Cardon Research Papers

in Agricultural and Resource Economics

Research Paper 2004-16
September 2004

Quantile Regression Analysis of Asymmetrically Distributed Residuals in Consumer Demand Equations

Lester D. Taylor
University of Arizona

Department of Agricultural and Resource Economics
College of Agriculture and Life Sciences
The University of Arizona

This paper is available online at http://ag.arizona.edu/arec/pubs/workingpapers.html

Preliminary: For comment only. July 2004

Abstract

Quantile Regression Analysis of
Asymmetrically Distributed Residuals
In Consumer Demand Equations

Lester D. Taylor
University of Arizona

Apart from examining for autocorrelation and heteroscedasticity, applied econometricians seldom give much attention to the properties of the stochastic terms of their regression models. Most of the time, least-squares estimation (in some form or another) is employed, based upon a belief (often no more than implicit) that the conditions needed for the validity of the Gauss-Markov and Classical Normal Central-Limit Theorems are ever present. In fact, there are a lot of reasons as to why real-world error terms may not behave in ways that these conditions require. Residuals from Engel curves and demand functions estimated from data from the BLS quarterly consumer expenditure surveys provides a rather striking instance of this, in that the distribution of these residuals are almost invariably asymmetrical and fat-tailed. The focus of the present exercise is on the use of quantile regression as a robust corrective.

This paper is part of an on-going analysis of data from the quarterly BLS consumer expenditure surveys. Earlier exercises include Taylor (2003, 2004a, 2004b, and 2004c). I am grateful to Sean McNamara of the American Chambers of Commerce Researchers Association (ACCRA) for making EXCEL files of ACCRA surveys available to me and to the Cardon Chair Endowment in the Department of Agricultural and Resource Economics at the University of Arizona for financial support. Construction of data sets and econometric estimation have all been done in SAS.

Quantile Regression Analysis of Asymmetrically Distributed Residuals In Consumer Demand Equations

Lester D. Taylor
University of Arizona

I. Introduction

Apart from examining for autocorrelation and heteroscedasticity, applied econometricians seldom give much attention to the properties of the stochastic terms of their regression models. Most of the time, least-squares estimation (in some form or another) is employed, based upon a belief (often no more than implicit) that the conditions needed for the validity of the Gauss-Markov and Classical Normal Central-Limit Theorems are ever present. In fact, there are a lot of reasons as to why real-world error terms may not behave in ways that these conditions require. If, for example, even one of the factors subsumed in an error term has an infinite variance, then neither of the theorems can be assumed to hold. Given the pervasiveness of phenomena (both natural and socio-economic, including in economics the distributions of income and wealth and of short-term speculative price changes) that are empirically described by scaling distributions with exponents greater than -2, that such phenomena may occasionally find presence in the error terms of econometric models is obviously to be expected. However, the concern of the present exercise is not with absence of second moments, per se, but rather with a phenomenon that has revealed itself steadily over the last couple of years in connection with my research using data from the BLS quarterly consumer expenditure surveys, namely, a rather marked asymmetry in the distributions of residuals from estimated Engel curves and consumer demand functions.

The graph in Figure 1 is typical. The graph in this figure is a kernel-smoothed density function of the residuals from a double-logarithmic equation fitted by ordinary least squares to 8056 observations on housing expenditures of individual households from the four quarterly BLS CES surveys for 1996. Quite clearly, the distribution is not symmetrical, but rather is left-skewed (i.e., has a longer tail on the left than on the right), and has peak density well to the right of its OLS mean of 0. As illustration of the analysis to follow, the graphs in Figures 2 and 3 show residuals from the same model for two quantile regressions. The graph in Figure 2 is for the residuals from a median quantile regression, which corresponds to estimation by minimizing the sum of absolute errors, while the graph in Figure 3 is for the residuals from a quantile regression estimated at the mode of the density function for the residuals from the median regression.[1]

---

[1] The model is a simple double-logarithmic equation that relates the logarithm of a households's housing expenditures to the logarithms of the household's total consumption expenditure and the logarithms of prices for food consumed at home, housing, utilities, transportation, health care, and miscellaneous expenditures, plus a list of socio-demographical variables. The data set will be described in Section III below and in greater detail in the appendix. The kernel-smoothing employed is described in a footnote in Section IV.

