Measurement and sampling in the Living Conditions Survey 2008/09

Martin Wittenberg

Director, DataFirst

6 July 2017

Introduction

- Measurement of consumption/expenditures
  - Why it is important
  - IES/LCS series
- The LCS 2008/09
  - A puzzle
  - The financial crisis?
- Sampling issues
  - Impact of reweighting the sample
- International comparisons
- Conclusions

Measurement of consumption/ expenditure

Why are consumption surveys important?

- Main instruments for measuring poverty and inequality
  - Consumption is thought to be better measured than income in developing countries
- Food expenditures particularly important

Consumption surveys in South Africa

- First nationally representative survey with microdata released was the Project for Statistics on Living Standards and Development (PSLSD)
  - 1993
  - Run on template of the World Bank's Living Standards Measurement Surveys
  - Multipurpose survey
  - Sample of around 8000 households

- Since 1995 Statistics SA has run Income and Expenditure Surveys (IESs) and since 2008 Living Conditions Surveys (LCSs)
  - 1995 IES was linked to 1995 October Household Survey
  - Around 30 000 households
  - Much more detailed consumption schedule
  - 2000 IES – linked to the Labour Force Survey
    - "broke" the rotating panel
  - Lots of criticism about its accuracy (or lack thereof)

Redesign of the IES

- 2005 IES was run as a year-long survey
  - Diary method instead of recall
  - Not linked to other questions (e.g. labour market outcomes, education, health) and not linked to other surveys
  - Break in the series: is the data comparable to previous years?

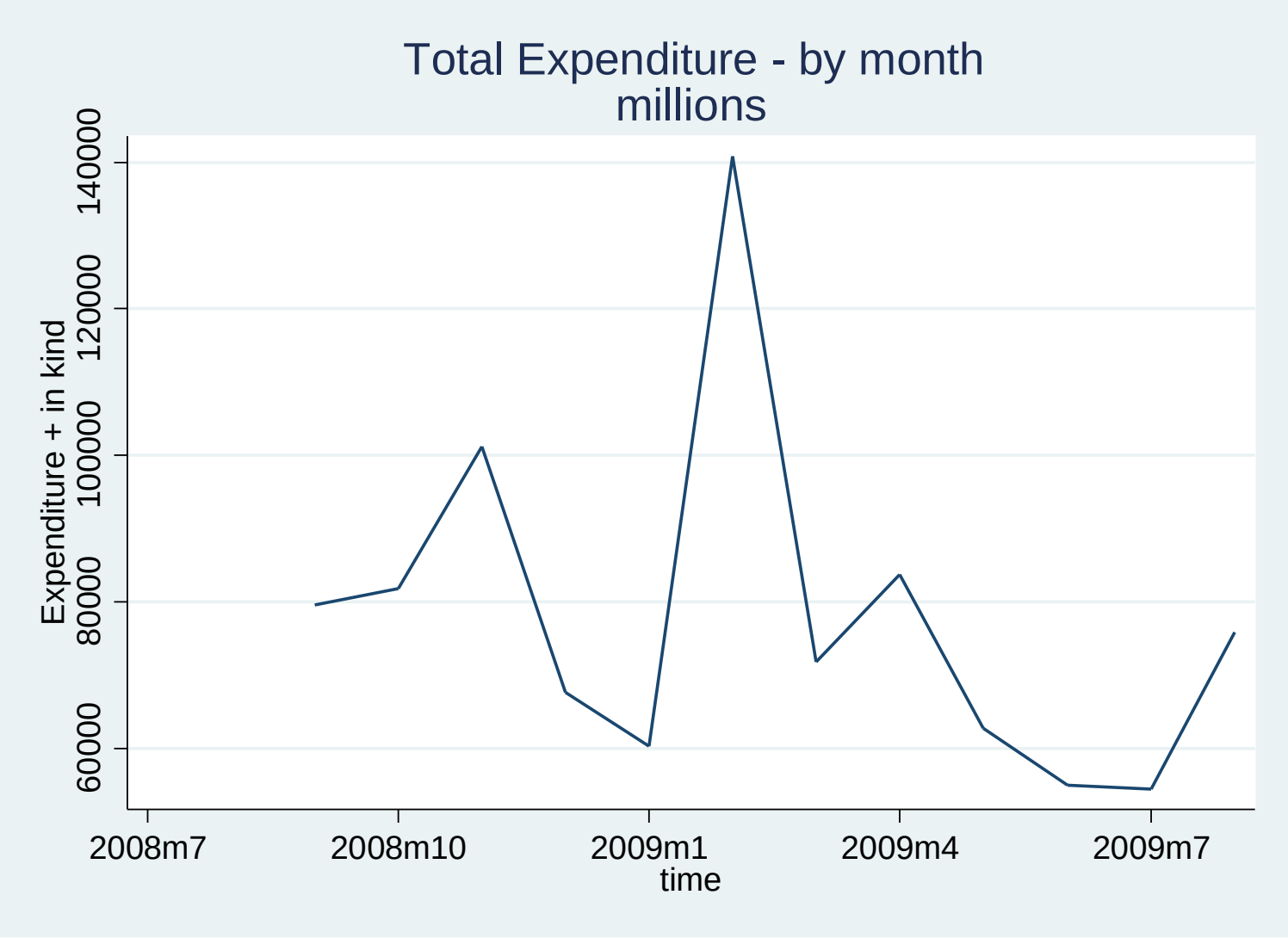
- 2008/09 LCS
  - Diary method
  - Lots of additional modules
  - Main instrument for thinking about poverty, well-being
  - Around 25 000 households

