The CPI Variance System

Owen J. Shoemaker
U.S. Bureau of Labor Statistics
2 Massachusetts Ave NE
Washington DC 20212

BLS
U.S. BUREAU OF LABOR STATISTICS
125 YEARS

OVERVIEW

- Introduction to CPI's Highly Stratified AREA & ITEM Structures
- CPI's Variance Estimation System(s)
  - STRATIFIED RANDOM GROUP (SRG)
  - JACKKNIFE (for SRCs only)
- How CPI Variances are Calculated
- CE/CPI Weights and Bridging Factors across Pivot Years

125 YEARS BLS

CPI's AREA STRUCTURE

- 38 Index Areas (Large PSUs)
  - 31 A-Sized Metropolitan Areas (A102, etc) which are **Self-Representing** Index Areas
  - 4 B-Sized Regional Areas (X100, X200, X300, X499) and 3 C-Sized Regional Areas (D200, D300, D400) which are **Non-Self-Representing** Areas (East, Midwest, South, and West Regions), each consisting of at least 4 "little PSUs", each representing a Stratum
  - Plus 16 larger Area Aggregates, including All-US (0000), The 4 Regions, All the A's, etc.

125 YEARS BLS

CPI's ITEM STRUCTURE

- 211 Basic Item Strata
- 72 Expenditure Classes (EC's)
- 8 Major Groups
  - APPAREL, EDUC/COMM, FOOD, HOUSING, MEDICAL, RECREATION, TRANSPORTATION, OTHER GOODS & SERVICES
- All Items (SA0) plus 57 Higher Level Aggregates including SA0L1E (All Items Less Food and Energy – CORE)
- 87 Special Relative Calculations (SRC's)

125 YEARS BLS

CPI's ESTIMATION SYSTEM Lower Level

- The PRC produces 8,018 (38 x 211) Full Sample 1-Month Price Relatives, plus 20,467 (97 x 211) Replicate PRELs
  - 2 Replicates in each A-Sized City (with NYC, Chicago & LA getting 2 extra Reps)
  - 4 Reps in X100, 5 Reps in X200, 11 Reps in X300, and 3 Reps in X499
  - D200, D300 & D400 each get 2 Reps
  - BLS uses a Modified Laspeyres but mostly a Geomeans formula to calculate the PRELs

Nature of CPI Replicates

- Each replicate contains and so represents the entirety of the ITEM (large or small) that is being estimated
- In the Self-Representing ("A" Cities) INDEX AREAs, the 2 or 4 replicates are *randomly* selected samples (done at the basic Item Stratum level) from the Full Sample
- In the Non-Self-Representing INDEX AREAs ("X"s and "D"s), each replicate is the composite estimate of two paired ODD and EVEN "little" PSUs

125 YEARS BLS

CPI's ESTIMATION SYSTEM Higher Level

- The PRC then uses these 28,485 PRELs to update their corresponding indexes: IX(a,i,r,t) = IX(a,i,r,t-1) x PREL(a,i,r,t)
- The CPI Index and Variance Production Systems then convert these updated indexes into Cost Weights (CW=IX*AW) and proceed to produce nearly 160,000 1-, 2-, 6- and 12-Month Percent Price Changes & Standard Errors each month

Price Change Formula for Full Sample

$$PC(A, I, t, t-k) = \left(\frac{CW(A, I, 00, t)}{CW(A, I, 00, t-k)} - 1 \right) \times 100$$

CW(A,I,00,t) is the cost weight for AREA=A, ITEM=I, REPLICATE=00

CW(A,I,00,t-k) is corresponding CW in month t–k (k= 1, 2, 6, or 12)

Price Change Formula for Replicates

$$PC[(A,I,00-(a,I,00)+(a,I,r),t,t\text{-}k] \quad =$$

$$\left(\frac{CW(A,I,00,t) - CW(a,I,00,t) + CW(a,I,r,t)}{CW(A,I,00,t-k) - CW(a,I,00,t-k) + CW(a,I,r,t-k)} - 1 \right) \times 100$$

$$= \left(\frac{\sum_{b \in A} CW(b,I,00,t) - CW(a,I,00,t) + CW(a,I,r,t)}{\sum_{b \in A} CW(b,I,00,t-k) - CW(a,I,00,t-k) + CW(a,I,r,t-k)} - 1 \right) \times 100$$

125 YEARS BLS

Stratified Random Group (SRG) Formula

$$Var_{SRG}(A,I,t,t-k) = \sum_{a\in A}\frac{1}{N_a(N_a-1)}\times$$

$$\sum_{r\in R_a}\left(PC\left[(A,I,00)-(a,I,00)+(a,I,r),t,t-k\right]-PC(A,I,t,t-k)\right)^2$$

and the Standard Error of the k-month is simply computed as:

SE (A,I,t,t-k) = SQRT (Var(A,I,t,t-k))

