Distr.
GENERAL

CES/AC.49/2003/6
23 September 2003

Original: ENGLISH

STATISTICAL COMMISSION and UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)

INTERNATIONAL LABOUR ORGANISATION (ILO)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/ILO Meeting on Consumer Price Indices
(Geneva, 4-5 December 2003)

APPAREL HEDONIC REGRESSION MODELS: IMPROVEMENTS AFTER MORE THAN A DECADE OF USE IN THE U.S. CONSUMER PRICE INDEX

Invited paper submitted by the U.S. Bureau of Labor Statistics*

Abstract

Hedonic regression models have been used in the U.S. Consumer Price Index (CPI) for Apparel since January 1991. This paper describes some of the ways in which their use has evolved over that period. Hedonic regression models are an integral component of apparel commodities index calculations. Numerous enhancements have led to better models and increased use in the CPI. The difficulty associated with formulating and maintaining hedonic regressions for apparel items is discussed. Data preparation, model specification, and model stability are also addressed.

---

* Prepared by Ms. Nicole Rope, Bureau of Labor Statistics, United States of America. The author would like to thank John Greenlees, Charles Fortuna and Paul Liegey for helpful comments and also thank all her colleagues for their excellent work in developing apparel hedonic regression models. The views expressed in this paper are solely those of the author and do not necessarily reflect the policies of the Bureau of Labor Statistics.

GE.03-32224

I. HISTORY AND BACKGROUND

1. Hedonic regression models have been used in the U.S. Consumer Price Index (CPI) for Apparel since January 1991. This paper describes some of the ways in which their use has evolved over that period.

2. Because the CPI is a measure of the change in prices paid by consumers for a fixed market basket of goods and services through time, it is important to make adjustments for any changes in the quality of priced goods or services that take place from one month to the next. Item replacement and quality change are particularly common for apparel commodities due to apparel marketing tactics. New fashions are constantly offered to consumers, although often these new styles are simply a re-bundling of existing characteristics into the latest fashionable item. It is primarily for that reason that the U.S. Bureau of Labor Statistics (BLS) first began extensive use of hedonic models for quality adjustment in the apparel components of the CPI.

3. These econometric models estimate implicit prices for individual characteristics bundled together to form apparel commodities. This allows the CPI apparel analyst to calculate the value of quality change between two items. The parameter estimates from the hedonic regression model are used to adjust the price of the old item for use in index calculations when the new replacement item and old item differ in quality.

4. It should be noted that BLS does not use time dummies or other direct methods to quality adjust price indexes. Although models with time dummies are common in hedonic analysis, the approach used in the CPI is what a recent book (National Research Council 2002) has called the Indirect Method. An example of how a quality adjustment is calculated and applied is in Appendix 1.

5. The Apparel major group encompasses 16 item strata, or basic item index categories. Within these strata there are 36 entry level items (ELIs), which are more detailed subdivisions used in sampling. Heterogeneous ELI's are further divided into clusters. Apparel models are generally specified at the ELI/cluster level. Since 1991, 32 apparel ELI/clusters have been modeled. Not all regression models were used for quality adjustments and some were used only for short time periods since parameter estimates were believed to be unstable. Infrequent substitutions and small sample sizes led to other models' brief usage. A list of all apparel ELIs and ELIs with models is in Appendix 2.

6. The ability to use quality adjustments to adjust prices to account for quality change has led to an increased use of directly compared prices in apparel index calculations. Prior to the use of quality adjustments, apparel analysts compared prices for replacement items only four or five out of every ten times; in 2002 the rate improved to eight or nine out of every ten times. Quality adjustments were used in the apparel index almost 20,000 times during 1991 to 2002.

7. There is significant difficulty associated with formulating and maintaining hedonic regressions for apparel items. Preparing the data for analysis is a time consuming process and much thought must be placed on model specification. One of the biggest obstacles is accounting for how much value a consumer places on fashion. The degree to which an apparel

item is perceived to be fashionable is subjective and difficult to measure, but since it is invariably a price determinant it cannot be ignored in model specification. Multicollinearity is also a frequent problem since many of the variables are closely related to one another. There is also uncertainty as to the stability of the parameter estimates of the resulting model due to the dynamics of the apparel market.

II. DATA

8. CPI field economists in apparel stores (outlets in index terminology) throughout the United States collect the data that are used to calculate the apparel hedonic regression models. Statisticians select the outlets via sampling procedures and allocate the number of observations for collection in a given outlet. The apparel sample is designed to allocate pairs of observations to each outlet to account for the seasonal nature of apparel retailing. The field economist must select one item available during the spring/summer selling season months and one item available during the fall/winter selling season months. A unique item is selected in the outlet for inclusion in the sample with probability proportional to sales. The field economists then describe that unique item on data collection forms designed by apparel economists. Over the past decade, these data collection forms have been fine-tuned in order to enhance the quality of data obtained by the field economists. Apparel economists use their market knowledge and the hedonic regression model results themselves when designing the data collection form. Variables that have proven to be statistically significant are included on the data collection form.

9. Apparel economists review all of the descriptions of the unique items in their sample. They verify that all relevant information has been obtained and ensure that all descriptions are consistent. For instance, if a vest is described as having long sleeves or if a solid color shirt is also reported as having a plaid design the apparel economist will ask the field economist to clarify and correct the description. This is likely the most time consuming part of the hedonic modeling process. However, the emphasis on having a 'clean' data set pays off on an improved model (most econometricians are familiar with the "garbage in, garbage out" effect). Over the years the apparel economists have worked on improving the data verification process. They have improved the design of the data collection documents in order to combat common errors, have developed programs to automatically find inconsistencies, use spreadsheet tools to become more familiar with the data, and maintain a dialog with the field economists to ensure that problems are addressed. These improvements have led to a considerable reduction in the amount of time spent on data preparation compared to the efforts of a decade ago, and the quality of data has improved substantially.

