

Scanner Data, Time Aggregation and the Construction of Price Indexes

Lorraine Ivancic¹, Kevin J.Fox² and W. Erwin Diewert³

¹ Centre for Applied Economic Research, University of New South Wales

² School of Economics and Centre for Applied Economic Research, University of New South Wales

³ Department of Economics, University of British Columbia

Time aggregation and scanner data

- Scanner data increasingly available
 - Contains highly detailed information on consumer purchases
 - Statistical agencies in Netherlands, Norway and Switzerland currently using scanner data
- Increasing number of ways data can be aggregated
- Existing literature shows time aggregation likely to be important (Reinsdorf (1999), Bradley et al (1997), de Haan and Opperdoes (1997), Dalen (1997))
- Limitation of existing studies: small number of product categories
 - Difficult to make generalisations about findings

Scanner data set

- Data collected by A.C. Nielsen
- Period covered: 02/02/97 – 26/04/98 (65 weeks)
- 111 stores located within the Brisbane area
- Item categories include:

Biscuits	Detergent	Margarine	Sugar
Bread	Frozen peas	Oil	Spreads
Butter	Honey	Pasta	Tin tomatoes
Cereal	Jams	Pet food	Toilet paper
Coffee	Juices	Soft drinks	

- Data aggregated to weekly data
- Additional information: description, EANAPN (unique identifier for each item)

Index number estimation

- Direct and chained indexes estimated
- Two types of chained indexes:
 - Flexible chained: basket of goods allowed to change
 - Fixed chain: basket of goods same as direct indexes
- Types of indexes estimated:
 - Laspeyres, Paasche, Fisher, Törnqvist and Walsh

Aggregation methods

- Average price and total quantities aggregated at:
 - weekly,
 - monthly; and
 - quarterly intervals.
- Items treated as:
 - different goods if they were not located in the same store (ie. *no* item aggregation over stores); and
 - the same good no matter which store they were in (ie. item aggregation over stores).
- In total 6 different aggregation methods

Index number results: Laspeyres flexible chained indexes (Base = 100)

	Quart	Week		Quart	Week
Biscuits	100.65	318.33	Margarine	111.94	13897.59
Bread	106.16	3146.25	Oil	94.10	132.41
Butter	102.80	193.00	Pasta	101.97	790.75
Cereal	102.36	361.49	Pet food	102.53	263.49
Coffee	113.72	543.34	Soft drinks	111.82	46575.10
Detergent	103.50	227.96	Spreads	105.51	140.14
Frozen Peas	101.92	300.51	Sugar	107.20	176.18
Honey	105.05	128.45	Tin tomatoes	103.15	212.26
Jams	101.40	294.13	Toilet paper	107.31	11955.97
Juices	103.51	821.30			

Index number results: Fisher flexible chained indexes (Base = 100)

	Week	Quart.		Week	Quart.
Biscuits	79.86	97.91	Margarine	79.35	104.06
Bread	99.32	104.00	Oil	80.89	91.33
Butter	96.53	100.83	Pasta	77.68	100.11
Cereal	84.47	100.18	Pet food	95.04	100.49
Coffee	87.79	110.30	Soft drinks	74.28	104.01
Detergent	91.99	102.06	Spreads	99.66	104.39
Frozen Peas	89.48	100.55	Sugar	89.90	106.14
Honey	101.29	104.21	Tin tomatoes	88.12	101.32
Jams	81.48	99.93	Toilet paper	79.86	100.43
Juices	90.94	100.76			

Index number results: Fisher direct indexes (Base=100)

	Week	Quart.		Week	Quart.
Biscuits	101.01	99.01	Margarine	103.88	103.85
Bread	105.27	103.72	Oil	86.16	91.95
Butter	99.64	100.63	Pasta	102.78	100.88
Cereal	103.22	100.41	Pet food	102.84	100.88
Coffee	113.67	110.41	Soft drinks	107.09	104.04
Detergent	104.14	102.68	Spreads	106.29	104.29
Frozen Peas	101.42	100.82	Sugar	106.97	106.56
Honey	105.06	104.52	Tin tomatoes	100.47	101.70
Jams	101.53	101.18	Toilet paper	94.45	99.86
Juices	101.55	101.45			



Index number results: summary

- Time aggregation has huge impact on all index number estimates
 - Expect this for chained or non-superlative BUT
 - Even direct and/or superlative indexes affected
- Weekly chained indexes often unreasonable and exhibit large amount of chain index drift
- Unclear how much of quarterly and monthly chained indexes is drift and how much is actual price change
- Want drift free estimate of price change - may get us closer to 'truth'

GEKS method

- Multilateral index typically used for cross country comparisons
 - Satisfies circularity or transitivity
- GEKS: geometric mean of all ratios of bilateral Fisher indexes where each entity is taken in turn as base

$$\text{GEKS}_{jk} = \prod_{l=1}^M \left[P_{jl} / P_{kl} \right]^{1/M}$$

- P_{jl} = Fisher index between country j and l, $l=1\dots m$
- P_{kl} = Fisher index between country k and l, $l=1\dots m$
- GEKS satisfies multiperiod identity test and is free of chain index drift
- Modify formula: replace countries with time periods