The Living Conditions Survey 2008/09

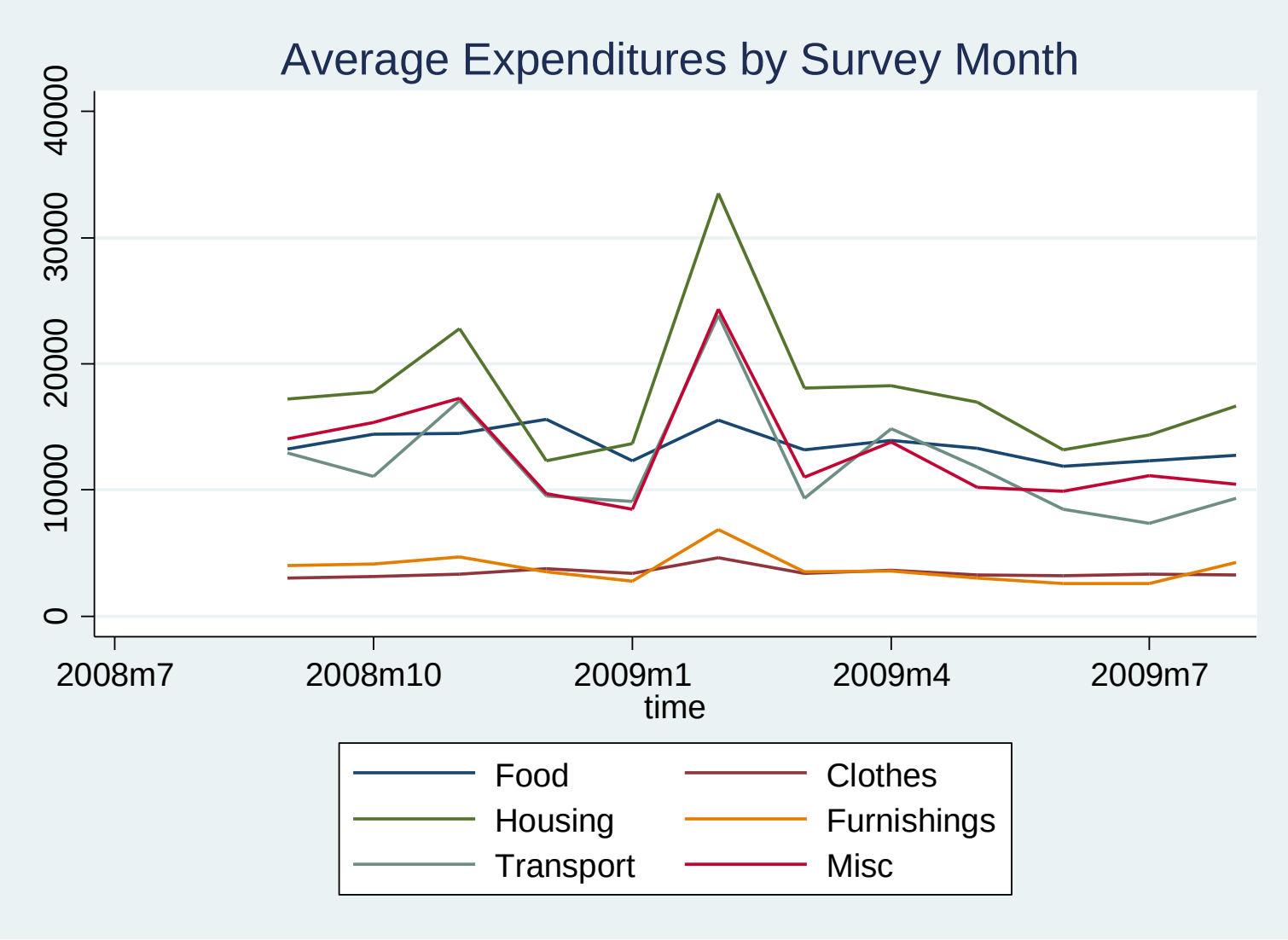
Design

- Around 3000 primary sampling units
- Fieldwork from September 2008 to August 2009
- Coincided with the international financial crisis
- Sampling:
  - “The sample was evenly split into four rotations (quarters) with national representativity in each rotation. Each rotation (consisting of a sample for three months) was then evenly split into monthly samples. Ultimately, the sample was evenly spread over the 12 survey periods (one month each).”
- Households were interviewed over a period of one month
  - Diary
  - Some recall expenditures– for previous 11 months (e.g. clothing) or 12 months (e.g. gym memberships) in the household questionnaire

