Deflating Statistical Series: An Example Using Aggregate U.S. Demand for Textile End-Use Categories

By Thomas M. Bell, Joseph M. Roop, and Cleve E. Willis*

Analysts frequently adjust price, income, or other data to eliminate the influence of inflation or differences in size The authors of this article examine economic and statistical reasons for deflating time-series and cross-sectional data prior to estimating demand relations Signs and magnitudes of regression coefficients change when aggregate demand equations for textiles are estimated from time-series data Questions of heteroskedasticity, multicollinearity, and homogeneity are addressed The demand equations are disaggregated by end use category—apparel, household, and industrial demand

Keywords

Deflation
Demand analysis
Econometrics
Textile demand

Analysts often deflate data on prices, income, and other variables to eliminate the effects of inflation or household size in demand analyses In time-series analyses, for example, they frequently deflate consumption by population, and investment by volume of sales In cross-sectional studies, household income is often deflated by size of household and sales by size of firm

Our purpose here is to present some reasons for deflating statistical series and to demonstrate the results —namely, that signs and magnitudes of regression coefficients change—when we use aggregate demand equations and time-series data for textiles

ECONOMIC REASONS FOR DEFLATING

Variables that shift demand functions must be used if we are to isolate price-quantity relationships (*19*) [1] To measure consumer demand from time-series data, Foote divides shift variables into four classes (1) consumer income or other measures of the general level of demand on a national basis, (2) the general price level (3) supplies or prices of competing products, and (4) population (*6*, p 27)

Let us focus on Foote's second category, the general price level Assume that demand is homogeneous of degree zero for all prices and income, as economic theory suggests We impose this assumption by deflating each price and income variable by the general level of prices [2] We express demand for commodity y as

$$y = b_0 + b_1 X_1 + b_2 X_2 + u \quad (1)$$

where

X_1 = own price
X_2 = consumer income
b_i = unknown parameters, and
u = error term

The use of real income and relative prices is the "Marshallian" method, alternatively, we could use normalized prices (*9*) or a "mixed" demand curve specification (*15*) Utility theory requires that competing- and complementary-good prices be included We ignore them here to simplify the presentation However, this argument precedes the functional form

From economic theory, price of y relative to other commodity prices influences consumption of y, thus X_1 should be the relative price of y This is the verbal statement of homogeneity in the multigood world

Operationally, we obtain a measure of the relative price of y by deflating its absolute price by an index of other prices The relevant price becomes $*X_1 = X_1/K$, and the relative income measure is $*X_2 = X_2/K$ The original variables are expressed in nominal terms and K is an index of the general price level, such as the consumer price index (CPI)

Deflating by an index that contains the price of the dependent variable makes the resultant regression coefficients subject to bias Bias is also introduced when the index is included as a separate variable

*Thomas M Bell is senior commodity analyst with Merrill Lynch, and Joseph M Roop is senior economist with Evans Economics, Inc Both Bell and Roop were with ESCS when this article was prepared Cleve E Willis is a professor in the Department of Food and Resource Economics, University of Massachusetts

[1] Italicized numbers in parentheses refer to items in References at the end of this article

[2] We assume that (a) the good is relatively unimportant in the consumer's budget or that (b) the price movements of substitutes correspond approximately with the general price level

Analysts frequently adjust price, income, or other data to eliminate the influence of inflation or differences in size

This has led some investigators to construct special index numbers which eliminate the price(s) of the good(s) included in the analysis and to use these to deflate the price variables included in the study Obviously, the unadjusted measure of the general price level should generally be used to deflate the income variables

Brennan suggests that it is often better to deflate by some other index than that of all prices (2, p 379) In the demand analysis for an agricultural commodity, Brennan deflates the price in question by an index of agricultural prices only Further, if only one other commodity is a strong substitute for the good in question, it may be desirable to deflate the own price by the price of that substitute No unambiguous rule for the choice of the appropriate index can be given This choice is determined by the investigator's judgment and knowledge of the behavior of the subject being studied and the economic theory involved