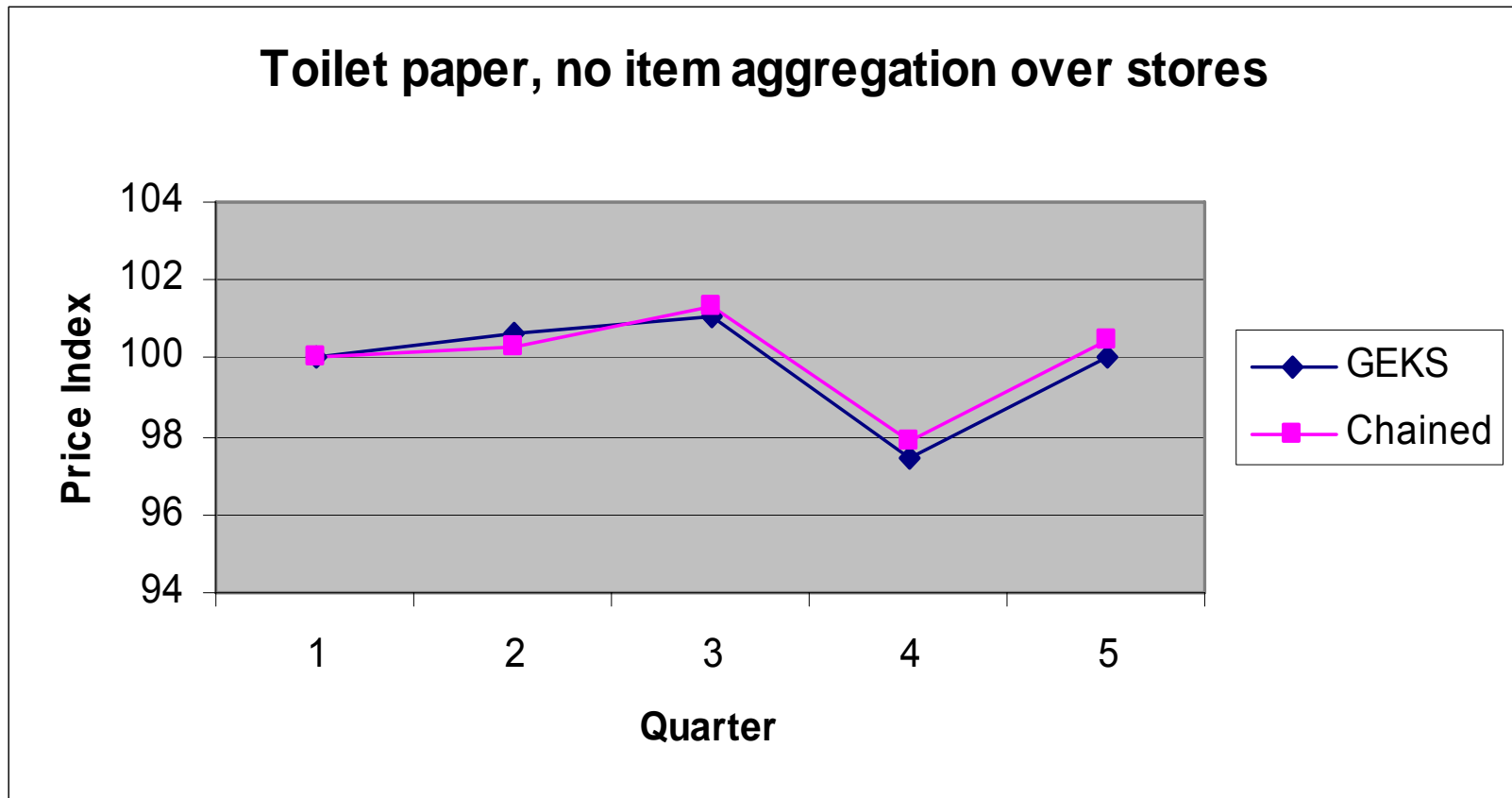
Calculating GEKS

- Example: for monthly index:
 - Compute Fisher ideal indexes that compare all n months with the base month
 - Use data on all items which appears in both periods for Fisher indexes (maximise matching across time)
 - From this we obtain n separate monthly time series
 - Take the geometric average of the n time series
 - Resulting price series is free of drift