Figure 1

Kernel-Smoothed Density Function
OLS Residuals for Housing
BLS-CES Surveys 1996

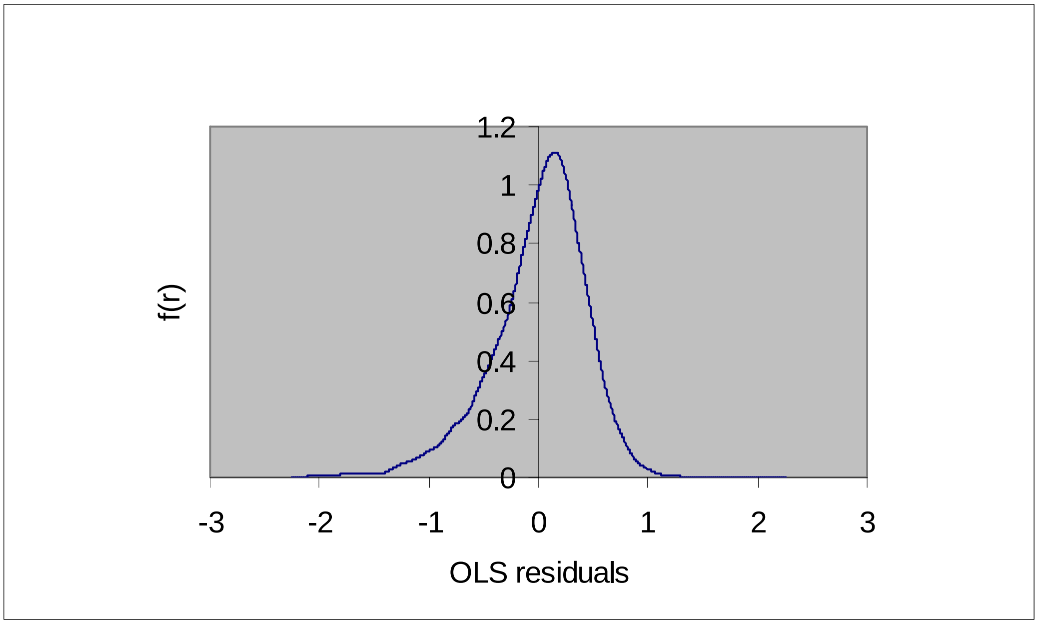


Figure 2

Kernel-Smoothed Density Function
Median Regression Residuals for Housing
BLS-CES Surveys 1996

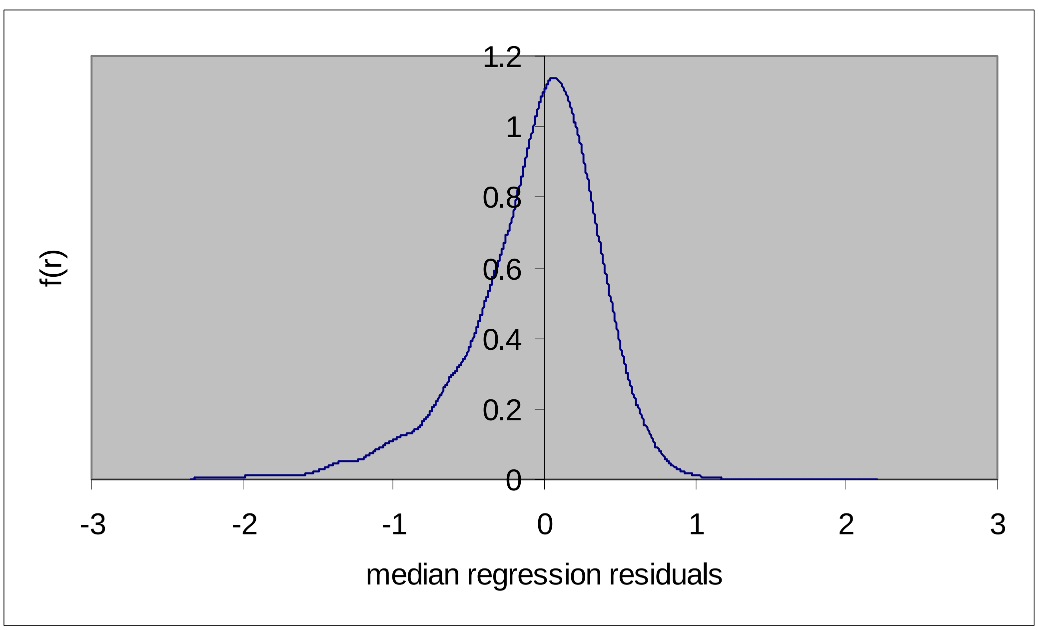
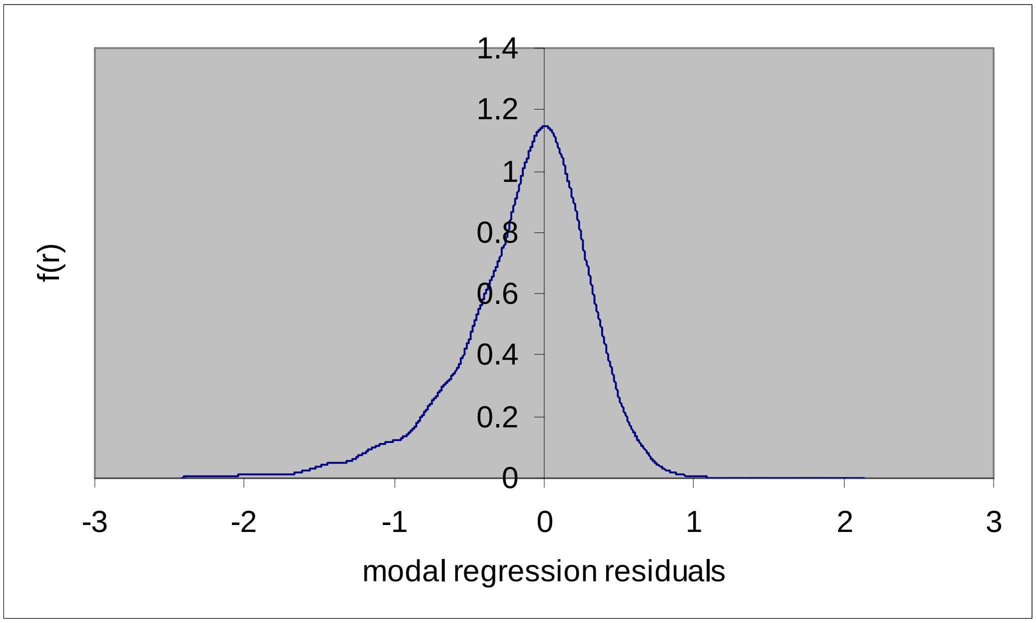