III. MODEL SPECIFICATION

10. Once a cross section of data is cleaned, variables are created from the apparel item's characteristics. Dummy variables are constructed for all characteristic data with the exception of the fiber content variables. The functional form for the regression model is specified as a semi-logarithmic relationship between price and characteristics, implying a rising supply price per characteristic unit (Griliches, 1971) :

$$\log(\text{regular price}) = \beta_0 + \Sigma\beta_i X_i + \varepsilon_i.$$

11. This functional form is recommended in the hedonic model literature, for example, Griliches (1971) and Triplett (1971). The natural logarithm of regular (i.e., non-sale) price is the dependent variable, and the coefficient β_i is interpreted as a measurement of the percentage change in price associated with a unit change in the quality specification X_i, assuming all the other values for the remaining explanatory variables are held constant. The β_0 value is assumed to be the value of the base item without any of the quality characteristics that enhance or add value to the commodity (Georges and Liegey, 1988). Sale prices are replaced with regular prices in order to give equal importance to fall/winter and spring/summer items regardless of the season from which the cross-section of data has been derived.

12. Determining the best set of regressors for an hedonic regression model is a difficult task. In the early applications of CPI apparel models, model specifications were based on results of stepwise regressions and on choosing characteristic variables that were highly correlated with the dependent variable. Only variables deemed significant according to the t-statistic were included in the model. Unfortunately this approach can easily result in models that are biased or misspecified. Liegey (1993) notes that there are difficulties determining the best set of characteristics to explain price for each stratum because of the influence that fashion has on price.

13. Apparel economists are now more knowledgeable about the intricacies of the items they are modeling than they were a decade ago and are now better able to determine an initial model specification via a priori knowledge. Emphasis is placed on using their knowledge to construct the regression model instead of relying solely on results of statistical tests. The data are further analyzed when results run contrary to expectations. Due to the nature of the data the preliminary regression results have the potential to be misleading. This is especially true if regressions are run that include all potential variables. High correlations between independent variables will cause standard errors to increase and potentially even lead to insignificant parameter estimates for key variables. Based on the results of this type of model, someone unfamiliar with the item being modeled could unwittingly conclude that an important variable does not belong. Also, including extraneous variables in the model specification will lead to high variances (Kennedy, 1998).

14. Apparel economists have also become adept at specifying models that mitigate the effect of multicollinearity. Early models eliminated highly correlated variables in order to obtain a model with more precise parameter estimates. Since the parameter estimates are used to adjust prices in the CPI for differences in quality the parameter estimates must be as precise as possible; however, eliminating variables leads to omitted variable bias and biased estimates. For example, style and type of closure tend to be highly correlated. A 1992 model for women's shirts, blouses and other tops has two variables that serve to explain style and type of closure (pullover and open-front shirt) and the 2002 version of the model has five (vest, open-front blouse, open-front shirt, pullover blouse, and other pullover). Multicollinearity is avoided by combining style and type of closure into a single variable and model specification is thereby optimized.

15. Some of the other improvements involve better brand and outlet category definitions. Inclusion of brand category variables has been debated since the first apparel models. Armknecht and Weyback (1989) considered including brand category variables in their preliminary apparel regression models but decided not to use them due to "their significant instability." In 1992, 60 percent of the models included some type of brand category variable. By 2002 all models have at least one brand category or brand name variable (the athletic footwear model is the only model that includes variables for actual brand names). Improved definitions and more accurate coding of brands into their respective categories have resulted in more useful and realistic parameter estimates. Over the years it even appears that the parameter estimates themselves have become more stable. Simply comparing models for several items showed that the parameter estimates for the brand category variables are nearly the same for the same item over different time periods. Apparel economists issued new type of business designations for apparel outlets in 1996. These improved designations, combined with more thorough data cleaning, led to more frequent inclusion of type of business variables in apparel models and resulting parameter estimates that make more intuitive sense. The parameter estimates for type of business are not actually used for quality adjustments; however, they do help the overall model specification in that they control for the effects that the different business practices have on price.

16. One other notable difference between early and recent models is the more frequent inclusion of country (or region) of origin variables in model specification. This information was added to most apparel data collection forms during the latter half of the 1990s. These variables mainly serve as proxies for quality of the materials and workmanship. Consumers are aware that some countries produce apparel of superior quality and other countries sacrifice quality in order to provide lower priced items. Italian made shoes and clothing have long been recognized as having unsurpassed quality and the models that include Italy (or in some cases the Western Europe region) as a variable have supported this belief by their positive, significant parameter estimates. For the most part, country of origin variables are used as control variables and not used for quality adjustments. However, there are instances where country or region does serve to explain quality differences between items, so that country of origin parameter estimates are sometimes applicable for quality adjustments. For example, higher quality brands and better quality fibers are often labeled as originating from North America and certain Asian or European countries. Parameter estimates from these countries have proven to be more stable than others.

IV. MODEL EVALUATION

17. Prior to using the parameter estimates from a regression model to quality adjust substitutions, the overall quality of the model is evaluated. The primary measure of quality is whether the resulting parameter estimates make sense. The modeler verifies that the signs associated with the parameter estimates are in the expected direction and that the ranks of the parameter estimates match a priori expectations. Any parameter estimates that conflict with a priori expectations are further investigated until they are explained. Influential observations are also investigated to ensure their accuracy and also to verify that they are representative of the item being modeled. The model is tested for multicollinearity by examining correlations between explanatory variables and through the tolerance statistic. Parameter estimates from

highly correlated variables are not used for quality adjustments since multicollinearity causes parameter estimates to be imprecise. Only variables with parameter estimates significant at the 5 percent level are actually used for quality adjustments.