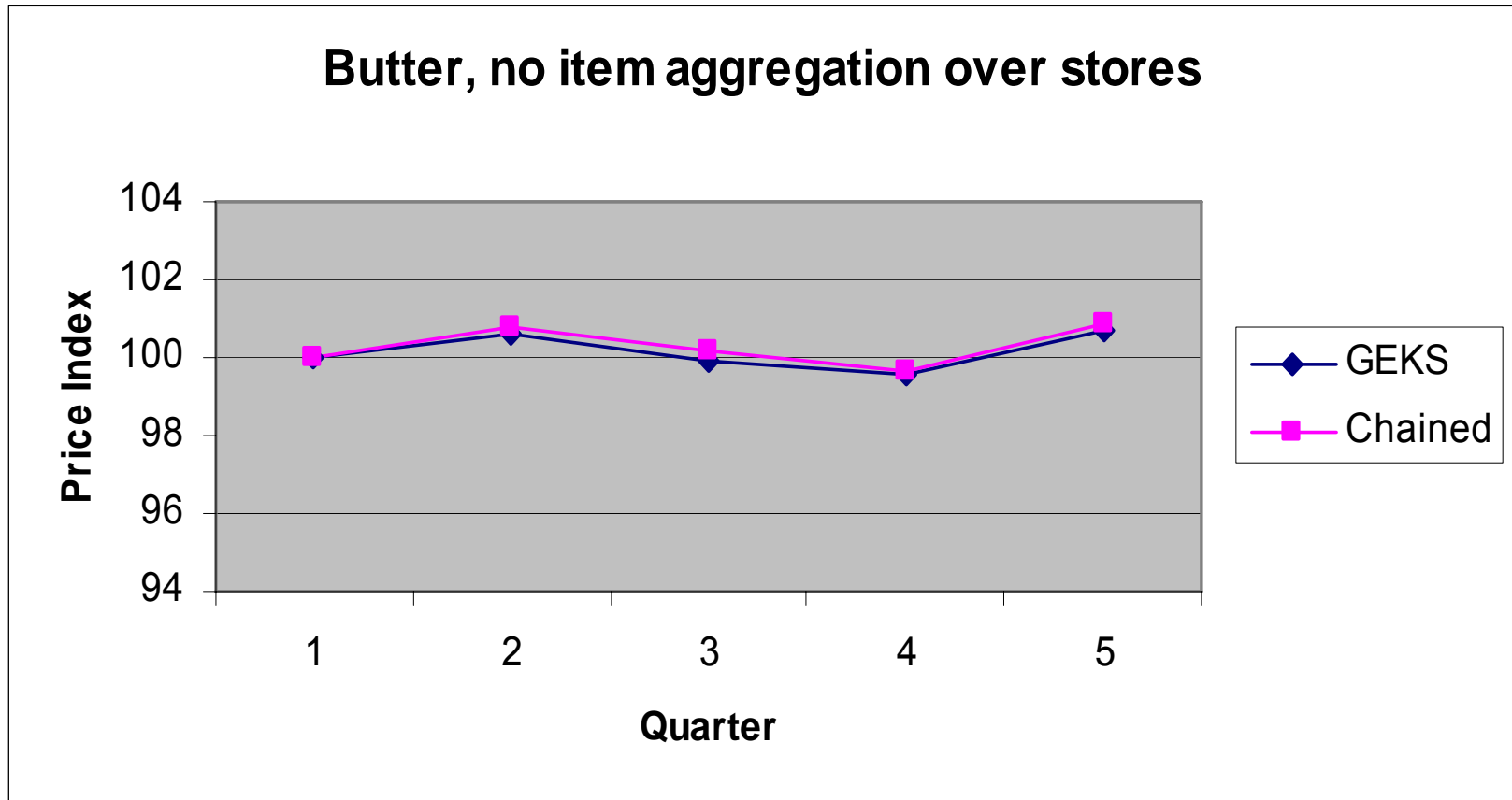
GEKS estimation method

- GEKS indexes estimated for 2 item categories:
 - toilet paper and butter
- GEKS indexes estimated between periods:
 - Quarterly: 1-2, 1-3, 1-4 and 1-5
 - Monthly: 1-2, 1-3 ...1-14 and 1-15
- Aggregation methods:
 - quarterly and monthly time aggregation
 - item aggregation over stores and no item aggregation over stores