Figure 3

Kernel-Smoothed Density Function
Modal Regression Residuals for Housing
BLS-CES Surveys 1996



At this point, three things are to be noticed about these graphs. The first thing is that the asymmetrical density function of Figure 1 is definitely preserved in the quantile regressions. However, since the only material change is in the minimizing metric employed in estimation, this is probably to be expected. The second thing to notice is a lengthening of the lower tail of the density function for the quantile equations, especially for the residuals from the modal regression.[2] This latter, of course, is a reflection of the strong OLS sensitivity to outlier observations. Finally, the third thing to notice is the movement of the peak (or mode) of the densities to 0. The modal density, obviously, has the mode right at 0, and nearer to 0 for the median regression residuals than for the OLS residuals. All this, of course, is simply a reflection of the well-known relations to one another of the mean, median, and mode in skewed distributions.

While it might be thought that the use of a robust method of estimation might lead to the disappearance of the asymmetrical OLS residuals in this context, this quite clearly is not the case. Quantile regression, which is being increasingly viewed as one of the most powerful robust regression methods, appears only to lead to a sharpening of the asymmetry. Thus, the big question

---

[2] Outliers affect quantile regressions only in terms of numbers above and below the regression hyperplane at the specified quantile, while OLS regressions are affected by both numbers and value.

remains: Why such asymmetries? Are these a true implication of nature? Or are they simply a consequence of misspecification? Unfortunately, I do not know the answer to this question, but actually having an answer is not the concern of this exercise. My concern in what follows is to take the asymmetry at face value, and to suggest that quantile regression provides a useful instrument for mitigating its effects. Accordingly, in Section III an exercise involving quantile-regression estimation of an additive double-logarithmic demand system applied to a data set consisting of observations for individual households on six exhaustive categories of expenditure from the quarterly BLS consumer expenditures surveys will be undertaken. However, before doing this, a brief overview of the theory and mechanics of quantile regression is in order.

II. Quantile Regression

Since the seminal articles of 1978 of Bassett and Koenker, quantile regression has emerged as a powerful robust alternative to least-squares procedures in situations in which the assumptions underlying least-squares estimation are questionable. As a point of departure, consider the standard regression model, which postulates a relationship between a dependent variable y and N independent variables x (viewed as a non-stochastic row vector), in terms of the conditional expectation of y, given x:

$$E(y|x) \quad = \quad g(x), \tag{1}$$

for some function g. If g(x) is assumed to be linear in x, then we have the conventional linear regression model,

$$y \quad = \quad X\beta + u, \tag{2}$$

where u is assumed to be an unobservable stochastic term of mean zero with density function f(u). If f(u) has a finite variance, then, as is well-known, the Gauss-Markov for Least Squares states that the estimator of β that minimizes the quantity:

$$\varphi(y, X, \beta) \quad = \quad (y - X\beta)'(y - X\beta)\,, \tag{3}$$

with respect to β,

$$b \quad = \quad (X'X)^{-1} X'y\,, \tag{4}$$

is the minimum-variance unbiased estimator of β in the class of all unbiased estimators linear in y. If, in addition, f(u) should be distributed normally (even if only asymptotically), then all of the apparatus of conventional statistical inference is available for hypothesis testing and confidence interval construction.

However, as Koenker and Bassett note,[3] one of the dark secrets of applied statistics and econometrics is that faith in the validity of Gauss-Markov and Normality assumptions is often misplaced, in which case it is wise to consider non-linear, or even, biased, estimators that may be superior to least squares. One such estimator, which in Gauss and Laplace has essentially the same progenitors as least squares, is the one associated with minimizing the sum of absolute errors (as opposed to the sum of squared errors), or LAE. Historically, LAE was never able to compete with least squares because of mathematical intractability and lack of a suitable sampling theory. Fortunately, this is no longer the case, for in the 1960's it was recognized that LAE estimators can be obtained as solutions to linear programming problems,[4] and, thanks to Koenker and Bassett, a workable sampling theory is now available -- not only for LAE, but also for its generalization to quantile regression.