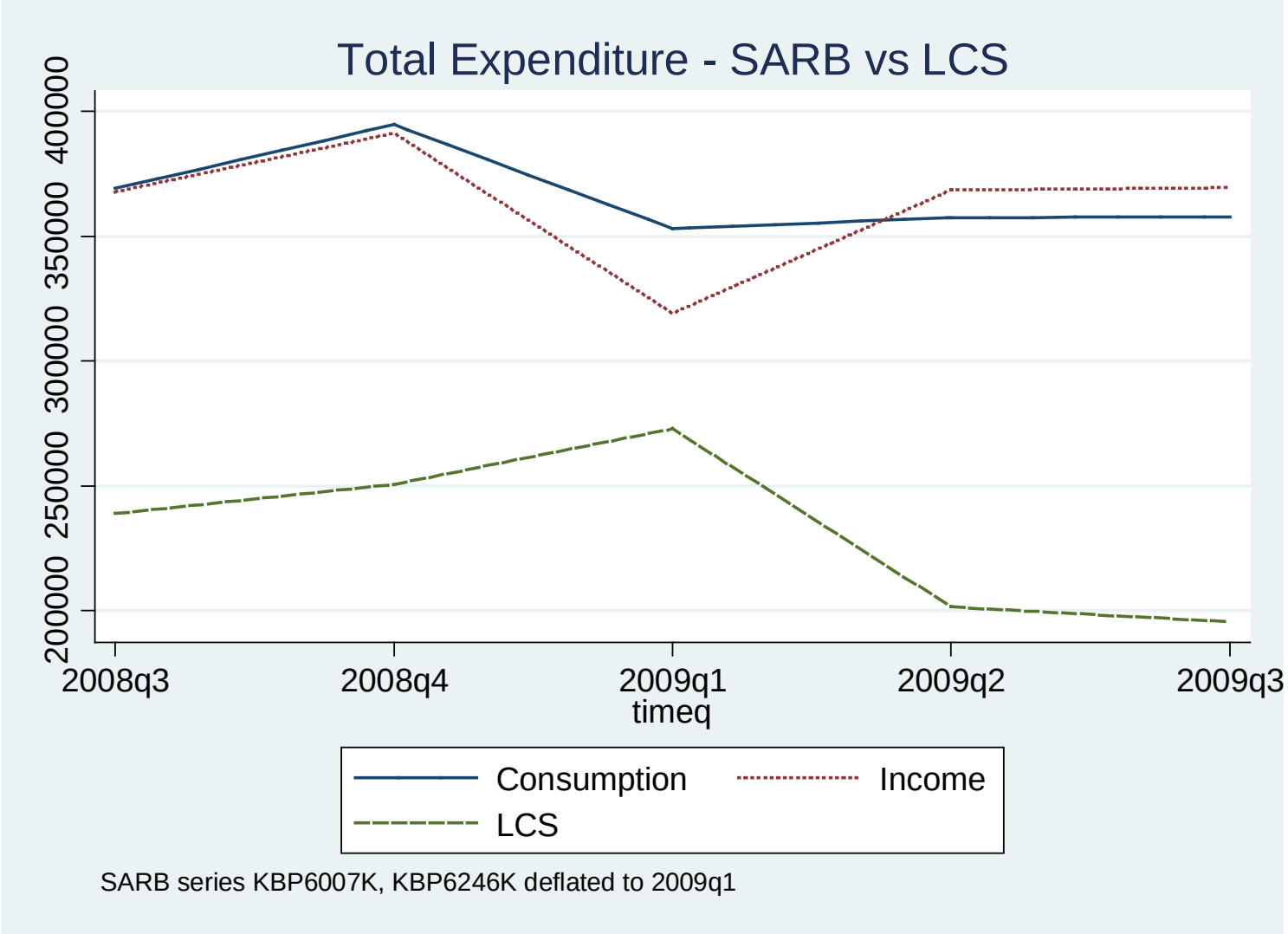
A puzzle


Total Expenditure - by month
millions
Expenditure + in kind
60000
80000
100000
120000
140000
2008m7
2008m10
2009m1
2009m4
2009m7
time

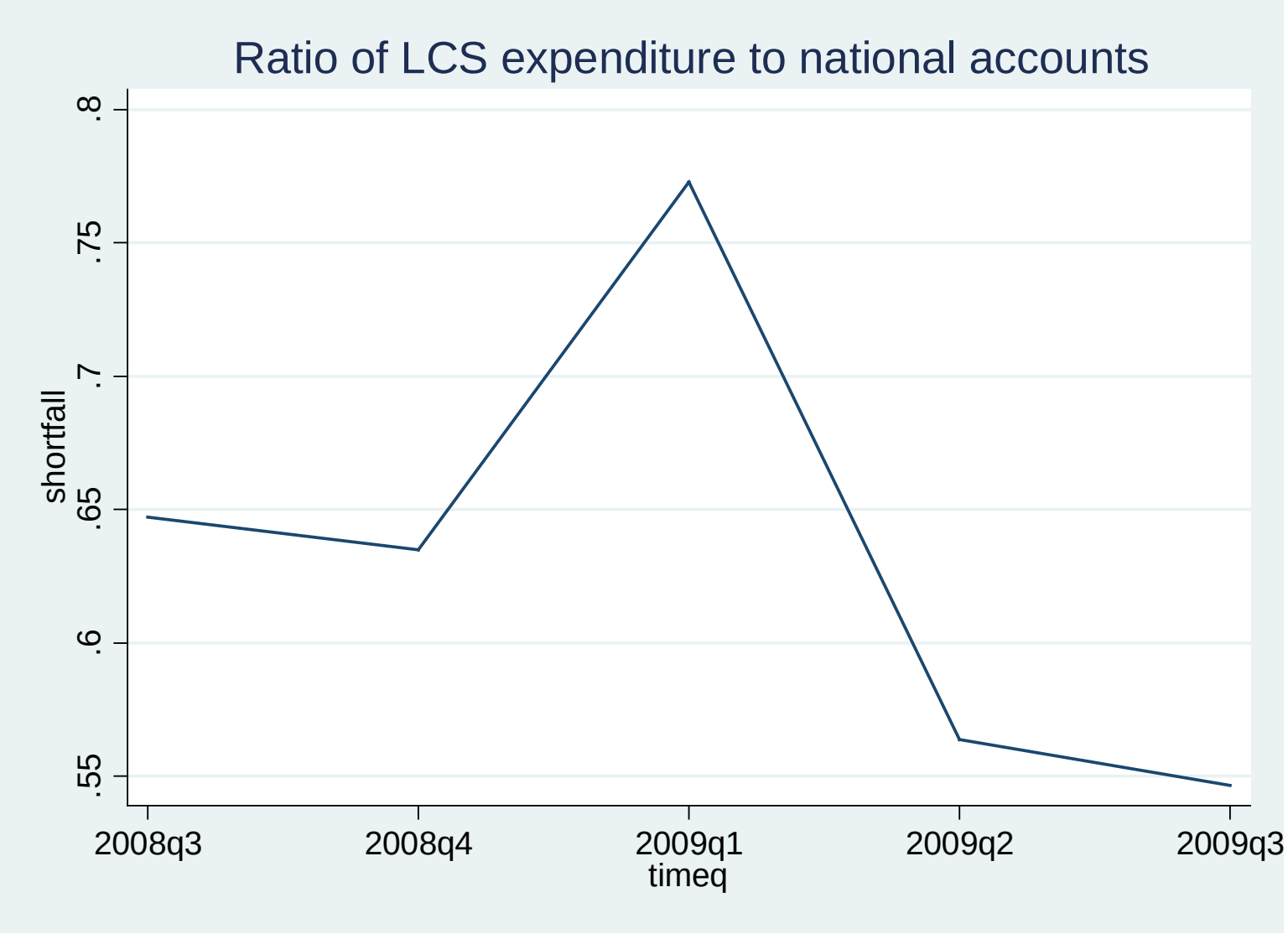
The peak occurs in many items



How does this correspond to macro conditions?