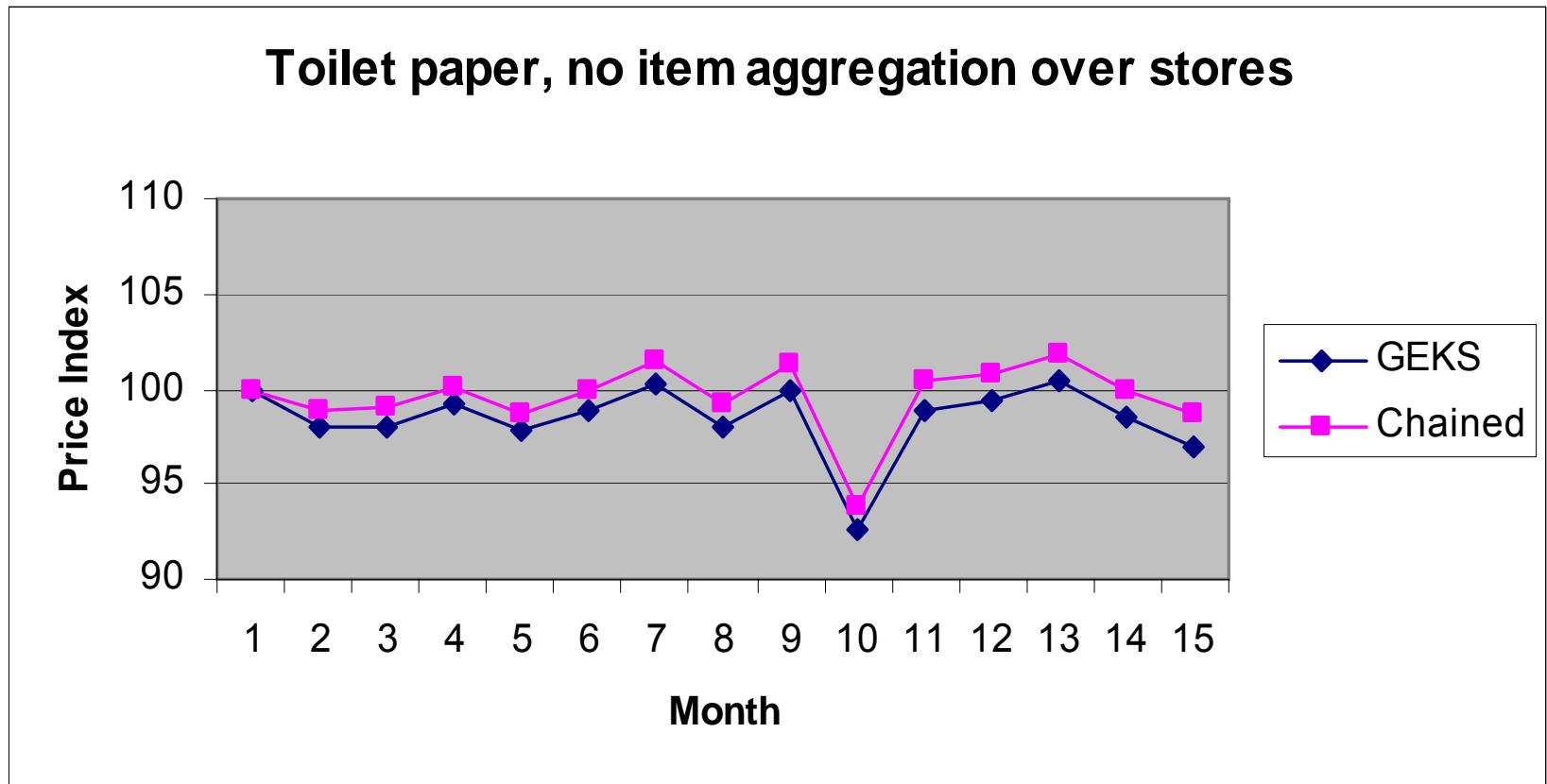
Quarterly comparisons



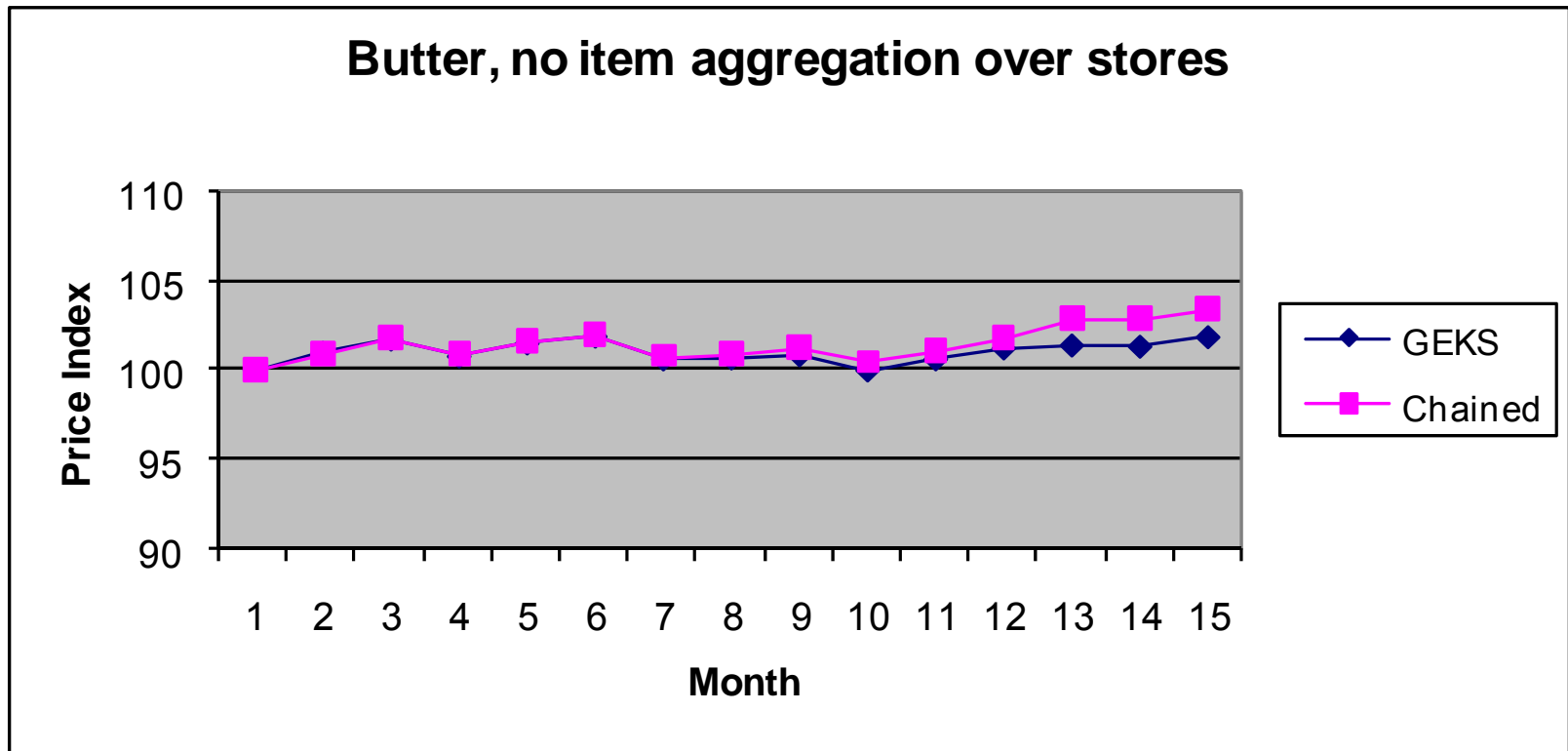
Quarterly comparisons



Monthly comparisons



Monthly comparisons



Rolling window GEKS

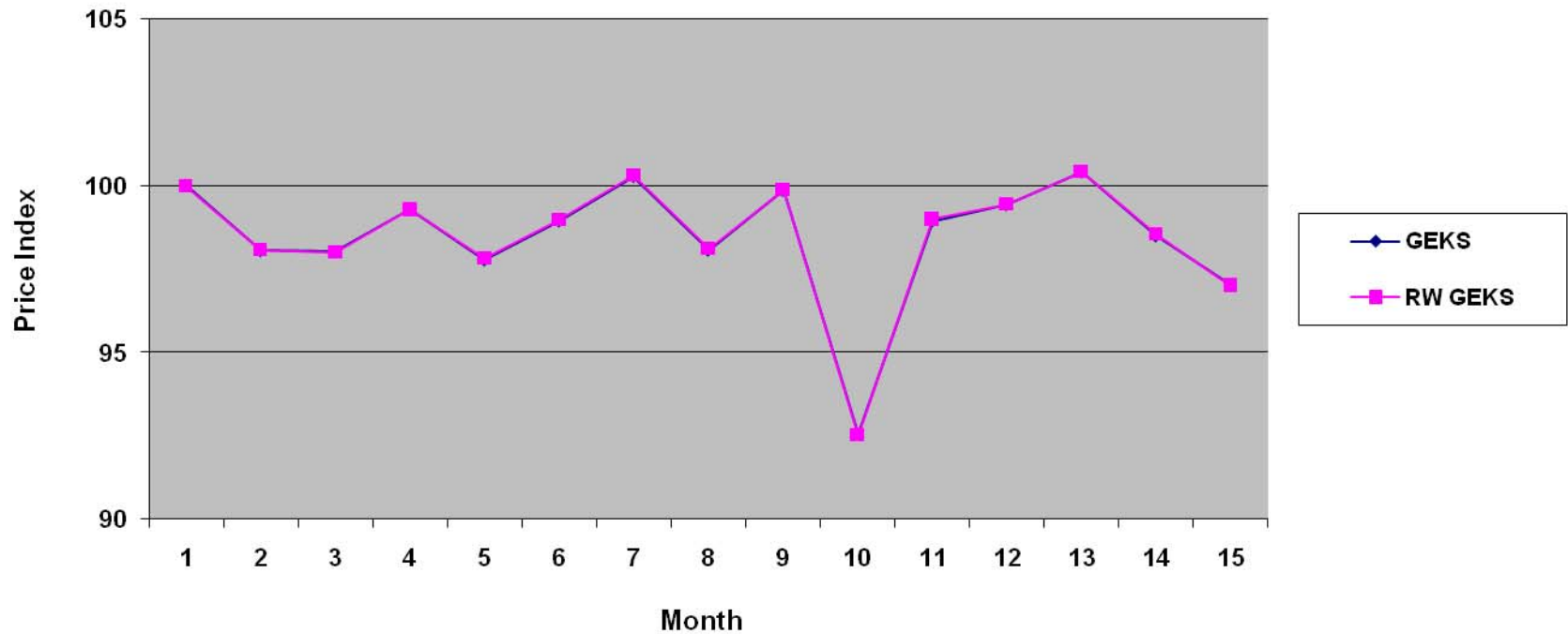
- Drawback GEKS: when new period of data available all previous parities are recomputed
 - Unacceptable for statistical agency
- Propose Rolling Window GEKS (RWGEKS)
 - Use rolling window to continuously update price series
 - No need to revise previous period parities
 - 'Natural' choice for window: 13 months
 - 13 month window → Rolling Year GEKS (RYGEKS)