For a single random variable x, the LAE estimator is simply the median of x, that is, the point on the distribution of x for which half of the values of x lie above and half below. With the linear model, $y = X\beta + u$, the LAE estimator of β also corresponds to a median, though in this case to the "median" hyperplane , defined by Xb, for which half the values of y will lie "above" Xb and half will lie "below".[5] Quantile regression (QR) is a generalization of the LAE estimator to quantiles other than the median. Specifically, the quantile regression for the q^{th} quantile ($0 \leq q \leq 1$), the QR vector, b_q , for q will be given by minimizing with respect to β the expression:[6]

$$\theta(\beta|y, X, q) = \sum\rho_q|y - X\beta| , \tag{5}$$

where:

$$\rho_q = \begin{cases} q & \text{if } y - X\beta \leq 0 \\ 1 - q & \text{if } y - X\beta > 0. \end{cases} \tag{6}$$

---

[3] Koenker and Bassett (1978, p. 35).

[4] See Bassett and Koenker (1978, 1982), Koenker and Bassett (1978) , Hahn (1995), Buchinsky (1994a, 1994b, 1995, 1998), Horowitz (1998), Bassett and Chen (2000), and Mello and Perrelli (2003).

[5] What "above" and "below" mean in this context is that half of the residuals will be positive and half will be negative. In saying "half-and-half", however, I am ignoring the fact that (degeneracy aside) that N of the residuals (with N independent variables) will be exactly zero. The median for a single random variable can according be viewed as the LAE regression of the variable on a vector of 1's. Assuming that the linear model has an intercept (i.e., that one of the columns of X is a vector of 1's), the intercept with LAE can accordingly be interpreted as shifting the (N-1)-dimensional hyperplane defined by the N-1 non-constant variables so as to "split" the residuals half-and-half between positive and negative.

[6] See Bassett and Koenker (1978).

The solution vector b_q to this minimization problem is obtained straightforwardly as the solution to a linear programming problem.

Just as the LAE regression, which corresponds to q = 0.5, has half of the residuals positive and half negative, the quantile regression for q (and T observations) will have qT residuals less than or equal to 0 and (1 - q)T residuals greater than 0. This is true for all q. Under fairly lenient regularity conditions, the QR estimators are distributed asymptotically normal with mean vector β_q and covariance matrix $\omega\,(X'X)^{-1}$, where

$$\omega = q(1 - q)/f(q)^2 \tag{7}$$

and f(q) is the value of the density function of u at q.[7] These asymptotic results hold even if u should lack a second moment.

III. Quantile Regression Applied to CES Demand Equations

We now turn to the main purpose of this paper, which is investigation, using quantile regression, of asymmetrically distributed residuals from consumer demand functions estimated with data from the U. S. BLS consumer expenditure surveys. However, unlike in the introduction, which illustrated the phenomenon using residuals from a simple double-logarithmic demand function, we will now utilize a system of additive double-logarithmic demand equations, as applied to six exhaustive categories of expenditure. The six categories of expenditure are food consumed at home, housing, utilities, transportation, health care, and miscellaneous expenditures.

Despite their often superior fit and convenience of estimation, simple double-logarithmic demand functions (i.e., equations in which logarithms of goods purchased are related linearly to logarithms of income and prices) have the important drawback that they are not additive, in that the sum of predicted expenditures from them do not satisfy the budget constraint. The demand functions to be employed here start out double-logarithmic as just described, but are made additive using a mathematical device first introduced by Houthakker in his famous *Econometrica* paper of 1960 on additive preferences.[8]

The mathematical device in question enables any non-additive function $\theta_i(y)$ to be made additive in terms of y by the transformation,

---

[7] Provided, of course, that f(q) is not 0. See Koenker and Bassett (1978). A number of finite sample approximations, including bootstraps, to the asymptotic covariance matrix have been proposed and investigated in Monte Carlo studies. Buchinsky (1995) provides a good discussion and evaluation of these.

[8] A more detailed discussion of this model is given in Taylor (2004c).