STATISTICAL OR ECONOMETRIC REASONS FOR DEFLATING

Some econometric considerations affect the choice of a deflator Karl Pearson's early work on ratios having a common denominator showed that correlations between ratios can reflect spuriously high estimates of the relationship between the numerators (*13*) Kuh and Meyer showed that correlation among deflated series may also be spuriously low (*12*) The question of spurious correlation does not arise, of course, if the maintained hypothesis is in ratios Kuh and Meyer further demonstrated two necessary and sufficient conditions for the correlation of ratios to yield correct estimates of the undeflated partial correlations These are that (1) the coefficient of variation (the ratio of the standard deviation to the mean) of the deflating variable is small, and (2) the variables deflated are linear homogeneous functions of the deflator (*12*, p 405) The degree of bias depends on the relative size of $r(X,Y)$ compared to $r(X,Z)$ and $r(Y,Z)$, where Z is the deflator and $r(\ )$ is the correlation operator Hence, when cross-sectional data are deflated because of size, economic relationships will probably approximate the homogeneity requirement, so the ratio estimates generally should not be seriously biased

Another focus of attention is the spherical attributes of the residuals If the usual homoskedasticity is assumed for the undeflated series, deflating leads to heteroskedasticity because deflation of the included variables transforms the error term The assumption of homoskedasticity is seldom appropriate for undeflated cross-sectional data

Small observations are typically associated with small variances and large observations with large variances David and Neyman demonstrated that least squares produce efficient unbiased estimates only if the residual sum of squares to be minimized is appropriately weighted (*4*) That is, we assume the usual Markov assumptions are met and the variance of the conditional distribution of the dependent variable is a weighted average of the unknown population variance (with weights w) Then, the most efficient unbiased estimate of the regression parameters (b_i) are produced by minimizing $(w(y - X_b))^t(w(y - X_b))$ The matrix w is diagonal with elements $w_i^{-1/2}$, in the homoskedastic case, w is an identity matrix This derivation of b is an Aitken generalized least squares estimator (*10*, p 214)

Suppose the simple deflation is such that $w_i = D_i^2$ (D_i the deflator) Deflating yields efficient, unbiased estimators when the undeflated residuals are heteroskedastic At worst, deflation will usually be superior to assuming falsely that a constant unitary weight is appropriate (see *12*, p 407, *3*, *11*) A further advantage is that extreme observations will have less effect on the estimation

Deflating statistical series to achieve a favorable specification of an econometric model may be appropriate for both theoretical and empirical reasons Deflating an otherwise spherical relationship may induce heteroskedasticity although cross-sectional data frequently need to be deflated Multicollinearity is also affected—most of the consequences are wellknown (*10*, p 160, *14*, pp 46-52, and *18*, pp 127-128) The actual size and comparisons of $r_i(\ )$ determine the magnitude of the bias of partial correlation introduced when deflation is used

AN EXAMPLE [3]

We now apply the deflating method by considering demand for three categories of textiles by end uses (apparel, household, and industrial) Each end use includes four fiber types (noncellulosic, cellulosic, cotton, and wool), with estimates of fiber content of purchases expressed in cotton equivalent pounds Main explanatory variables of these categories are nominal disposable income, the end-use price index, an implicit deflator for all goods except the end use in question, and population (Fiber types and classifications appear in appendix table 1, data are in appendix table 2 Problems of quality changes and aggregation are ignored)

In functional form, the demand for each end-use category is expressed as

w_i – ½

$$Q_i = f(CPI_i, PD_i, INCN, POP),$$

i = A, H, I, end-use category) (2)

where

[3] This example does not illustrate the heteroskedasticity arguments For a summary of tests for homoskedasticity, see (*10*, pp 214-221)