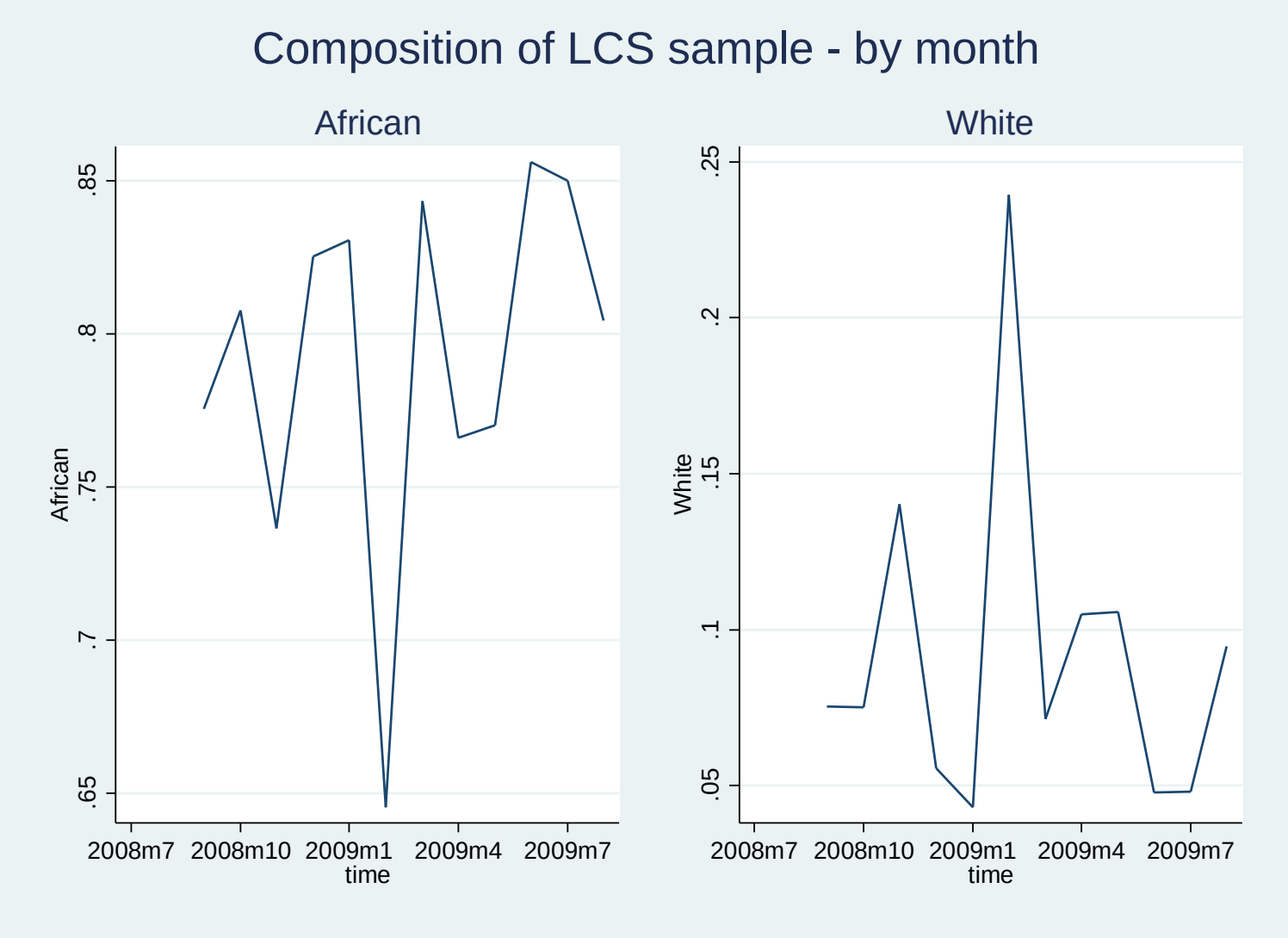


The gap


Ratio of LCS expenditure to national accounts
shortfall
.55
.6
.65
.7
.75
.8
2008q3
2008q4
2009q1
2009q2
2009q3
timeq

Sampling Issues

A straight-forward explanation



Not a random effect

Month	Year	African	[95% Conf.	Interval]	White	[95% Conf.	Interval]
9	2008	3 431 345	3 140 238	3 722 451	333 569	215 135	452 003
10	2008	3 353 251	3 056 984	3 649 518	311 750	188 712	434 789
11	2008	2 963 781	2 650 643	3 276 918	564 875	396 742	733 007
12	2008	3 493 601	3 230 897	3 756 306	235 777	116 830	354 724
1	2009	3 430 916	3 197 592	3 664 240	178 291	80 320	276 262
2	2009	2 621 546	2 274 614	2 968 478	972 612	749 025	1 196 199
3	2009	3 452 347	3 184 262	3 720 432	291 972	175 770	408 174
4	2009	3 190 518	2 895 359	3 485 676	437 654	297 941	577 368
5	2009	2 686 678	2 434 236	2 939 120	368 742	238 872	498 611
6	2009	3 388 284	3 151 927	3 624 641	189 095	105 744	272 446
7	2009	3 243 666	3 016 728	3 470 603	183 975	103 704	264 246
8	2009	3 516 897	3 228 857	3 804 937	413 534	264 242	562 827