(8) $$g_i(y) = \frac{y\theta_i(y)}{\sum \theta_k(y)} \,,$$

since $\sum g_i(y) = y$. The application of this transformation to the double-logarithmic expenditure function,

(9) $$p_i q_i \quad = \quad A_i y^{\beta_i} \prod_{j=1}^{n} p^{\gamma_j} \;, \; i, j = 1, \ldots, n \,,$$

then gives an additive system of functions:

(10) $$f_j(y, p) = \frac{A_j y^{\beta_j} \prod_{k=1}^{n} p^{\gamma_k}}{\sum A_j y^{\beta_j} \prod_{k=1}^{n} p^{\gamma_k}} \,, \quad j = 1, \ldots, n.$$

The denominator in this expression for $f_j(y, p)$ is obviously a very complicated function of prices (p) and income (y), indeed so much so that estimation of the functions directly is pretty much intractable. However, following Houthakker's 1960 derivation of the indirect addilog model, the messy denominators can be eliminated through division of $f_j(y, p)$ by $f_i(y, p)$, so that:

(11) $$\frac{p_j q_j}{p_i q_i} = \frac{A_j y^{\beta_j} \Pi p^{\gamma_k}}{A_i y^{\beta_i} \Pi p^{\gamma_k}} \,.$$

Upon taking logarithms, this expression becomes:

(12) $$\ln p_j q_j - \ln p_i q_i \quad = \quad a_{ij} + (\beta_j - \beta_i)\ln y + \sum(\gamma_{jk} - \gamma_{ik})\ln p_k \,, \quad i, j, k = 1, \ldots, n, \; j \neq i \,,$$

where $a_{ij} = \ln A_j - \ln A_i$.

Expression (12) is thus seen to be consist of n - 1 double-logarithmic equations, in which the "dependent" variables are logarithmic differences, and the "independent" variables are the logarithms of income and the n prices. The coefficients that are estimated in the these equations are not β_j and γ_{jk}, but rather $(\beta_j - \beta_i)$ and $(\gamma_{jk} - \gamma_{ik})$, which would appear to leave the individual β's and γ's unidentified. However, by making use of a variety of additivity and other constraints, estimates of these underlying parameters can be obtained.[9]

[9] The particular identifying restrictions imposed will be discussed below.

We now turn to the empirical analysis. The equations in expression (12), with n = 6, are applied, using both OLS and QR, to a data set consisting of 8056 observations from the four quarterly U. S. BLS consumer expenditure surveys for 1996, augmented with price data obtained from quarterly cost-of-living surveys conducted by the American Chambers of Commerce Researchers' Association (ACCRA).[10] The 'left-out" category in the equations is miscellaneous expenditures. Although the model is formulated in terms of expenditures, the equations have been estimated with $\ln q_j - \ln q_i$ as the dependent variables, rather than $\ln p_j q_j - \ln p_i q_i$.[11] The independent variables in the equations are, as noted in footnote 1, the logarithms of total expenditure and of the six prices, plus socio-demographical and regional variables, such as age and education of head of household, family size, ethnicity, and region of residence.[12]

The estimated regression coefficients and t-values (asymptotic for the QR equations) for the five estimating equations are tabulated in Table 1, while the kernel-smoothed distributions of residuals are given in Figures 4 - 8.[13] The coefficients in the quantile regressions are generally very similar, as are also the coefficients on total expenditure for OLS and QR. Often, however, there are rather marked differences between the OLS and QR price coefficients.[14] On the other hand, the real

---

[10] See http://www.ACCRA.com for descriptions of these surveys. A list of the items included in the surveys is given in the appendix.

[11] Since the information collected in the CES surveys refer to $p_j q_j$, rather than to q_j and p_j separately, the logarithms of "quantities" have been calculated as $\ln p_j q_j - \ln p_j$, where p_j is the ACCRA price for category j. Because both expenditures and prices are in logarithms, the only difference in estimation (with $\ln q_j - \ln q_i$ as the dependent variables, rather than $\ln p_j q_j - \ln p_i q_i$) is that the coefficients on $\ln p_j$ and $\ln_i$ differ by 1 (or -1).

[12] A full listing of the variables included is given in the appendix.

[13] To save space, graphs for the QR median residuals are suppressed. As in Figures 2 and 3, the QR median and mode residuals are very similar, the only real difference being movement of the mode to 0 in the modal regression.The kernel-smoothed density functions $[g(u_i)]$ have been calculated using the unit normal density function as the kernel weighting function and a 'support' of k = 1000 intervals. Silverman's rule-of-thumb,

$$h = (0.9)\min[\text{std. dev., interquartile range}/1.34](N^{-1/5}),$$

has been used for the smoothing parameter h. Two standard references for kernel density estimation are Silverman (1986) and Wand and Jones (1995). Ker and Goodwin (2000) provide an interesting practical application to the estimation of crop insurance rates.

[14] It must be kept in mind that the coefficients in Table 1 refer to elasticity *differences*, rather than elasticities themselves. Tables of the latter will be presented in the next section.