Calculating RYGEKS

- For monthly RYGEKS index:
 - Compute GEKS index between month 1 – 13 as done previously ($GEKS_{1-13}$)
 - For next entry (chain link) in price series, month 1 is dropped from rolling window and month 14 is added to our rolling window
 - GEKS index is then calculated between periods 13-14 using all data from months 2-14 ($GEKS_{13-14}$)
- To obtain RYGEKS index for month 14:
 - $RYGEKS(14) = GEKS_{1-13} \times GEKS_{13-14}$

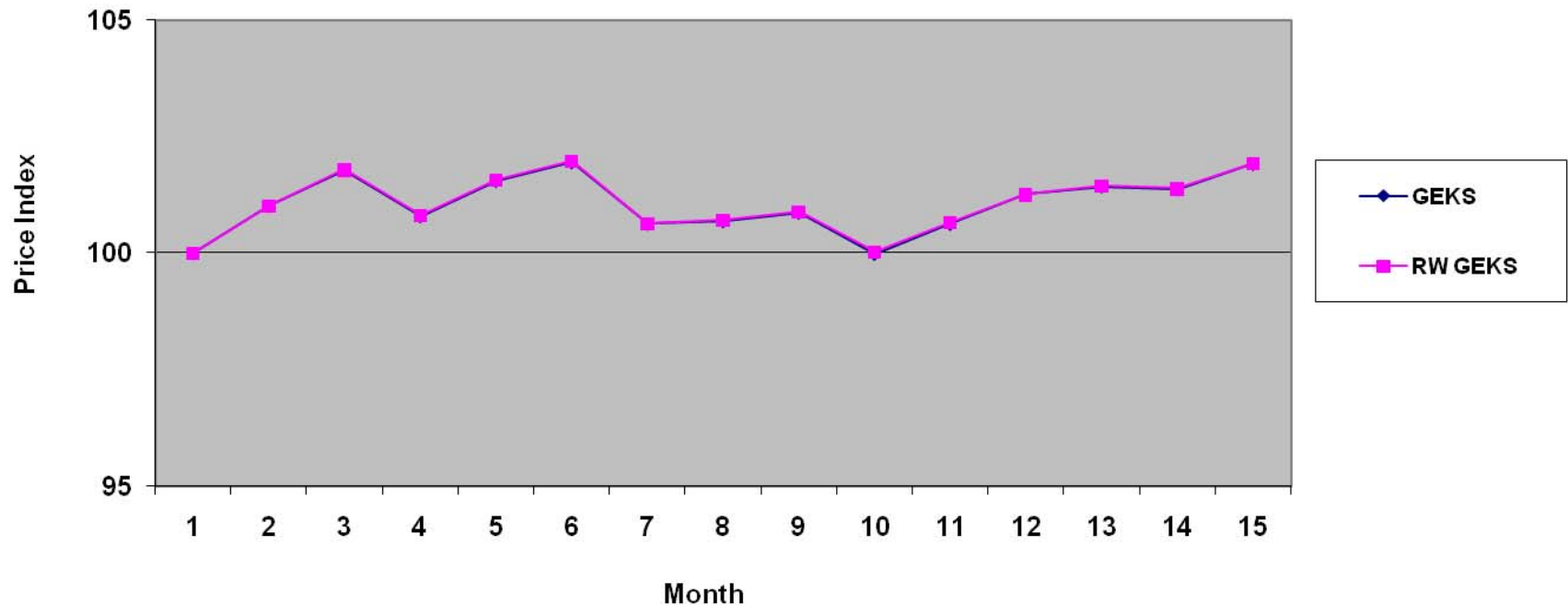
GEKS and RYGEKS: Toilet paper

Toilet paper, no item aggregation over stores



GEKS and RYGEKS: Butter

Butter, no item aggregation over stores



GEKS and official CPI figures

- Australian CPI: quarterly CPI estimates
- GEKS indexes :
 - Scanner data for Brisbane (Official CPI: Australia)
 - Match 6 scanner data item categories with official CPI sub heading groups
 - 4 quarters of scanner data matched with official series
 - Calculate quarterly GEKS indexes (series too short for RYGEKS)
 - 2 aggregation methods:
 - Item aggregation over stores
 - No item aggregation over stores

ABS CPI and GEKS indexes (April 97 – March 98)