V. STABILITY

18. Stability of the hedonic regression model continues to be a concern, as demonstrated by the women's suits model. This model was tested for stability using the Chow test. The previous version of the regression model (from 1998) was run on the most recent data set (2001) and also on the data set pooled from both time periods. The calculated F value is significant, thus the models from the two time periods are different. More models were run to determine if any one group of variables caused the instability. Since Armknecht and Weyback (1989) believed that the brand category variables caused instability, the Chow test was applied to the 1998 model without any brand variables. This still led to the same conclusion. In fact, removing other variable categories believed to cause instability (country of origin, type of business and fiber) did not change the outcome. The only model that remained stable according to the Chow test results is a model that included only fiber variables. However, the parameter estimates in the fiber only model varied widely between the models.

19. Another method of gauging stability of the parameter estimates is to simply compare parameter estimates from the same model from different time periods. Often the parameter estimates are nearly identical and (provided the base variable does not change) the sign of the parameter estimate is almost always the same. However, there are enough instances where the parameter estimates do change drastically to indicate that the models are not stable over time. Fiber parameter estimates seem to fluctuate the most. Comparing models over time shows that not only do the parameter estimates change over time, but the ranks also change. For example, in women's sweaters and sweater vests, three fibers (polyester, ramie, and acrylic) switch ranks in all three time periods. (See Table 1.) The fiber variables also have more 'turnover' in the model specification since fibers are included and excluded from model specification based on their prevalence in the current sample. Parameter estimates for fiber are thought to be the most commonly used parameter estimates for quality adjustments. Due to their frequent use their reliability is quite important. Thought should be given as to whether a different approach needs to be taken to account for fiber quality in apparel hedonic regression models.

20. Stability issues are likely to be a long term problem for apparel regression models. Since it is unknown how long fiber parameter estimates remain stable, previous recommendations to update models every 12 – 18 months should be reexamined. At a minimum, models should be updated as soon as possible if new features exist or if fiber mix changes.

VI. WOMEN'S SUITS TEST MODELS

21. In order to evaluate whether data cleaning and model specification efforts are worthwhile additional regression models were run and compared to the final model for women's suits and suit components. First a model was run on the minimally cleaned data set

for women's suits and suit components. (The final, official model used the same but cleaned data set.) (See Appendix 3.) The statistical software is unable to process the data sets used for modeling without some preparation. Observations were cleaned if incorrect data formats would have prevented them from being processed by the software (for instance, text entries were deleted from the fiber amounts since these variables must be numeric and text entries that could not be read by the statistical software were overwritten). Regular prices must be entered for sale priced items. Also, items that were not representative or ineligible for pricing as a suit and items with too much missing information were deleted from the data set. The resulting model is still quite good according to the adjusted R^2 (.7489), but seven of the parameter estimates that are significant in the "final" model are no longer significant. Most of the parameter estimates still made sense; in other words, the signs remained the same and usually the parameter estimates' rank within a category did not change. There are changes with the parameter estimates for fiber category variables that did not follow a priori reasoning. For instance, triacetate switched ranks with wool in the minimally cleaned model suggesting that triacetate is a bigger price factor than wool. The parameter estimates are less precise due to an increase in the standard errors for almost all of the parameter estimates. Also, since minimal effort went into cleaning the data set, new price determining variables were not identified. This led to three fewer variables in the model – variables that were found to be significant in the final model. By not including these important variables the minimally cleaned model is misspecified and therefore the parameter estimates are biased.

22. A second test model was run on the thoroughly cleaned women's suits data set. This model included virtually all available variables in the model specification. (See Appendix 3.) The resulting model has virtually the same adjusted R^2 as the final model (0.8615). Only two of the parameter estimates that are significant in the final model are no longer significant and one parameter estimate became significant in the all variable model. As expected, the standard errors for most parameter estimates notably increased. Contrary to the minimally cleaned data set model, none of the parameter estimates switched ranks within their categories between the two models. Contrary to a priori expectations, a few variables that were not included in the final model are significant in the all variable model. The parameter estimate for petite sizes is positive and significant but this is due to higher quality characteristics generally found in petite sized suits; these suits are more likely to be made from better quality fibers, sold in full-price outlets, and have brands that are more valued by consumers. However, misses sized suits with the same characteristics as petite sized suits are comparably priced; in other words, petite sized women are not paying a premium for their suits. Unexpected significant parameter estimates should not be used for quality adjustment purposes unless a valid reason explains their importance. Even if models are specified with every potential variable, analysis still needs to occur to ensure that resulting parameter estimates and significance make sense.

VII. MODEL USAGE AND IMPROVEMENTS

23. Improved model specification and higher quality data have led to vastly improved models. In 1992, on average the adjusted R^2 for models used in the apparel index was 0.6190. In 2002, the average adjusted R^2 rose to 0.7924. The number of variables rose notably from 1992 to 2002 – 1992 models had on average 12 variables in the regression model and 2002 models average 40 variables. Also, current models rely less on hard-to-define style name

variables and more on actual tangible attributes for model specification. Table 2 delineates implemented improvements and Graph 1 illustrates the increase in the average adjusted R^2. The increased ability of the explanatory variables to explain the variation in the dependent variable has led to more frequent use of quality adjustments.

24. Lower level data show that women's apparel models are responsible for most of the increase in the proportion of quality adjusted substitutions. (See Table 3.) Women's apparel items are more complex than men's apparel items and are also influenced much more by fashion. The regression models for women's apparel reflect the increased complexity — women's apparel regression models have on average 39 variables included in model specification compared to an average of 25 variables for men's apparel regression models. Women's suits and women's dresses, arguably the most complex apparel items, are the most frequently applied regression models — in 2002, 60 percent and 61 percent of substitutions for women's suits and women's dresses were quality adjusted.

VIII. CONCLUSION

25. By several measurable standards, apparel regression models have greatly improved since their inception in the U.S. CPI. Apparel economists now rely more on their commodity knowledge to develop regression models. This has led to more appropriate model specification and has resulted in more relevant and usable parameter estimates. The effort expended on ensuring data quality has led to parameter estimates that are more precise than models from a decade ago.