Table 1

Additive Double-Log Regression Equations
CES Surveys 1996
(t-ratios in parentheses)

Variable	Food OLS	Food QR(Med.)	Food QR(Mode)	Housing OLS	Housing QR(Med.)	Housing QR(Mode)	Utilities OLS	Utilities QR(Med.)	Utilities QR(Mode)
pfood	-1.2085	-0.7179	-0.7668	-0.9667	-0.8665	-0.8494	-0.9727	-0.8703	-0.7235
	(-3.45)	(-2.10)	(-2.26)	(-2.69)	(-2.45)	(-2.40)	(-2.51)	(-2.44)	(-2.07)
phous	0.2840	0.1006	0.1074	-0.6254	-0.6717	-0.6665	0.2548	0.1471	0.1289
	(3.19)	(1.16)	(1.25)	(-6.85)	(-7.48)	(-7.43)	(2.59)	(1.62)	(1.45)
putil	0.3134	0.2227	0.2266	0.5077	0.3997	0.4046	-0.4839	-0.6233	-0.7446
	(2.79)	(2.03)	(2.08)	(4.40)	(3.52)	(3.56)	(-3.89)	(-5.45)	(-6.64)
ptrans	-0.0960	-0.1484	-0.1435	0.2314	0.1815	0.1727	0.0509	-0.0430	-0.0605
	(-0.76)	(-1.20)	(-1.17)	(1.78)	(1.42)	(1.35)	(0.36)	(-0.33)	(-0.48)
phlthcare	0.0026	0.2563	0.2198	0.2337	0.1938	0.2031	-0.3291	-0.0755	-0.1435
	(0.02)	(1.56)	(1.35)	(1.35)	(1.14)	(1.19)	(-1.77)	(-0.44)	(-0.85)
pmisc	0.8421	0.4984	0.6040	0.6745	0.7991	0.7673	1.5005	1.4356	1.4604
	(2.38)	(1.45)	(1.76)	(1.86)	(2.24)	(2.15)	(3.84)	(3.99)	(4.14)
totexp	-0.9106	-0.8944	-0.8998	-0.4279	-0.3668	-0.3685	-0.8428	-0.8381	-0.8377
	(-43.46)	(-43.84)	(-44.43)	(-19.91)	(-17.35)	(-17.44)	(-36.40)	(-39.34)	(-40.14)
R^2	0.3962	0.3933	0.3931	0.2289	0.2267	0.2264	0.3172	0.3144	0.3137
mode			0.4784			0.4962			0.4559

Variable	Transportation OLS	Transportation QR(Med.)	Transportation QR(Mode)	Health Care OLS	Health Care QR(Med.)	Health Care QR(Mode)
pfood	-1.4176	-0.5409	-0.3084	-1.9658	-1.8155	-1.7564
	(-2.99)	(-1.17)	(-0.68)	(-3.86)	(-3.42)	(-3.33)
phous	0.1731	0.0125	-0.0040	0.0525	0.0635	0.0098
	(1.44)	(0.17)	(-0.03)	(0.41)	(0.47)	(0.07)
putil	0.1650	0.0136	0.0389	0.4332	0.1625	0.0882
	(1.08)	(0.09)	(0.27)	(2.65)	(0.95)	(0.52)
ptrans	-1.2485	-1.1576	-1.2697	0.0494	-0.0228	-0.0171
	(-7.28)	(-6.96)	(-7.81)	(0.27)	(-0.12)	(-0.09)
phlthcare	- 0.5675	-0.3620	-0.2003	-0.6667	-0.8908	-0.9595
	(-2.49)	(-1.63)	(-0.92)	(-2.72)	(-3.49)	(-3.78)
pmisc	3.0436	2.2411	1.9042	1.7214	2.1671	2.3660
	(6.35)	(4.82)	(4.20)	(3.35)	(4.05)	(4.44)
totexp	0.1704	-0.1018	-0.1318	-0.7661	-0.7730	-0.7765
	(6.00)	(-3.70)	(-4.90)	(-25.16)	(-24.37)	(-24.59)
R^2	0.0705	0.0560	0.0531	0.3064	0.3039	0.3039
mode			0.4545			0.5209

Figure 4

Kernel-Smoothed Density Functions
OLS and QR(mode) Residuals
Food Consumed at Home