Table 1—Correlation matrix

Variable	QA	QH	QI	CPIA	CPIH	WPII	PDA	PDH	PDI	INCN
QA	1 00	0 97	0 60	0 83	0 73	0 67	0 81	0 82	0 84	0 87
QH	97	1 00	50	92	83	79	89	89	91	93
QI	60	50	1 00	24	12	10	21	22	25	29
CPIA	83	92	24	1 00	97	94	99	99	99	99
CPIH	73	83	12	97	1 00	.98	99	99	98	97
WPII	67	79	10	94	98	1 00	96	96	95	94
PDA	81	89	21	99	99	96	1 00	1 00	1 00	99
PDH	82	89	22	99	99	96	1 00	1 00	1 00	99
PDI	84	91	25	99	98	95	1 00	1 00	1 00	1 00
INCN	87	93	29	99	97	94	99	99	1 00	1 00
POP	95	94	48	88	80	74	88	88	90	91
QAPC	99	94	66	76	64	59	74	74	76	80
QIPC	– 41	– 48	44	– 64	– 67	– 62	– 68	– 67	– 66	– 64
QHPC	97	1 00	52	91	81	77	87	88	89	92
INCNPC	86	93	28	99	97	95	99	1 00	1 00	1 00
INCRA	96	99	46	94	87	82	92	93	94	96
INCRH	96	99	46	94	87	82	93	93	94	96
INCRI	94	98	41	96	90	86	95	95	95	97
CPIAR	– 76	– 75	– 23	– 80	– 83	– 80	– 88	– 88	– 88	– 87
CPIHR	– 94	– 94	– 45	– 88	–.80	– 75	– 89	– 89	– 91	– 91
WPIIR	– 92	– 88	– 47	– 79	– 71	– 61	– 80	– 80	– 83	– 83
CV	21	46	07	17	15	14	23	23	22	48

* Variable definitions in addition to those in table 2 are

QAPC=QA/POP
QIPC=QI/POP
QHPC=QH/POP

INCNPC=INCN/POP
INCRA=INCNPC/PDA
INCRH=INCNPC/PDH
INCRI=INCNPC/PDI

CPIAR=CPIA/PDA
CPIHR=CPIH/PDH
WPIIR=WPII/PDI

Q_i = quantity consumed of the ith end-use category of total fibers (millions of pounds),
CPI_i = consumer price index for the ith end-use category (1967=100), WPI used for industrial category,
PD_i = implicit deflator for all except the ith good in question,
INCN = nominal disposable income (billion dollars),
POP = U S population (millions)

The population impact can be removed by deflating the quantity and income variables by population, and the real income and relative price impacts can be arrived at by deflation of own price and per capita income by the appropriate deflators [4] We consider first the impact of converting to per capita measures The simple correlation coefficients appear in table 1

We first compare simple correlation coefficients between raw and deflated series, and then compare simple correlation coefficients of deflated series with partial correlation coefficients of raw series to determine the extent and magnitude of the bias among these measures [5]

[4] Approximate because the implicit deflator does not contain all prices

[5] The relationship between partial correlation coefficients and regression coefficients is discussed in (*10*, pp 61-65, 132-135, and *18*, pp 131-138) See also (*5*, pp 192-197)

and coefficients of variability *

POP	QAPC	QIPC	QHPC	INCNPC	INCRA	INCRH	INCRI	CPIAR	CPIHR	WPIIR
0 95	0 99	−0 41	0 97	0 86	0 96	0 96	0 94	−0 76	−0 94	−0 92
94	94	− 48	1 00	93	99	99	98	− 75	− 94	− 88
48	66	44	52	28	46	46	41	− 23	− 45	− 47
88	76	− 64	91	99	94	94	96	− 80	− 88	− 79
80	64	− 67	81	97	87	87	90	− 83	− 80	− 71
74	59	− 62	77	95	82	82	86	− 80	− 75	− 61
88	74	− 68	87	99	92	93	95	− 88	− 89	− 80
88	74	− 67	88	1 00	93	93	95	− 88	− 89	− 80
90	76	− 66	89	1 00	94	94	95	− 88	− 91	− 83
91	80	− 64	92	1 00	96	96	97	− 87	− 91	− 83
1 00	90	− 58	93	91	96	96	95	− 87	− 99	− 97
90	1 00	− 30	95	79	92	92	89	− 68	− 89	− 87
− 58	− 30	1 00	− 45	− 64	− 54	− 54	− 57	67	59	56
93	95	− 45	1 00	92	98	98	97	− 72	− 93	− 87
91	79	− 64	92	1 00	96	96	97	− 86	− 91	− 82
96	92	− 54	98	96	1 00	1 00	1 00	− 80	− 95	− 90
96	92	− 54	98	96	1 00	1 00	1 00	− 81	− 95	− 90
95	.89	− 57	97	97	1 00	1 00	1 00	− 81	− 94	− 87
−.87	− 68	67	− 72	− 86	− 80	− 81	− 81	1 00	88	82
− 99	− 89	59	− 93	− 91	− 95	− 95	− 94	88	1 00	96
− 97	− 87	56	− 87	− 82	− 90	− 90	− 87	82	96	1 00
08	14	08	40	40						