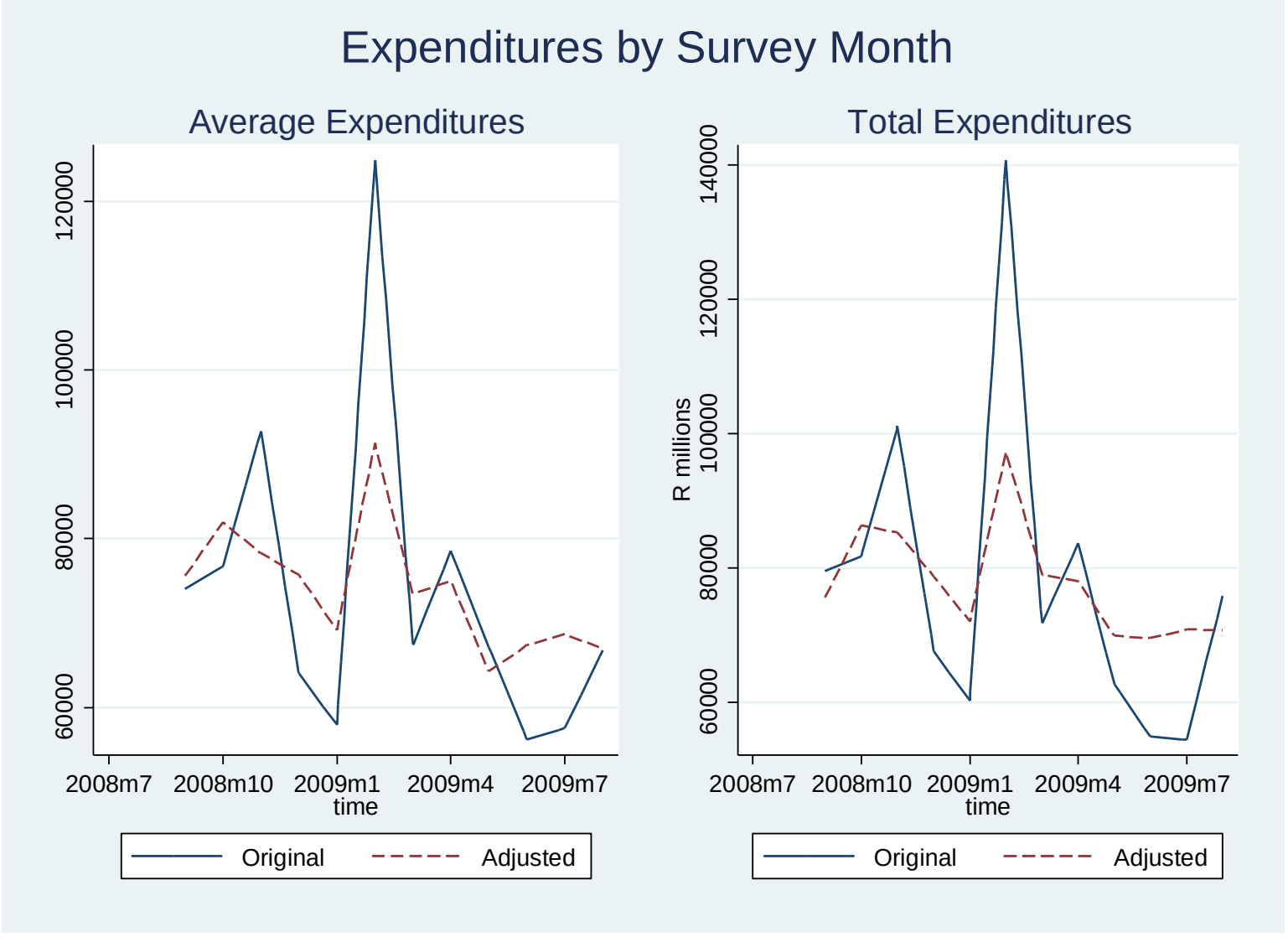
Every quarter of field work

Quarter	African	[95% Conf.	Interval]	Coloured	[95% Conf.	Interval]
1	9 748 376	9 221 340	10 300 000	1 143 394	889 908	1 396 881
2	9 546 063	9 023 659	10 100 000	1 177 528	929 658	1 425 397
3	9 329 542	8 855 642	9 803 443	1 077 208	831 854	1 322 563
4	10 100 000	9 707 287	10 600 000	1 010 155	792 747	1 227 564
Quarter	Indian	[95% Conf.	Interval]	White	[95% Conf.	Interval]
1	498 413	269 374	727 452	1 210 194	965 770	1 454 618
2	315 639	167 358	463 920	1 386 680	1 093 199	1 680 160
3	242 024	131 160	352 888	1 098 368	873 756	1 322 980
4	200 667	90 217	311 117	786 605	593 325	979 885

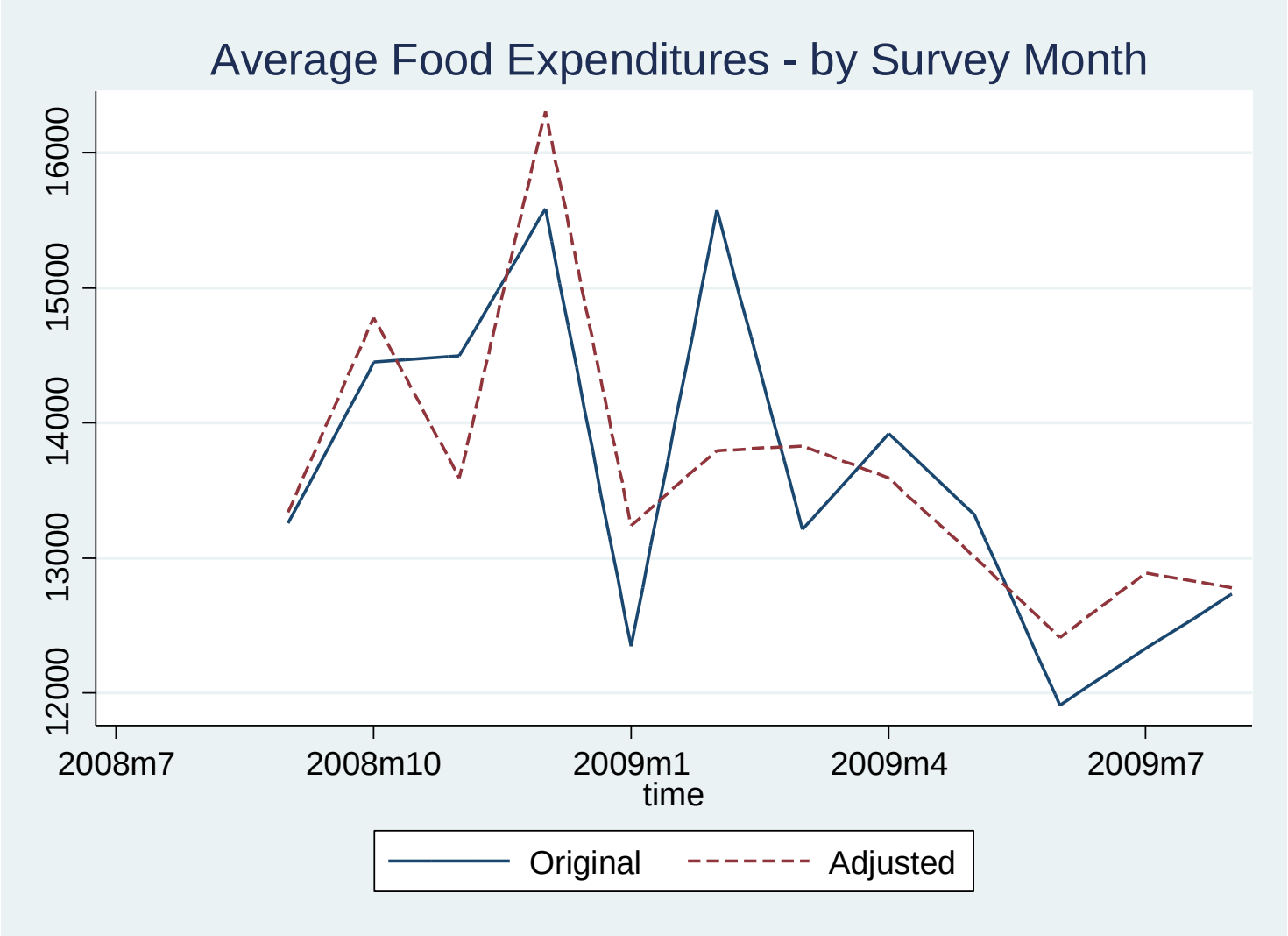
What can we do about this?