26. Further major improvements to model fit are unlikely given that consumer's demand for certain apparel fashions is difficult to quantify. It is unlikely that there are any new major variable categories that would enhance model fit. Effort now needs to be directed at maintaining existing models, updating models on a timely basis, and adding new models where needed. Also, as part of the BLS research program on hedonics, the use of predictive models in index calculations is being evaluated, what the National Research Council (2002) termed the Direct Characteristics Method. This entails calculating regression models each period and using the resulting parameter estimates to predict the price of every item based on its characteristics.

Table 1

Women's Sweaters and Sweater Vests

	Parameter Estimate / Time Period			Rank (from base) / Time Period		
Fiber	T	T-1	T-2	T	T-1	T-2
Spandex	-0.00609			-5		
Polyester	-0.00463	-0.00204	-0.00224	-4	-1	-2
Ramie	-0.00281	-0.00318	-0.00551	-3	-3	-3
Acrylic	-0.00243	-0.00245	-0.00211	-2	-2	-1
Nylon	-0.00184			-1		
Cotton	Base	Base	Base			
Silk	0.00193	0.00169	0.00176	+1	+1	+1
Wool	0.00217	0.00260	0.00349	+2	+3	+3
Rayon	0.00297	0.00197	0.00263	+3	+2	+2
Cashmere	0.00926	0.01139	0.01138	+4	+4	+4
Linen	0.01144			+5		
Metallic	0.02063			+6		
Mohair			0.01234			+5

Table 2

Year:	Improvement:
1991	Apparel index implemented hedonic models for quality adjustment purposes
1995	Improved model specification procedures
1996	Updated type of business definitions
1997	Updated brand category definitions
1998	Widespread use of country of origin variables

Graph 1

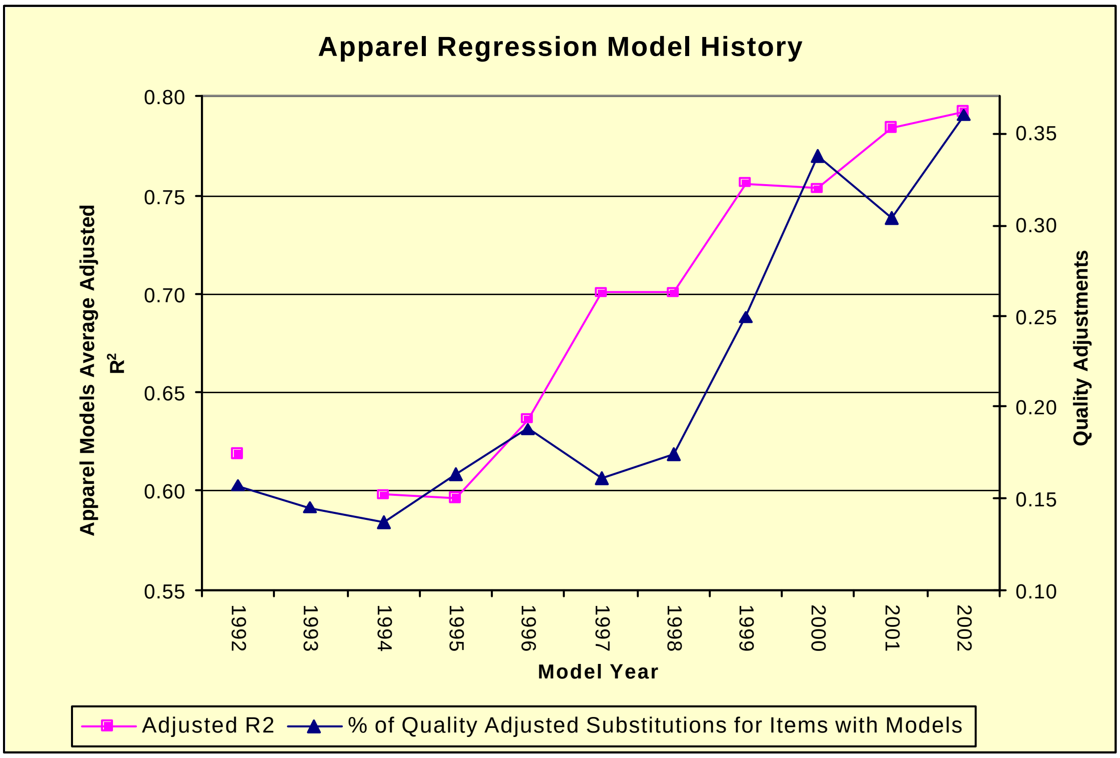

Table 3

ELI		Adjusted R^2	Proportion of Quality Adjusted Substitutions in 2002	Number of Regression Model Variables
AA011	Men's suits and formal wear	0.6778	28.1%	26
AA012	Men's sport coats and tailored jackets	0.7416	27.1%	16
AA013	Men's outerwear	0.7542	20.1%	30
AA031	Men's shirts	0.6526	16.4%	27
AA041	Men's pants and shorts	0.7500	26.0%	26
AC011	Women's outerwear	0.7254	18.8%	37
AC021	Women's dresses	0.8127	61.2%	49
AC031	Women's shirts, blouses etc	0.7073	42.0%	38
AC031	Women's sweaters	0.8049	42.6%	41
AC031	Women's tailored and untailored jackets	0.8154	43.0%	38
AC032	Women's pants and shorts	0.7459	40.5%	35
AC032	Women's skirts	0.8216	47.0%	34
AC033	Women's suits	0.8619	60.2%	42
AE011	Men's athletic footwear	0.8382	9.1%	28
AE011	Men's dress and casual shoes	0.8104	14.5%	24
AE031	Women's athletic footwear	0.8382	7.0%	28
AE031	Women's dress and casual shoes	0.8828	31.5%	42

References

Armknecht, P.A., and D. E. Weyback. 1989. "Adjustments for quality change in the U.S. consumer price index." Journal of Official Statistics (2). pp 107-123.