	GEKS Indexes		Official CPI figures
	Item aggregation over stores	No item aggregation over stores	
Cereal	100.09	100.21	97.51
Bread	101.47	101.40	102.41
Butter	99.25	99.86	99.89
Juices	100.10	100.30	100.99
Sugar	98.08	98.34	105.35
Soft drinks	99.43	99.64	103.43
Geomean	99.73	99.95	101.56

Results: ABS CPI and GEKS indexes

- Very little difference between 2 GEKS series
- Five out of six item categories: GEKS less than official CPI figures
- Some differences between official series and GEKS quite large, eg. soft drinks: approx 4%.
- Difference in Geomean of official CPI and GEKS:
 - No item agg. over stores: 1.61 percentage points
 - Item agg. over stores: 1.83 percentage points
- Results indicate may be substantial amount of substitution bias in official figures

Country Product Dummy (CPD) Method

- Another multilateral index method
- CPD method is transitive
- CPD: obtain standard errors on coefficients
- Standard CPD model:

$$\ln P_{ic} = \sum_{i=1}^I \pi_i D_i + \sum_{c=1}^C \eta_c D_c + \varepsilon_{ic}$$

Where: $\ln P_{ic}$ = natural logarithm of price item i in country c

D_i = dummy variable for item i , where $i=1\dots I$

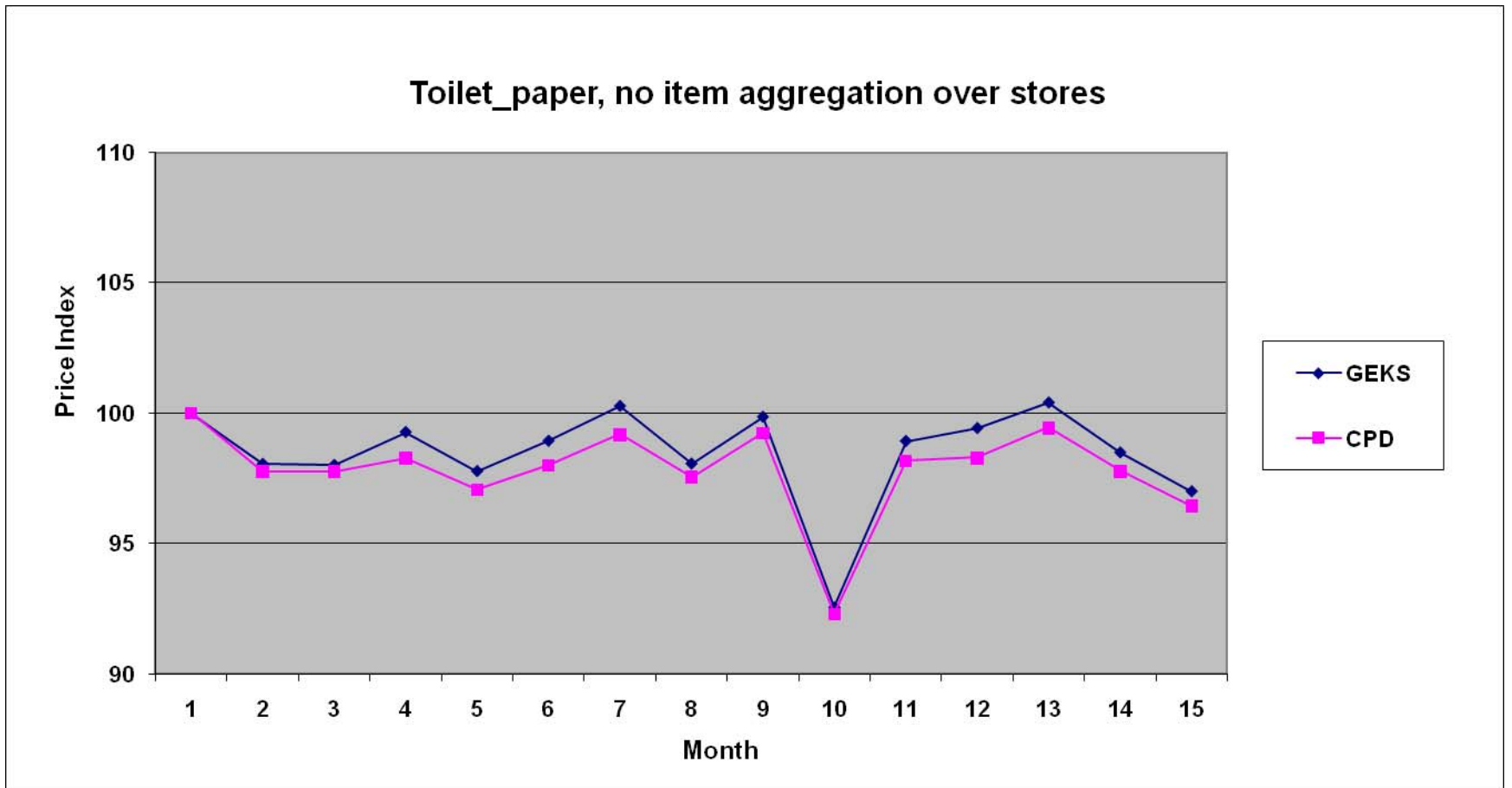
D_c = dummy variable for country c , where $c=1\dots C$



CPD method (cont.)