- Post-stratifying each monthly sample to the same racial breakdown
  - Reweighting February “whites” down and *vice versa*
- A rather crude adjustment
  - Ignoring any other covariates
  - Not doing the adjustments on the design weights

What impact does this have?



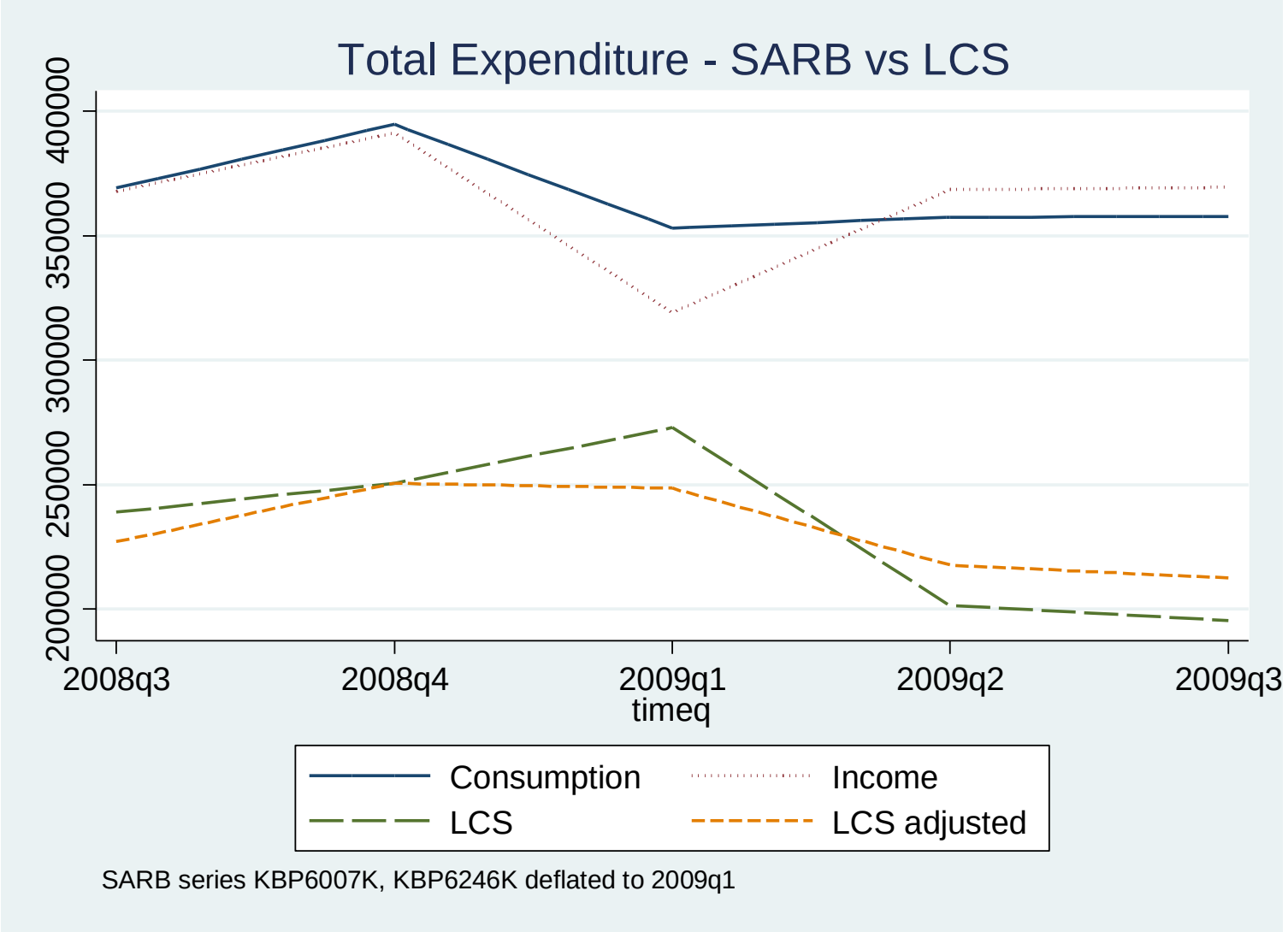
And for food


Average Food Expenditures - by Survey Month
16000
15000
14000
13000
12000
2008m7
2008m10
2009m1
2009m4
2009m7
time
Original
Adjusted