Fixler, Dennis, Charles Fortuna, John Greenlees and Walter Lane. 2000. "The Use of Hedonic Regressions to Handle Quality Change: The Experience of the U.S. CPI," in Proceedings of the Ottawa Group Fifth Meeting. Reykjavik: Statistics Iceland.

Georges, E.V., and P.R. Liegey, Jr. 1988. "An examination using hedonic regression techniques to measure the effects of quality adjustment on apparel indexes. Internal report. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Griliches, Zvi. 1971. "Introduction: Hedonic Price Indexes Revisited". In Price Indexes and Quality Change, ed. Z. Griliches. Cambridge, Mass.: Harvard University Press.

Kennedy, Peter. 1993. A Guide to Econometrics. MIT Press, Cambridge Mass.

Liegey, P.R. Jr. 1993. "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences". In Price Measurements and Their Uses. The University of Chicago Press, 1993. pp 209-26.

Liegey, P.R. Jr. 1994. "Apparel price indexes: effects of hedonic adjustment". Monthly Labor Review. pp 38-45.

National Research Council. 2002. At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes. Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes, Charles L. Schultze and Christopher Mackie, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Shepler, Nicole. 1994. "Analysis of Hedonic Regression: Applied to Women's Apparel in the Consumer Price Index". Manuscript. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Thompson, Bill. 1993. "Stability of Regression Models". Manuscript. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Triplett, J. E. 1971. "The Theory of Hedonic Quality Measurement and Its Use in Price Indexes." BLS Staff Paper 6.

Appendix 1

Quality Adjustment Example
(women's sweaters)

	New Item (time period=t)	Old Item (time period t-1)
Price:	$120.00	$100.00
Characteristics:	Cardigan (parameter estimate=0.10965)	Pullover
	Long sleeves	Long sleeves
	Machine knit	Machine knit
	100% cotton	100% cotton
	National brand	National brand
	Machine wash	Machine wash
	Multicolor	Multicolor
	Misses size range	Misses size range
	Single rib knit	Single rib knit
	No adornment	No adornment
	USA origin	USA origin

Quality Adjustment Calculation

Adjusted price of old item	= (price of old item) * ($e^{?\text{parameter estimate changes}}$)
	= ($110) * ($e^{(0.10965)}$)
	= $111.589
Price change used in index calculations	= (price of new item)/(adjusted price of old item)
	= $120/$111.589
	= 7.5 percent

Women's Sweaters (AC031– 01) Final Model

Variable Category	Variable Name	Parameter Estimate	Standard Error	T Statistic	Tolerance Statistic
	Intercept	3.84863	0.04864	79.12	.
Fiber:	Metallic	0.02063	0.01724	1.20	0.90003
	Linen	0.01144	0.00476	2.40	0.84222
	Cashmere	0.00926	0.00146	6.32	0.59281
	Rayon	0.00297	0.00136	2.18	0.83318
	Wool	0.00217	0.00083875	2.58	0.52381
	Silk	0.00193	0.00102	1.90	0.57488
	Cotton	Base			
	Nylon	-0.00184	0.00177	-1.04	0.78141
	Acrylic	-0.00243	0.00041498	-5.86	0.65191
	Ramie	-0.00281	0.00107	-2.62	0.52444
	Polyester	-0.00463	0.00170	-2.72	0.90726
	Spandex	-0.00609	0.00877	-0.69	0.80737
Sweater Style:	All cardigan sweaters	0.10965	0.03220	3.41	0.82207
	All pullover sweaters	Base			
Brand/Label Category:	Exclusive	0.71067	0.34760	2.04	0.73036
	National/Regional	Base			
	Miscellaneous	-0.30491	0.07328	-4.16	0.55191
	Private label	-0.32956	0.03972	-8.30	0.53360
Knitting Method:	Hand knit	0.45563	0.14020	3.25	0.90605
	Machine knit	Base			
Sleeve Length:	Long sleeved	Base			
	Short sleeved	-0.17348	0.04156	-4.17	0.81691
	Sleeveless	-0.29788	0.04491	-6.63	0.75729
Fabric Design:	Multicolored	0.23561	0.03459	6.81	0.87687
	Solid color	Base			
Cleaning Method:	Dry clean	0.20823	0.07488	2.78	0.33613
	Hand wash	0.11326	0.04517	2.51	0.65693
	Machine wash	Base			
Size Range:	Women’s plus	0.09158	0.05628	1.63	0.90947
	Petites/Misses/Maternity	Base			
	Juniors	-0.26349	0.05113	-5.15	0.83209
Body Knit:	Crochet/Loose/Open weave	0.29886	0.09842	3.04	0.84728
	Cable knit	0.09450	0.03917	2.41	0.90978
	Single/rib knit	Base			
Details/Features:	Adorned (embroidery, appliqué, sequins, beads, glitter, rhinestones)	0.16150	0.04434	3.64	0.69739
	No features	Base			
Country of Origin:	Western Europe	0.22184	0.10607	2.09	0.80055
	Asia	0.07567	0.03477	2.18	0.67396
	USA	Base			
	Mexico	-0.31882	0.12750	-2.50	0.91502
Type of Outlet:	Independent/Boutique	0.41539	0.10722	3.87	0.71383

	Apparel department	0.39709	0.07680	5.17	0.78145
	Mail order/Catalog	0.24995	0.08182	3.05	0.68845
	Full price women's	0.18870	0.06267	3.01	0.77708
	Full price family	0.11785	0.06204	1.90	0.72830
	Full price department	Base			
	Discount family	-0.26048	0.11542	-2.26	0.84131
	Off price family	-0.47431	0.10712	-4.43	0.87017
	Discount department	-0.68987	0.04500	-15.33	0.63181
	Off price department	-0.74367	0.08080	-9.20	0.61816
Control Variables:	B size PSU	-0.11578	0.03329	-3.48	0.74785
	C size PSU	-0.22690	0.05391	-4.21	0.81343
	West region	-0.04681	0.03570	-1.31	0.83324