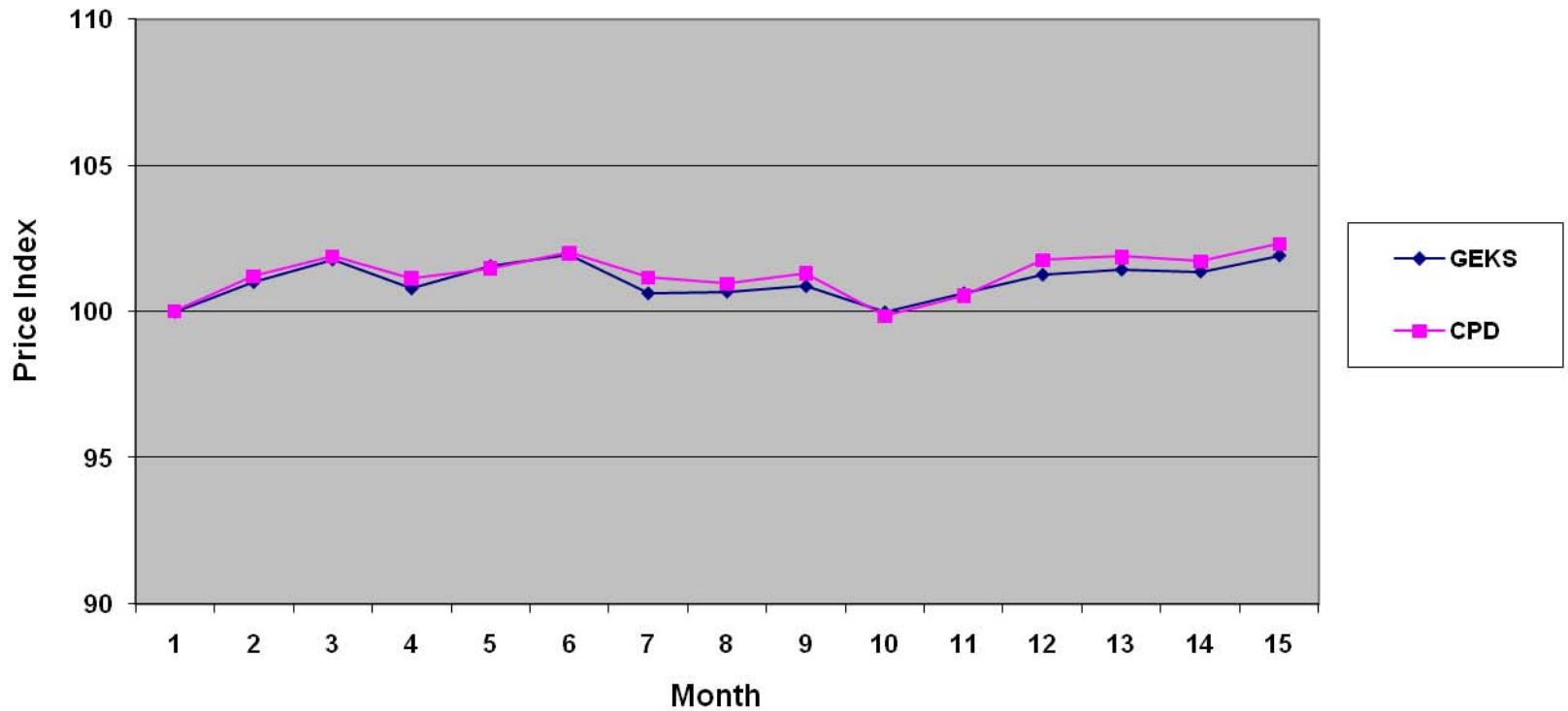
- We estimate CPD models for two item categories:
 - butter and toilet paper
- Weights included in our model
 - observations weighted by expenditure share
- Sample size varies across time so new items allowed to enter sample
- Aggregation methods:
 - monthly time aggregation
 - item aggregation over stores; and NO item aggregation over stores

CPD and GEKS: Toilet paper



CPD and GEKS: Butter

Butter, no item aggregation over stores





Results CPD and GEKS

- Results very similar for GEKS and CPD methods
- CPD appears to be a good alternative to GEKS if standard errors are required



Conclusions

- Weekly aggregation not appropriate when scanner data is used
- GEKS method can provide us with drift free estimate of price change
 - Possible benchmark estimate of price change
- Initial results suggest quarterly or monthly time aggregation may be appropriate
- GEKS results suggest possible bias in official CPI figures may not be negligible