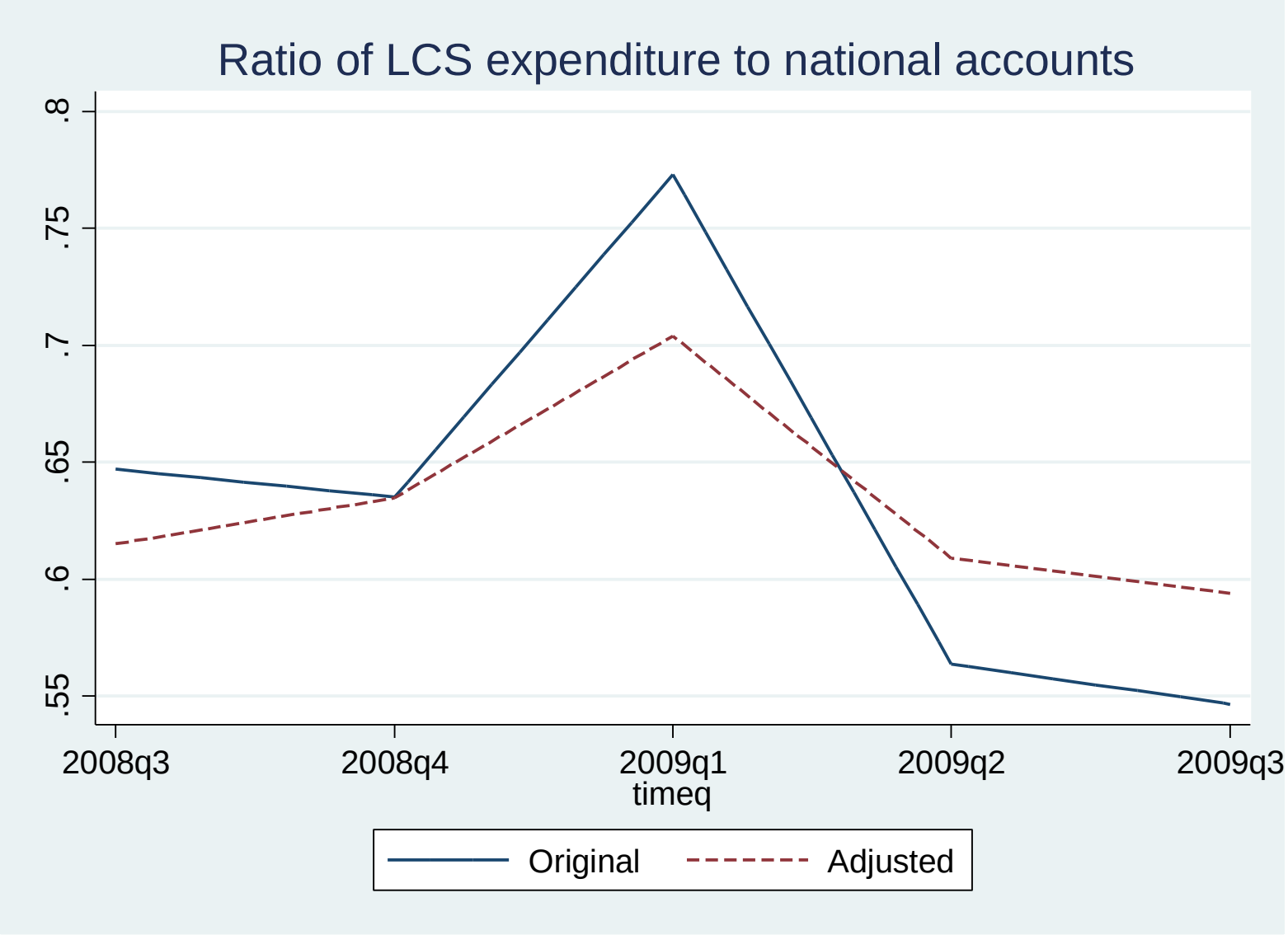
Comment

- Big spike is now in December
- February spike has more or less disappeared
- BUT it is much more evident that food expenditures are much lower at the end
  - Is this a “macro effect” or not?

The macro comparisons



The ratios


Ratio of LCS expenditure to national accounts
.55
.6
.65
.7
.75
.8
2008q3
2008q4
2009q1
2009q2
2009q3
timeq
Original
Adjusted

International Comparisons

Dabalen *et al* “Is poverty in Africa overestimated because of poor data?”

- Look at consumption surveys collected by means of diaries over the course of a year
- Typical pattern: fewer items recorded, lower expenditures in the last few months of survey