R^2=0.8229	Adjusted R^2=0.8049	F value=45.89	Number of observations=447
Model completed: 09/26/2002	Data extracted: 0206/0207	Month first used for QA's: 200209	

Appendix 2

			Model Status
MAJOR GROUP: APPAREL COMMODITIES			
Men's apparel		**AA**	
	Men's suits, sport coats, and outerwear	*AA01*	
	MEN'S SUITS	AA011	In use
	MEN'S SPORT COATS AND TALORED JACKETS	AA012	In use
	MEN'S OUTERWEAR	AA013	In use
	Men's furnishings	*AA02*	
	MEN'S UNDERWEAR, HOSIERY AND NIGHTWEAR	AA021	Never modeled
	MEN'S ACCESSORIES	AA022	Never modeled
	MEN'S ACTIVE SPORTSWEAR	AA023	Never modeled
	Men's shirts and sweaters	*AA03*	
	MEN'S SHIRTS	AA031	In use
	MEN'S SWEATERS AND VESTS	AA032	Model no longer in use
	Men's pants and shorts	*AA04*	
	MEN'S PANTS AND SHORTS	AA041	In use
	Unsampled men's apparel	*AA09*	
	UNSAMPLED ITEMS	AA090	
Boy's apparel		**AB**	
	Boy's apparel	*AB01*	
	BOYS' OUTERWEAR	AB011	Model no longer in use
	BOY'S SHIRTS AND SWEATERS	AB012	Never modeled
	BOYS' UNDERWEAR, NIGHTWEAR, HOSIERY AND ACCESSORIES	AB013	Never modeled
	BOYS' SUITS, SPORT COATS, AND PANTS	AB014	Never modeled
	BOYS' ACTIVE SPORTSWEAR	AB015	Never modeled
	Unsampled boy's apparel	*AB09*	
	UNSAMPLED ITEMS	AB090	
Women's apparel		**AC**	
	Women's outerwear	*AC01*	
	WOMEN'S OUTERWEAR	AC011	In use
	Women's dresses	*AC02*	
	WOMEN'S DRESSES	AC021	In use
	Women's suits and s eparates	*AC03*	
	WOMEN'S TOPS	AC031	
	Women's Sweaters	AC031-01	In use
	Women's Shirts, Blouses, Other Tops	AC031-02	In use
	Women's Tailored and Untailored Jackets	AC031-03	In use
	WOMEN'S SKIRTS, PANTS, AND SHORTS	AC032	
	Women's Skirts	AC032-01	In use
	Women's Pants and Shorts	AC032-02	In use
	WOMEN'S SUITS AND SUIT COMPONENTS	AC033	In use
	Women's underwear, nightwear, sportswear and accessories	*AC04*	
	WOMEN'S UNDERWEAR AND NIGHTWEAR	AC041	Never modeled
	WOMEN'S HOSIERY AND ACCESSORIES	AC042	Never modeled
	WOMEN'S ACTIVE SPORTSWEAR	AC043	
	Women's exercise and sport suits	AC043-02	Model no longer in use
	Unsampled women's apparel	*AC09*	
	UNSAMPLED ITEMS	AC090	

Girls' apparel			**AD**	
	Girls' apparel		*AD01*	
		GIRLS' OUTERWEAR*	AD011	Model no longer in use
		GIRLS' DRESSES	AD012	Model no longer in use
		GIRLS' TOPS	AD013	
		Girls' Sweaters*	AD013-01	Model no longer in use
		Girls' Shirts, Blouses, or Tops*	AD013-02	Model no longer in use
		GIRLS' SKIRTS AND PANTS	AD014	
		Girls' Pants	AD014-01	Model no longer in use
		GIRLS' ACTIVE SPORTSWEAR	AD015	
		Girls' Swimsuits*	AD015-01	Model no longer in use
		Girls' Exercise and Sport Suits*	AD015-02	Model no longer in use
		GIRLS' UNDERWEAR, NIGHTWEAR, HOSIERY AND ACCESSORIES	AD016	
		Girls' Nightwear*	AD016-01	Model no longer in use
		Girls' Underwear*	AD016-02	Model no longer in use
	Unsampled girls' apparel		*AD09*	
		UNSAMPLED ITEMS	AD090	
Footwear			**AE**	
	Men's footwear		*AE01*	
		MEN'S FOOTWEAR	AE011	
		Men's Dress and Casual Shoes	AE011-01	In use
		Men's Athletic Footwear	AE011-03	In use
	Boys' and girls' footwear		*AE02*	
		BOYS' FOOTWEAR	AE021	Never modeled
		GIRLS' FOOTWEAR	AE022	Never modeled
	Women's footwear		*AE03*	
		WOMEN'S FOOTWEAR	AE031	
		Women's Dress and Casual Shoes	AE031-01	In use
		Women's Athletic Footwear	AE031-02	In use
		Women's Slippers	AE031-04	Model no longer in use
Infants' and toddlers' apparel			**AF**	
	Infants' and toddlers' apparel		*AF01*	
		INFANTS' AND TODDLERS' OUTERWEAR, PLAY AND DRESSWEAR, AND SLEEPWEAR	AF011	Never modeled
		INFANTS' AND TODDLERS' UNDERWEAR AND DIAPERS	AF012	Never modeled
Jewelry and watches			**AG**	
	Watches		*AG01*	
		WATCHES	AG011	Model never used
	Jewelry		*AG02*	
		JEWELRY**	AG021	Model no longer in use

*Maximum likelihood regressions using 12 months of panel data
**Jewelry also has a number of subset models for individual jewelry items (e.g., bracelets, rings, pendants, earrings, and necklaces)

Appendix 3

Women's Suits and Suit Components

(data extracted 2001-07 and 2001-08)