SRG's Provenance

- The Stratified Random Group (SRG) Methodology was developed here at BLS, by BLS, in the mid-90's
- Credit goes to Rick Valiant, along with Sylvia Leaver, with Bob Fay also contributing. Also David Swanson.
- SRG is a variation on Kirk Wolter's Random Group Variance Estimator
- SRG is an equivalent and similar methodology to Stratified Jackknife

SRG is (mostly) equivalent to Jackknife Methodology

$$For\ \hat{\theta} = f(\theta)\ with\ \ f\ \ linear,$$

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2 = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i^* - \bar{\hat{\theta}}\right)^2$$

$$For\ \hat{\theta} = f(\theta)\ with\ \ f\ \ non-linear,$$

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{\theta}_i - \bar{\hat{\theta}}\right)^2 \cong \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i^* - \bar{\hat{\theta}}\right)^2$$

125 YEARS BLS

Variance Calculation Example (A101 NYC Metro – SEFA Cereals)

Cost Weight Data

AREA	ITEM	REPLCATE	CW0107	CW0106
A101	SEFA	00	153,894,954.78	156,243,961.14
A109	SEFA	00	41,314,419.20	42,254,602.70
A109	SEFA	01	50,420,364.82	58,542,449.85
A109	SEFA	02	32,585,409.36	32,585,409.36
A109	SEFA	03	48,131,270.91	47,933,712.92
A109	SEFA	04	50,746,734.01	44,358,709.13
A110	SEFA	00	51,882,684.13	53,291,507.00
A110	SEFA	01	59,309,461.69	59,513,890.19
A110	SEFA	02	44,940,684.09	46,598,174.12
A111	SEFA	00	60,697,851.45	60,697,851.45
A111	SEFA	01	59,643,508.67	59,643,508.67
A111	SEFA	02	60,902,708.76	60,902,708.76

125 YEARS BLS

Variance Calculation Example (Cont.)

Replicate Price Changes

a	r	PC_REP = PC((A-a,00)+(a,r),I,t,t-1)
A109	01	(((153,894,954.78 - 41,314,419.20 + 50,420,364.82) / (156,243,961.14 - 42,254,602.70 + 58,542,449.85)) – 1)*100 = **-5.52415**
A109	02	(((153,894,954.78 - 41,314,419.20 + 32,585,409.36) / (156,243,961.14 - 42,254,602.70 + 32,585,409.36)) – 1)*100 = **-0.96116**
A109	03	(((153,894,954.78 - 41,314,419.20 + 48,131,270.91) / (156,243,961.14 - 42,254,602.70 + 47,933,712.92)) – 1)*100 = **-0.74805**
A109	04	(((153,894,954.78 - 41,314,419.20 + 50,746,734.01) / (156,243,961.14 - 42,254,602.70 + 44,358,709.13)) – 1)*100 = **3.14447**
A110	01	(((153,894,954.78 - 51,882,684.13 + 59,309,461.69) / (156,243,961.14 - 53,291,507.00 + 59,513,890.19)) – 1)*100 = **-0.70452**
A110	02	(((153,894,954.78 - 51,882,684.13 + 44,940,684.09) / (156,243,961.14 - 53,291,507.00 + 46,598,174.12)) – 1)*100 = **-1.73699**
A111	01	(((153,894,954.78 – 60,697,851.45 + 59,643,508.67) / (156,243,961.14 – 60,697,851.45 + 59,643,508.67)) – 1)*100 = **-1.51364**
A111	02	(((153,894,954.78 – 60,697,851.45 + 60,902,708.76) / (156,243,961.14 – 60,697,851.45 + 60,902,708.76)) – 1)*100 = **-1.50145**

Variance Calculation Example (Finish)

Full-Sample Price Change

A	I	PC_FULL = PC(A,I,t,t-1)
A101	SEFA	((153,894,954.78 / 156,243,961.14) – 1)*100 = **-1.50342**

Variance Computations

a	r	PC_REP	PC_FULL	SQR_DIFF= (PC_REP-PC_FULL)**2	Na	SQR_DIFF/ / (Na(Na-1))
A109	01	-5.52415	-1.50342	= 16.16627	4	= 1.34719
A109	02	-0.96116	-1.50342	= 0.29405	4	= 0.02450
A109	03	-0.74805	-1.50342	= 0.57058	4	= 0.04755
A109	04	3.14447	-1.50342	= 21.60288	4	= 1.80024
A110	01	-0.70452	-1.50342	= 0.63824	2	= 0.31912
A110	02	-1.73699	-1.50342	= 0.05455	2	= 0.02728
A111	01	-1.51364	-1.50342	= 0.00010	2	= 0.00005
A111	02	-1.50145	-1.50342	= 0.00000	2	= 0.00000
						3.56593