Impact of collecting surveys over a year: Dabalen *et al*

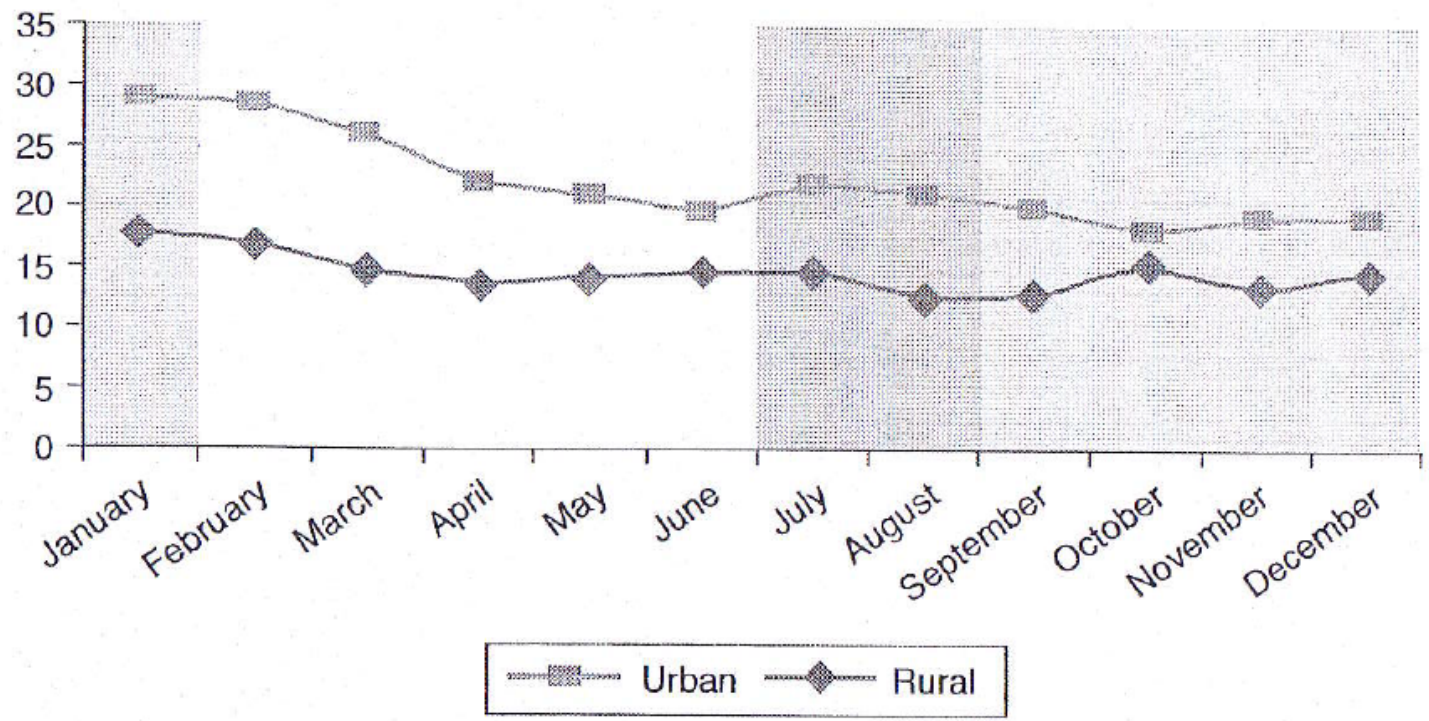


Figure 5.1 Average number of food item expenditures, Sierra Leone, 2011

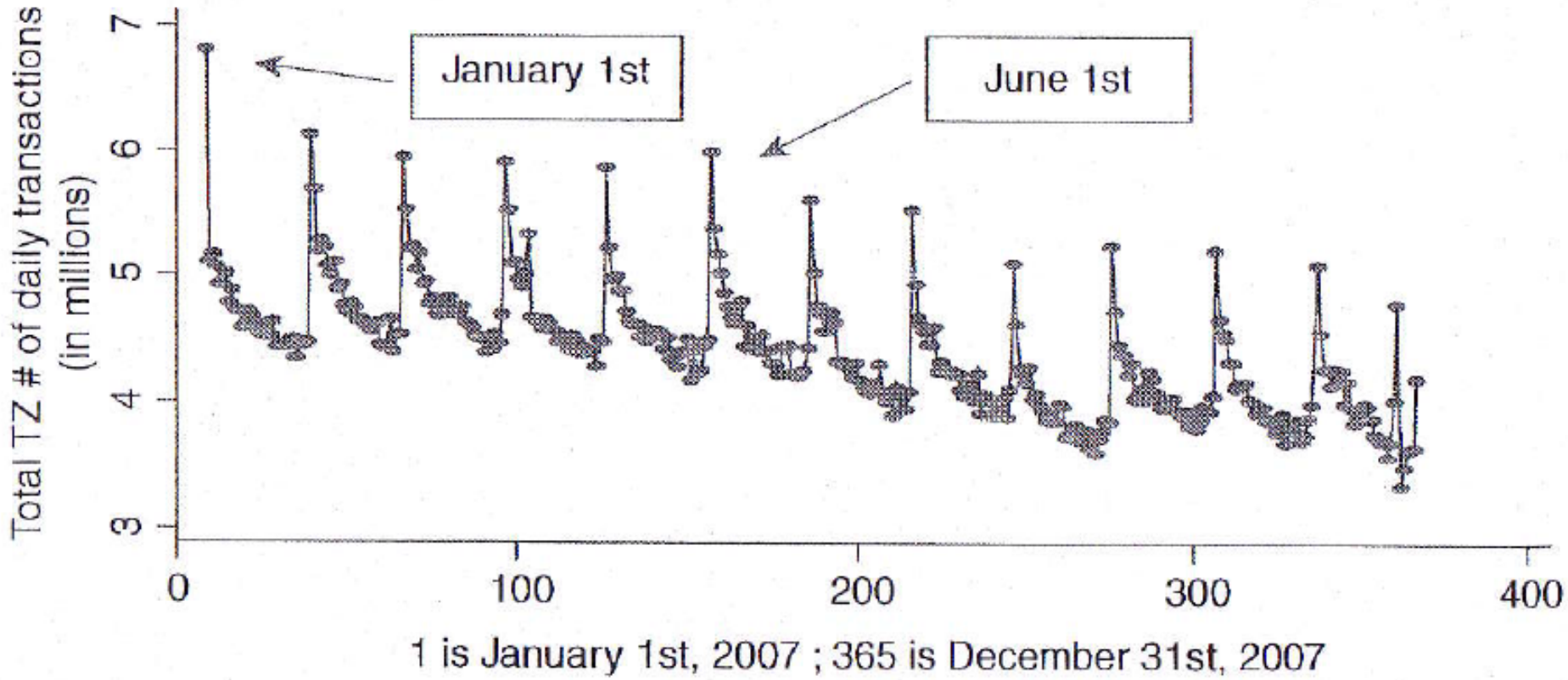


Figure 5.2 Total number of daily transactions in mainland Tanzania (in millions)

What is the mechanism?

- Not interviewee fatigue
  - Different samples at the end
- Fieldworkers getting to understand “the rules”
  - Avoiding extra work

Median total expenditure by month and week (unweighted)

Category	Week 1	Week 2	Week 3	Week 4	Total
Sep	299	192	185	258	**229**
Oct	310	182	164	274	**227**
Nov	292	200	175	251	**227**
Dec	285	212	242	371	**275**
Jan	222	187	191	281	**216**
Feb	345	220	194	280	**250**
Mar	292	182	174	267	**223**
Apr	306	197	185	257	**231**
May	288	188	178	298	**230**
Jun	291	167	153	250	**210**
Jul	288	182	167	279	**219**
Aug	288	173	160	257	**215**
Total	**289**	**189**	**180**	**276**	**229**

What is the mechanism in SA?

- Record fewer items?
- Make less effort to get into gated communities?

Conclusion

Diary method

- Is supposedly the “gold standard” for accuracy
- But is considerably more expensive
- AND there is now evidence (from both SA and elsewhere) that the quality of fieldwork may change over the year of gathering data

- BUT
  - Changing the method would break the series again