Our example demonstrates that both the magnitude and signs of regression coefficients may change because of deflating

Table 2—Regressions for retail demand for textile fibers, 1955 76

Equation	Dependent variable	Independent variable				Equation statistics			
		Ln c	Ln (CPI/PD)	Ln ((INCN/POP)/PD)		R^2	DW	SEE	
A1	$Ln(Q_A/POP)$	−0 37 [1] (−0 22)	0 44 (1 35)	0 95 (7 88)		0 87	1 05	0 05	
H1	$Ln(Q_H/POP)$	−5 25 (−1.44)	0 92 (1 39)	2 95 (8 18)		0 97	1 15	0 07	
I1	$Ln(Q_I/POP)$	1 57 (0 85)	0 23 (0 68)	−0 14 (−0 67)		0 33	1 73	0 07	
		Ln c	Ln CPI	Ln (INCN/POP)	Ln PD				
A2	$Ln(QA/POP)$	3 65 (2 60)	−0 57 (−1 89)	1 56 (10 39)	−1 74 (−5 83)	0 95	1 82	0 037	
H2	$Ln(Q_H/POP)$	−3 07 (−1 00)	0 35 (0 60)	3 39 (10 31)	−4 31 (−5 06)	0 98	1 51	0 057	
I2	$Ln(Q_I/POP)$	1 27 (0 80)	0 10 (0 36)	0 53 (1 76)	−1 25 (−2 01)	0 52	2 36	0 06	
A3	$Ln(C_A/POP)$	2 55 (18 17)	−0 34 (−5 19)	1 56 (10 36)	−1 90 (−8 89)				
H3	$Ln(Q_H/POP)$	0 43 (2 07)	−0 30 (−3 38)	3 09 (15 42)	−3 39 (−11 88)				
I3	$Ln(Q_I/POP)$	3 46 (12 80)	−0 27 (−2 48)	0 24 (1 10)	−0 51 (−1 59)				
		C	CPI	INCN/POP	PD				
A4	Q_A/POP	45 58 (9 15)	−0 19 (−2 79)	59 60 (7 53)	14 99 (−8 76)	0 92	1 51	0 76	
H4	Q_H/POP	45 87 (13 63)	−0 27 (−3 35)	21 76 (10 91)	−75 06 (−4 70)	0 99	1 74	0 82	
		LnC	$Ln(Q_A/POP)$	Ln(INCN/POP)	LnPD				ρ
A5	$Ln(CPI_A)$	4 625 (35 30)	−0 10 (−1 74)	0 35 (2 50)	0 25 (1 24)	0 99	1 08	0 01	0 78
		C	$Ln(Q_A/POP)$	$Ln(INCN/POP/PD_A)$					
A6	$Ln(CPI_A/PD_A)$	4 80 (16 87)	0 20 (1 33)	−0 46 (−3 52)		0 67	0 33	0 04	
A7	$Ln(CPI_A/PD_A)$	4 37 (15 35)	−0 01 (−0 25)	0 16 (0 71)		0 93	0 50	0 02	0 96

[1] *t* statistic

which differs from the unrestricted estimator by a linear function of g - RB (For a summary of this restricted estimation procedure, see *17* and 7) One restriction was imposed on each equation so that, after ignoring the intercept, the absolute value of the price and income coefficients would equal the absolute value of the implicit deflator coefficient [8] The resulting price elasticities were -0 34, -0 30, and -0 27 for apparel, household, and industrial use, respectively Corresponding income elasticities were 1 56, 3 09, and 0 24 The real income and relative price impacts appear to be more reasonable and the t-statistics are larger Although there is no guarantee, it appears likely this restriction will assure the "correct" signs, because of the dominance of the income and price deflator coefficients relative to the price coefficient Two categories of textile demand, household and industrial use, exhibited positive signs on price when equation set 2 was used Note, however, that the restrictions reversed the signs It might be necessary to use inequality restricted least squares to achieve the desired results