Final official model versus minimally cleaned data model

	Official Model (used 2001-10 to 2003-02)				Minimally Cleaned Data Model			
Variable Name	Parameter Estimate	Standard Error	T Statistic	Tolerance Statistic	Parameter Estimate	Standard Error	T Statistic	Tolerance Statistic
Intercept	5.07874	0.04433	114.56		5.24977	0.06543	80.23	.
Spandex	0.00952	0.00833	1.14	0.89259	0.01495	0.01129	1.32	0.88508
Tencel	0.00779	0.00401	1.94	0.90605	0.01123	0.00543	2.07	0.90177
Wool	0.00300	0.00039	7.77	0.64737	0.00351	0.00051	6.83	0.66417
Triacetate	0.00287	0.00146	1.97	0.87978	0.00501	0.00195	2.58	0.89981
Silk	0.00197	0.00067	2.92	0.83900	0.00350	0.00090	3.90	0.86032
Rayon	0.00085	0.00044	1.93	0.78474	0.00080	0.00059	1.35	0.79430
With top	0.22510	0.13954	1.61	0.92361	0.19847	0.18690	1.06	0.93637
With pants	0.10321	0.02859	3.61	0.71423	0.06771	0.03860	1.75*	0.71279
Exclusive	1.41056	0.16623	8.49	0.86614	0.73118	0.26291	2.78	0.94291
Boutique	0.17807	0.08423	2.11	0.54898				
Private	-0.29727	0.03221	-9.23	0.54581	-0.28875	0.04241	-6.81	0.64225
Miscellaneous	-0.52569	0.06078	-8.65	0.68189	-0.42157	0.06989	-6.03	0.75230
Sold separately	0.12541	0.03115	4.03	0.58605	0.13664	0.04191	3.26	0.58907
Machine wash	-0.07346	0.04686	-1.57	0.45805	-0.12174	0.06255	-1.95	0.46743
Juniors	-0.20334	0.07819	-2.60	0.60633	-0.13244	0.11076	-1.20*	0.72586
Jacket short sleeve	-0.10975	0.04190	-2.62	0.76217	-0.08626	0.05317	-1.62*	0.78217
Jacket not lined	-0.36581	0.05247	-6.97	0.28988	-0.51591	0.06821	-7.56	0.31197
Bottom not lined	-0.19332	0.04596	-4.21	0.31536	-0.15195	0.06247	-2.43	0.31039
Bottom no waistband	0.21380	0.04653	4.59	0.59210	0.14379	0.06096	2.36	0.61896
Bottom set on waistband	0.16509	0.03349	4.93	0.49020	0.07466	0.04305	1.73*	0.53622
Multicolor	0.04982	0.03038	1.64	0.88207	0.07288	0.04188	1.74	0.90780
Adornment	0.07586	0.03568	2.13	0.83961	-0.06077	0.04063	-1.50*	0.86789
Belt	0.10539	0.06669	1.58	0.79517	0.04719	0.09821	0.48	0.77365
Western Europe	0.27056	0.12992	2.08	0.85397	0.69429	0.16783	4.14	0.93073
Hong Kong	0.20072	0.09801	2.05	0.94313	0.28150	0.13169	2.14	0.95007
Korea	0.10051	0.05387	1.87	0.84186	0.04487	0.07124	0.63	0.87538
Southeast Asia	-0.07830	0.02945	-2.66	0.71930	-0.12742	0.03851	-3.31	0.75989
Central America	-0.26254	0.05841	-4.49	0.81570	-0.35739	0.07885	-4.53	0.81407
Bangladesh	-0.50470	0.14105	-3.58	0.90388	-0.60702	0.19202	-3.16	0.88709
Mexico	-0.58084	0.07430	-7.82	0.78555	-0.66365	0.10382	-6.39	0.77588
Caribbean	-0.66248	0.16068	-4.12	0.92706	-0.76535	0.21563	-3.55	0.93622
Independent	0.36199	0.07966	4.54	0.51091				
Apparel department	0.34696	0.05080	6.83	0.73003				
Full price women's	0.13897	0.04501	3.09	0.64181	0.14464	0.05134	2.82	0.70783
Catalog	0.08252	0.07593	1.09	0.79749	-0.06628	0.10062	-0.66	0.82607

Full price family	-0.18933	0.06988	-2.71	0.84021	-0.04761	0.05399	-0.88*	0.87339
Off price family	-0.39465	0.10026	-3.94	0.80262	-0.67479	0.10521	-6.41	0.86020
Discount family	-0.40816	0.07292	-5.60	0.73250	-0.33279	0.10628	-3.13	0.84293
Discount department	-0.53486	0.06443	-8.30	0.64772	-0.52075	0.09086	-5.73	0.68582
Off price department	-0.85314	0.05069	-16.83	0.82666	-0.85572	0.08085	-10.58	0.86615
B size city	-0.12077	0.02767	-4.36	0.80128	-0.15594	0.03651	-4.27	0.83680
C size city	-0.17146	0.06775	-2.53	0.80760	-0.12985	0.09075	-1.43*	0.81868
Adjusted R^2	0.8619				0.7489			
F Statistic	81.56				42.45			

