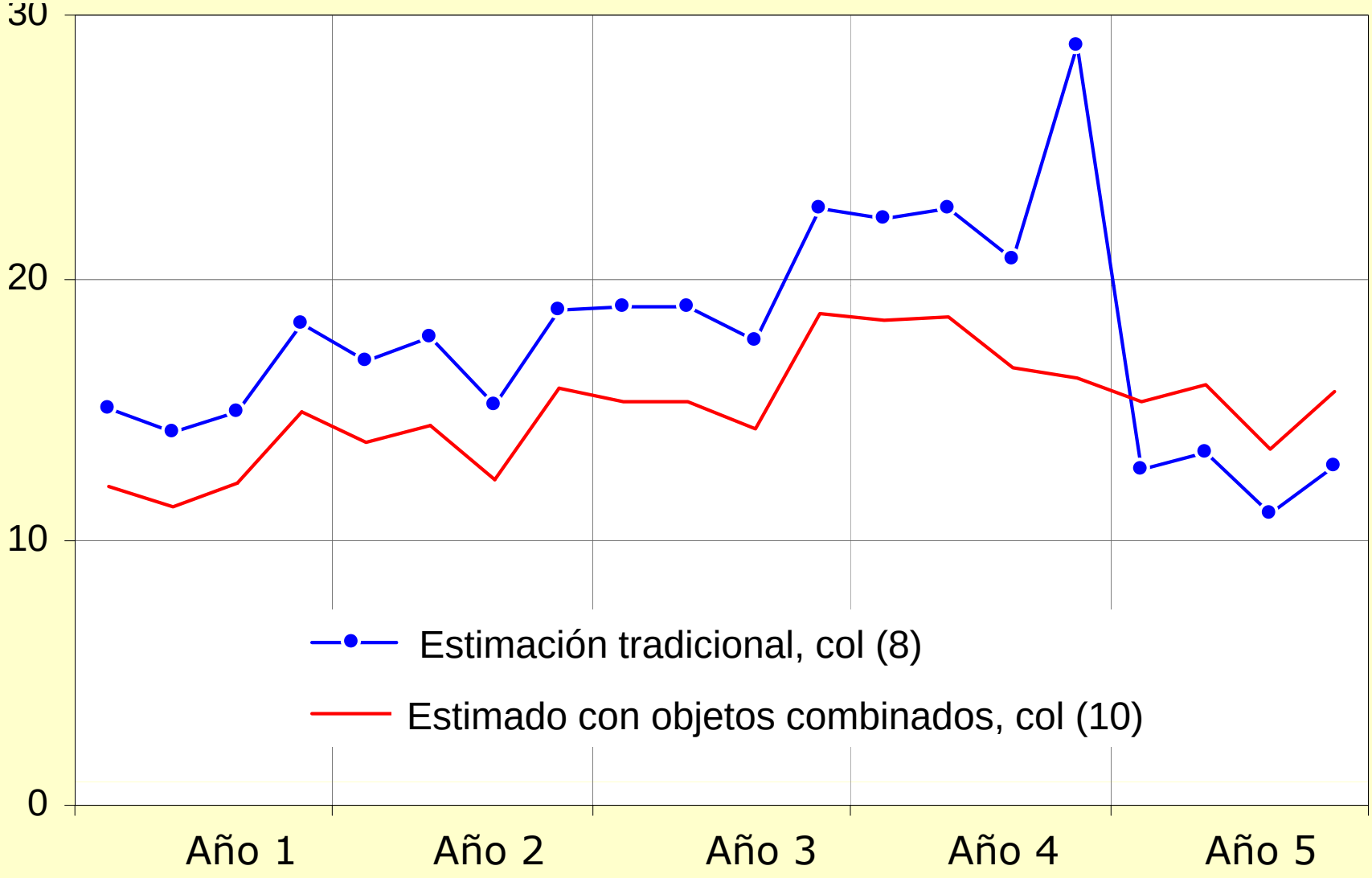
Trimes-tre del año		Industria R excl. X Ltd	X Ltd antes de compra	X Ltd después de compra	Partici-pación. X Ltd en la industria	Ponderador con cálculo tradicional	Contribución de XLT a Industria R, cálculo tradicional	Estimación tradicional Industria R	Contribución de LTD a R con objetos combinados	Industria R con objetos combinados
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
...		...	...		...	...	...	...	...	...
4	1	13 071	9 259		0.57	1	9 259	22 330	5 278	18 349
4	2	13 127	9 509		0.57	1	9 509	22 636	5 420	18 547
4	3	11 253	9 499		0.57	1	9 499	20 752	5 414	16 668
4	4	12 921		15 881	0.21	1	15 881	28 802	3 335	16 256
5	1	12 782		12 397	0.21	0	0 000	12 782	2 603	15 385
5	2	13 360		12 634	0.21	0	0 000	13 360	2 653	16 013
5	3	11 098		11 621	0.21	0	0 000	11 098	2 440	13 538
5	4	12 888		13 209	0.21	0	0 000	12 888	2 774	15 662

Azul

Rojo

Resultados en una industria, dos estimados

SEK miles de millones por trimestre



- Estos problemas de inconsistencias entre encuestas y cambios de nivel en las series de tiempo pueden resolverse usando ponderadores.
- En el Registro Empresarial Sueco hay información sobre ponderadores para diferentes industrias para cada establecimiento.
- De manera similar a la forma en que se utilizan los ponderadores para dar estimaciones en las encuestas por muestreo.
- Lo anterior se ilustra con el siguiente ejemplo:

1. Registro de unidades de locales

Unidad local	1° NACE	Empleados
LU1	DJ	218
LU2	DH	293
LU3	Dj	156
Suma:		667

2. Unidades de locales * Industria

Unidad local	NACE	Peso
LU1	DJ	1
LU2	DH	0.51
LU2	DJ	0.49
LU3	DJ	0.4
LU3	DH	0.3
LU3	DK	0.3

3. El cotejo de 1 y 2 por identificación de unidad de local da como resultado un registro con objetos combinados

Unidad local	1er NACE	Empleados	NACE	Ponderador	=(No. Emp.)* ponderador
LU1	DJ	218	DJ	1	218.00
LU2	DH	293	DH	0.51	149.43
LU2	DH	293	DJ	0.49	143.57
LU3	DJ	156	DJ	0.4	62.40
LU3	DJ	156	DH	0.3	46.80
LU3	DJ	156	DK	0.3	46.80

Estimado de 1	
DH	293
DJ	374
DK	0
Suma:	667

Estimado de 3	
DH	196
DJ	424
DK	47
Suma:	667