If a homogeneous degree-zero demand function is desired, with the functional form exhibited by equation set 2, then R = (0 1 1 1), and equation set 1 results,

$$ln(\text{QAPC}) = -0\ 37 + 0\ 44\ \ ln(\text{CPIA})$$

$$+ 0\ 995\ ln(\text{INCNPC})$$

$$- 1\ 39\ ln(\text{PDA}) \qquad (7)$$

[8] This restriction does not imply homogeneity

which is identical to equation A1 except for the t-statistics which are adjusted for degrees of freedom

CONCLUSIONS

We developed our argument for partial correlation coefficients, but it can be used for regression coefficients Deflating, for whatever reason, may have substantial impacts whether one deflates to maintain fidelity with the hypothesis formulated, as a preference for a particular functional form, to remove heteroskedasticity, or to improve what otherwise might be a severe multicollinearity problem

Our example demonstrates that both the magnitude and signs of regression coefficients may change because of deflating We simply call attention to these consequences as a reminder to those working with numbers The crude restrictions used to obtain "reasonable" estimates of the parameters suggest it may be appropriate to use some form of restricted estimation in conjunction with deflated series, if nothing more than as a check on the results

REFERENCES

(*1*) Bell, Thomas M , K Kobe, and S Evans "The Supply and Demand for Textile Fibers with Emphasis on Cotton " Forthcoming in Proceedings, Beltwide Cotton Conference, National Cotton Council, Memphis, Tenn , 1979

(*2*) Brennan, M J *Preface to Econometrics* Southwestern Publ Co , Cincinnati, 1965

(*3*) Brown, J , and others "A Study of Index Correlations " *J Royal Stat Soc* , 77 317-46, 1914

(*4*) David, F N, and J Neyman "Extensions of the Markoff Theorem of Least Squares " *Stat Res Memoirs*, 2 105-16, 1938

(*5*) Ezekiel, M , and K A Fox *Methods of Correlation and Regression Analysis* John Wiley & Sons, Inc , New York, 1966

(*6*) Foote, R F "Analytical Tools for Studying Demand and Price Structures " U S Dept Agr , Agr Hdbk 146, Washington, D C , 1958

(*7*) Goldberger, A S *Econometric Theory* John Wiley & Sons, Inc , New York, 1964

(*8*) Granger, C W J , and P Newbold "Spurious Regressions in Econometrics " *J Econometrics*, 2 111-120, 1974

(*9*) Houthakker, H S "Additive Preferences " *Econometrica*, 28 244 257, 1960

(*10*) Johnston, J *Econometric Method* McGraw Hill, 2nd ed , New York, 1972

(*11*) Katona, G , and others *Contributions of Survey Methods to*

Econometrics Columbia University Press, New York, 1954

(*12*) Kuh, E , and J Meyer "Correlation and Regression Estimates When the Data Are Ratios " *Econometrica*, 23 400-416, 1955

(*13*) Pearson, K "On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs " *Proc Royal Soc London*, 60, 1897

(*14*) Rao, P , and R L Miller *Applied Econometrics* Wadsworth Publishing Co , Inc , Belmont, Calif , 1971

(*15*) Samuelson, P "Using Full Quality to Show That Simultaneous Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand " *Econometrica*, 33 781 796, 1965

(*16*) Sato, K "Additive Utility Functions with Double-Log Consumer Demand Functions " *J Pol Econ* 80 102 124, Jan - Feb 1972

(*17*) Theil, H "On the Use of Incomplete Prior Information in Regression Analysis " *J Am Stat Assoc* 38 404-414, 1963

(*18*) Walters, A A *An Introduction to Econometrics* Macmillan, 2nd ed , London, 1970

(*19*) Working, E J "What Do Statistical Demand Curves Show?" *Qtr J Econ* 41 212 235, Feb 1927

Appendix table 1—End use categories and fiber classifications

Apparel	Household	Industrial	Fiber classifications
Blouses	Bedspreads	Abrasive products	Cotton
Coats	Blankets	Artist canvas	Wool
Diapers	Curtains	Automotive upholstery	Cellulosic
Dresses	Draperies	Awnings	Staple (rayon)
Jackets	Mattresses and pads	Bags	Yarn (acetate)
Jeans	Pillowcases	Bookbindings	Noncellulosic
Pajamas	Pillow ticking	Electrical insulation	Staple
Rainwear	Quilts	Flags and banners	Polyester
Robes	Sheets	Industrial hose	Nylon
Shirts	Tablecloths and napkins	Life jackets	Olefin
Sport clothes	Thread	Luggage and handbags	Acrylic
Suits	Towels and washcloths	Machinery belts	Yarn
Sweaters	Upholstery	Rope, cordage, and twine	Polyester
Work clothes		Sleeping bags	Nylon
		Tents	Olefin
		Umbrellas	Glass
		Wall covering fabric	