Final official model versus all variable model

	Official Model (used 2001-10 to 2003-02)				**All Variable Model**			
	Parameter Estimate	Standard Error	T Statistic	Tolerance Statistic	Parameter Estimate	Standard Error	T Statistic	Tolerance Statistic
Intercept	5.07874	0.04433	114.56		5.12640	0.06816	75.21	.
Spandex	0.00952	0.00833	1.14	0.89259	0.00795	0.00882	0.90	0.79917
Tencel	0.00779	0.00401	1.94	0.90605	0.00897	0.00429	2.09+	0.79677
Wool	0.00300	0.00039	7.77	0.64737	0.00292	0.00040	7.23	0.59293
Triacetate	0.00287	0.00146	1.97	0.87978	0.00263	0.00150	1.75*	0.83194
Silk	0.00197	0.00067	2.92	0.83900	0.00198	0.00070	2.82	0.78055
Nylon					0.00176	0.00236	0.75	0.85793
Rayon	0.00085	0.00044	1.93	0.78474	0.00087	0.00045	1.94	0.75700
Acrylic					-0.00005	0.00285	-0.02	0.93055
Acetate					-0.00022	0.00097	-0.23	0.78587
Linen					-0.00037	0.00129	-0.29	0.77433
Cotton					-0.00071	0.00150	-0.47	0.81525
With top	0.22510	0.13954	1.61	0.92361	0.23600	0.14390	1.64	0.87135
With pants	0.10321	0.02859	3.61	0.71423	0.09472	0.02954	3.21	0.67150
Exclusive	1.41056	0.16623	8.49	0.86614	1.40119	0.17045	8.22	0.82650
Boutique	0.17807	0.08423	2.11	0.54898	0.20250	0.08597	2.36	0.52868
Private	-0.29727	0.03221	-9.23	0.54581	-0.29305	0.03372	-8.69	0.49970
Miscellaneous	-0.52569	0.06078	-8.65	0.68189	-0.53558	0.06163	-8.69	0.66545
Evening style					-0.04124	0.05035	-0.82	0.67118
Sold separately	0.12541	0.03115	4.03	0.58605	0.13985	0.03350	4.18	0.50862
Machine wash	-0.07346	0.04686	-1.57	0.45805	-0.09536	0.05173	-1.84	0.37702
Hand wash					-0.03589	0.11588	-0.31	0.77206
Juniors	-0.20334	0.07819	-2.60	0.60633	-0.18994	0.08096	-2.35	0.56744
Maternity					0.11599	0.15969	0.73	0.70755
Women’s plus					0.02087	0.05839	0.36	0.76531
Petites					0.09112	0.03871	2.35	0.79361
Jacket short sleeve	-0.10975	0.04190	-2.62	0.76217	-0.11312	0.04327	-2.61	0.71692
Jacket not lined	-0.36581	0.05247	-6.97	0.28988	-0.35700	0.05857	-6.10	0.23337
Jacket part lined					-0.00667	0.13364	-0.05	0.67598

Jacket waist length					0.00401	0.03584	0.11	0.83495
Jacket double breasted					0.02486	0.04267	0.58	0.88259
Jacket zipper					-0.11559	0.07283	-1.59	0.86976
Bottom not lined	-0.19332	0.04596	-4.21	0.31536	-0.19199	0.04841	-3.97	0.28513
No waistband	0.21380	0.04653	4.59	0.59210	0.19062	0.06409	2.97	0.31319
Set on waistband	0.16509	0.03349	4.93	0.49020	0.14065	0.05665	2.48	0.17186
Part elastic set on waistband					-0.03113	0.05552	-0.56	0.28490
Multicolor	0.04982	0.03038	1.64	0.88207	0.04889	0.03119	1.57	0.83975
Adornment	0.07586	0.03568	2.13	0.83961	0.09114	0.03891	2.34	0.70856
Belt	0.10539	0.06669	1.58	0.79517	0.11221	0.06863	1.64	0.75348
Western Europe	0.27056	0.12992	2.08	0.85397	0.29879	0.13310	2.24	0.81621
Hong Kong	0.20072	0.09801	2.05	0.94313	0.21404	0.10192	2.10	0.87499
Korea	0.10051	0.05387	1.87	0.84186	0.09331	0.05499	1.70	0.81054
Southeast Asia	-0.07830	0.02945	-2.66	0.71930	-0.07459	0.03053	-2.44	0.67136
Central America	-0.26254	0.05841	-4.49	0.81570	-0.28444	0.06607	-4.31	0.63961
Bangladesh	-0.50470	0.14105	-3.58	0.90388	-0.49895	0.14763	-3.38	0.82789
Mexico	-0.58084	0.07430	-7.82	0.78555	-0.57150	0.07581	-7.54	0.75687
Caribbean	-0.66248	0.16068	-4.12	0.92706	-0.60449	0.16254	-3.72	0.90883
Eastern Europe					0.10714	0.07496	1.43	0.82111
Other region					-0.05236	0.13409	-0.39	0.57665
Africa					-0.23096	0.20422	-1.13	0.86203
Independent	0.36199	0.07966	4.54	0.51091	0.37622	0.08303	4.53	0.47173
Apparel department	0.34696	0.05080	6.83	0.73003	0.36899	0.05394	6.84	0.64960
Full price women’s	0.13897	0.04501	3.09	0.64181	0.14964	0.04712	3.18	0.58737
Catalog	0.08252	0.07593	1.09	0.79749	0.10379	0.08912	1.16	0.58086
Full price family	-0.18933	0.06988	-2.71	0.84021	-0.16608	0.07210	-2.30	0.79194
Off price family	-0.39465	0.10026	-3.94	0.80262	-0.38711	0.10196	-3.80	0.77861
Discount family	-0.40816	0.07292	-5.60	0.73250	-0.44077	0.07652	-5.76	0.66723
Discount department	-0.53486	0.06443	-8.30	0.64772	-0.49902	0.06967	-7.16	0.55583
Off price department	-0.85314	0.05069	-16.83	0.82666	-0.82898	0.05194	-15.96	0.79004
Midwest region					-0.06223	0.03744	-1.66	0.52824
Southern region					-0.08414	0.03592	-2.34	0.46142
Western region					-0.05417	0.04105	-1.32	0.45742
B size city	-0.12077	0.02767	-4.36	0.80128	-0.09684	0.03037	-3.19	0.66728
C size city	-0.17146	0.06775	-2.53	0.80760	-0.13527	0.06969	-1.94*	0.76581
Adjusted R^2	0.8619				0.8615			
F Statistic	81.56				54.51			