VARIANCE = **3.56593**

STANDARD ERROR = SQRT(3.56593) = **1.88837**

125 YEARS BLS

Example: A Final Set & Published Medians

ITEM	AREA	MONTH	PC01	PC02	PC06	PC12	SE01	SE02	SE06	SE12
SA0	0000	201301	0.2960	0.0260	0.5135	1.5949	0.0320	0.0487	0.0732	0.0853
SA0	0000	201302	0.8190	1.1174	0.7759	1.9782	0.0335	0.0411	0.0703	0.0870
SA0	0000	201303	0.2612	1.0823	0.5902	1.4738	0.0319	0.0401	0.0681	0.0875
SA0	0000	201304	-0.1037	0.1572	0.5250	1.0633	0.0300	0.0390	0.0641	0.0782
SA0	0000	201305	0.1777	0.0738	1.1832	1.3618	0.0309	0.0457	0.0564	0.0731
SA0	0000	201306	0.2400	0.4181	1.6998	1.7542	0.0322	0.0490	0.0669	0.0707
SA0	0000	201307	0.0397	0.2798	1.4400	1.9608	0.0335	0.0405	0.0648	0.0741
SA0	0000	201308	0.1203	0.1601	0.7370	1.5186	0.0345	0.0406	0.0574	0.0741
SA0	0000	201309	0.1161	0.2366	0.5912	1.1849	0.0507	0.0600	0.0684	0.0784
SA0	0000	201310	-0.2573	-0.1415	0.4365	0.9638	0.0401	0.0661	0.0713	0.0882
SA0	0000	201311	-0.2044	-0.4612	0.0534	1.2373	0.0450	0.0473	0.0626	0.0807
SA0	0000	201312	-0.0088	-0.2132	-0.1949	1.5016	0.0357	0.0630	0.0772	0.0858
	MED	2013	0.118	0.159	0.591	1.488	0.034	0.047	0.068	0.080

JACKKNIFE Formula (for SRCs only)

$$PC_{FS}(A,I,t,t-k)=\left(\frac{CW(A,I,00,t)}{CW(A,I,00,t-k)}-1\right)\times 100$$

$$PC_{JK}(A-a,I,t,t-k)=$$

$$\left(\frac{CW(A,I,00,t)-CW(a,I,00,t)}{CW(A,I,00,t-k)-CW(a,I,00,t-k)}-1\right)\times 100$$

Only for Areas (A) *above* basic-level areas.
Items (I) are *below* Item-Stratum level .
These SRCs come with NO replicate estimates.

125 YEARS BLS

JACKKNIFE Formula (Cont.)

$$Var_{JK}(A, I, t, t-k) =$$

$$\frac{N_A - 1}{N_A} \sum_{a \in A} \left(PC_{JK}\left(A - a, I, t, t-k\right) - PC_{FS}\left(A, I, t, t-k\right)\right)^2$$

$$and \quad SE(A, I, t, t-k) = SQRT(Var(A, I, t, t-k))$$

These JACKKNIFE estimates for the SRCs are upwardly biased. (Unlike the *unbiased* Stratified Jackknife and SRG estimates.)

125 YEARS
BLS

BRIDGING FACTOR Across Pivot Months

- CPI's Cost Weights are updated every **two years**, new usage beginning in every even-year January (with the previous December being the Pivot Month)
- Whenever any 1-, 2-, 6- or 12-month variance calculation *spans* this December pivot month, a **factor adjustment** is required for *each and every* full-sample and replicate price change calculation
- **Bridge Factor: $CW_{A,I,r,OLD} / CW_{A,I,r,NEW}$**

125 YEARS BLS

BRIDGING FACTOR Defined & Constructed (I)

$$Let \quad CW_{A,I,r,OLD} = IX_{A,I,r,DEC} \times AGGWT_{A,I,r,OLD}$$

$$and \quad CW_{A,I,r,NEW} = IX_{A,I,r,DEC} \times AGGWT_{A,I,r,NEW}$$

$$so \quad AGGWT_{A,I,r,OLD} = AGGWT_{A,I,r,NEW} \times \frac{CW_{A,I,r,OLD}}{CW_{A,I,r,NEW}}$$

$$PR_{A,I,r,t,t-k} = \frac{IX_{A,I,r,t}}{IX_{A,I,r,t-k}} = \frac{AGGWT_{A,I,r,t=NEW}}{\frac{\sum_{a,i} IX_{a,i,r,t-k} \times AGGWT_{a,i,r,t-k}}{AGGWT_{A,I,r,t-k=OLD}}}$$

(over)

BRIDGING FACTOR Defined & Constructed (II)

(*continuing*)

$$= \frac{\sum_{a,i} CW_{a,i,r,t}}{\sum_{a,i} CW_{a,i,r,t-k}} \times \frac{AGGWT_{A,I,r,NEW} \times \frac{CW_{A,I,r,OLD}}{CW_{A,I,r,NEW}}}{AGGWT_{A,I,r,NEW}}$$

$$= \frac{\sum_{a,i} CW_{a,i,r,t}}{\sum_{a,i} CW_{a,i,r,t-k}} \times \frac{CW_{A,I,r,OLD}}{CW_{A,I,r,NEW}}$$

125 YEARS BLS

Contact Information

Owen J. Shoemaker
Mathematical Statistician
Office of Prices and Living Conditions/
Division of Price Statistical Methods
shoemaker_o@bls.gov

BLS
U.S. BUREAU OF LABOR STATISTICS
125 YEARS