Appendix table 2—Data for demand analysis

Year	Q_A	Q_H	Q_I	CPI_A	CPI_H	WPI_I	POP	INCN	PD_H	PD_H	PD_I
	-- *Million pounds* ---			*(1967=100)*			*Million*	*Billion dollars*	*(1972=1)*		*(1972=1 0)*
1955	2 483 26	1 707 65	2,326 68	88 9	91 9	98 7	165 93	273 41	0 64	0 63	0 63
1956	2,471 38	1,583 73	2,324 50	89 8	93 5	98 7	168 90	291 25	65	64	64
1957	2,364 43	1,477 77	2,189 51	90 6	94 4	98 8	171 98	306 92	67	67	66
1958	2,316 00	1,491 36	2,046 97	90 4	92 9	97 0	174 88	317 13	69	68	68
1959	2,708 23	1,757 25	2,425 59	90 5	93 2	98 4	177 83	336 12	70	70	69
1960	2,675 22	1,710.82	2,212 84	91 5	94 5	99 5	180 67	349 37	72	71	70
1961	2,675 69	1 728 87	2,184 19	92 0	95 0	97 7	183 69	362 90	72	72	71
1962	2,918 79	1 955 51	2,361 10	92 1	94 9	98 6	186 54	383 88	74	73	72
1963	2,942 74	2,048 74	2,458 49	93 0	95 0	98 5	189 24	402 76	75	74	73
1964	3,091 70	2,310 81	2,565 80	93 8	95 3	99 2	191 89	437 03	76	75	74
1965	3,583 38	2,905 44	2,315 55	94 5	96 0	99 8	194 30	472 16	77	76	76
1966	3,752 54	3,155 55	2,577 50	96 2	97 3	100 1	196 56	510 40	79	79	78
1967	3,686 21	3,300 67	2,416 73	100 0	100 0	100 0	198 71	544 55	81	81	80
1968	3,876 47	3,742 45	2,701 16	105 7	103 7	103 7	200 71	588 14	84	84	84
1969	3,789 34	3,892 59	2,643 24	111 9	106 9	106 0	202 68	630 43	88	88	88
1970	3,786 05	3,928 15	2 436 06	116 3	109 2	107 2	204 88	685 94	92	92	92
1971	4,144 49	4,769 00	2,441 27	119 9	111 6	108 6	207 05	742 81	96	96	96
1972	4,427 74	5 180 08	2 516 04	122 3	113 6	113 6	208 85	801 30	1 00	1 00	1 00
1973	4 478 98	5,760 54	2,705 51	126 5	116 2	123 8	210 41	901 70	1 06	1 06	1 06
1974	4,015 12	4,828 67	2 334 63	135 7	131 5	139 1	211 90	984 60	1 18	1 18	1 18
1975	3 940 60	4 730 89	2,154 34	140 6	141 4	137 9	213 56	1,084 40	1 27	1 27	1 28
1976	4,317 90	5,285 50	2 479 47	144 9	148 3	148 0	215 14	1,185 80	1 34	1 34	1 35

Q_A, Q_H, Q_I = Quantity demanded by category (million pounds) End use percentages calculated from National Cotton Council of America data and applied to total domestic consumption figures from the Economics, Statistics, and Coopera tives Service

CPI_A, CPI_H, WPI_I = Consumer Price Index of Apparel minus foot wear, Consumer Price Index of Textile House-furnishings and Producer Price Index of Textile Products and Apparel respectively Bureau of Labor Statistics U S Department of Labor

POP = Total U S population (million) U S Bureau of the Census

INCN = Nominal personal disposable income (billion dollars), U S Department of Commerce

PD_A = Price deflator of services durables, food, gaso line and oil and other nondurables, U S Depart ment of Commerce

PD_H = Price deflator of services nondurables, auto and parts of other durable goods, U S Department of Commerce

PD_I = Price deflator of nondurables and services U S Department of Commerce