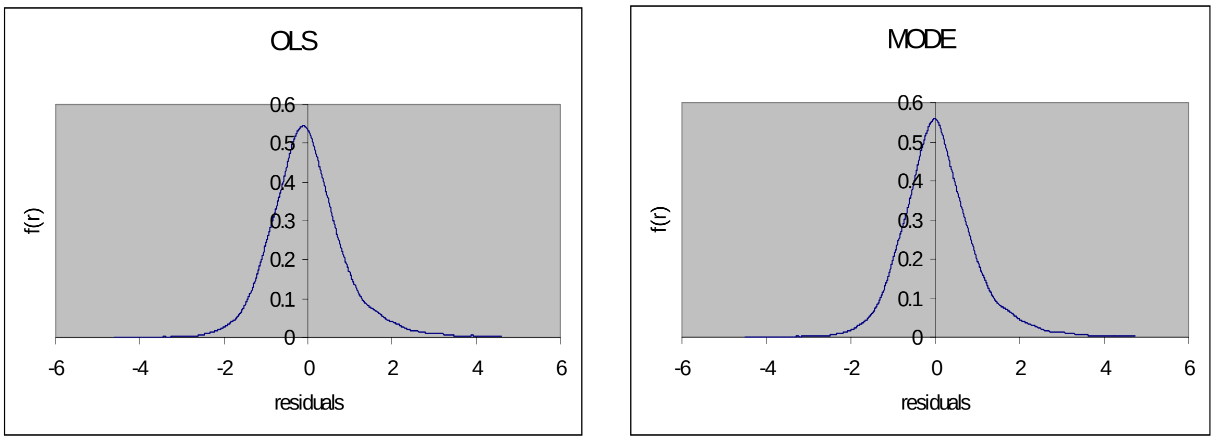


Figure 5

Kernel-Smoothed Density Functions
OLS and QR(mode) Residuals
Housing

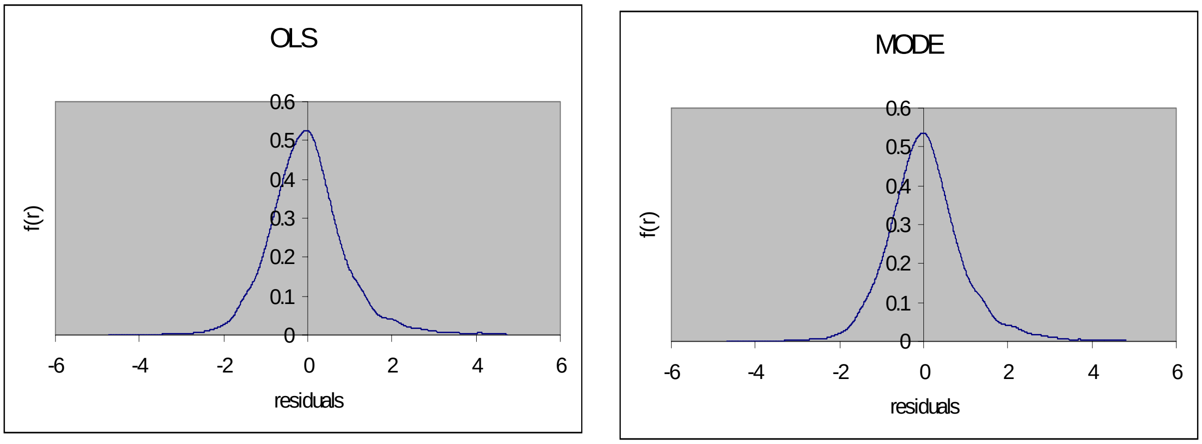


Figure 6

Kernel-Smoothed Density Functions
OLS and QR(mode) Residuals
Utilities

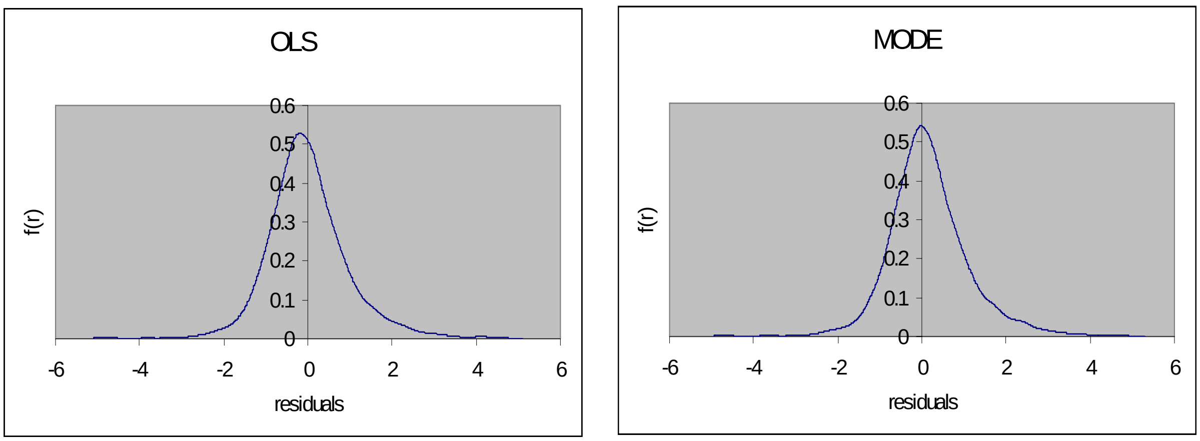


Figure 7

Kernel-Smoothed Density Functions
OLS and QR(mode) Residuals
Transportation

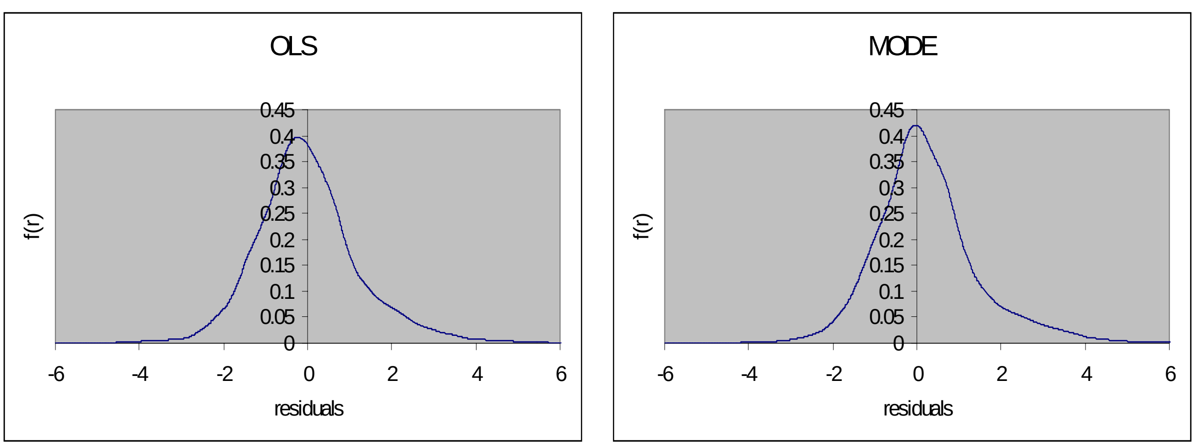
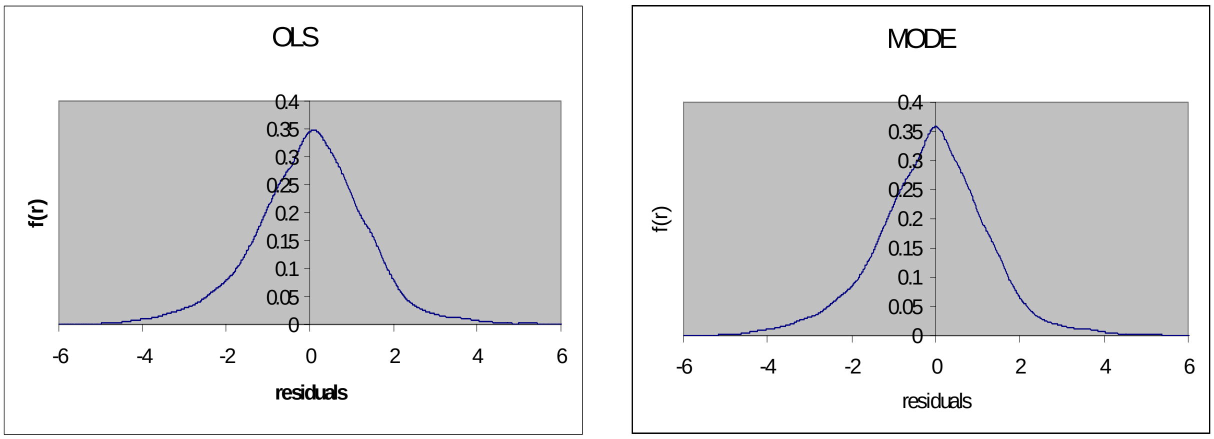


Figure 8

Kernel-Smoothed Density Functions
OLS and QR(mode) Residuals
Health Care



story (for now) is in Figures 4 - 8, for in these we see that all of the residuals distributions (except for the one for health care) are *skewed* and *all have a long tail (or tails).* These long tails are obviously very ominous for least-squares estimation.

What is particularly interesting about the skewness and long-tails in Figures 4 - 8 is that, unlike for the residuals in Figures 1 - 3, the error terms v_j underlying the residuals in these figures are themselves transforms, specifically, differences either of the form:

(13) $$v_j = u_j - u_i, \quad j \neq i$$

or

(14) $$v_j = \ln u_j - \ln u_i, \quad j \neq i,$$

where u_j denotes the error term attaching to the demand function for q_j in expression (9).[15] Accordingly, one might have thought that the skewness and fat tails manifested by the residuals in Figures 1 - 3 would be undone by the differencing, but Figures 4 - 8 show that this is obviously not the case.

---

[15] Expression (13) will hold if the error term appended to q_j is exponential, while (14) will if hold is the error term is multiplicative. At this point, since skewness and long tails in the estimating equations are facts, it makes little difference for present purposes which of the two specifications is taken to be correct.

What are the implications of the skewness and long tails displayed in these figures? In truth, I am not really sure. The only thing that is really clear is that long tails (since they carry the danger of infinite variance) are ominous for least squares estimation, but not for quantile regression. Skewness, on the other hand, would seem to be another matter, for there is nothing in the statement of the Gauss-Markov theorem that precludes error distributions being asymmetrical about their means (so long, that is, that they are not so skewed as to have an infinite variance). However, since quantile regression is clearly more robust than least squares in the presence of fat tails, and since it can be applied at any quantile, it seems reasonable (at least to me) to deal with skewness by moving the quantile of estimation to the mode.

IV. Price and Total Expenditure Elasticities[16]

The price and total expenditure elasticities calculated from the coefficients in Table 1 are tabulated in Tables 2 - 4.[17] Three things stand out in these tables:

---

[16] For other estimates of total expenditure and price elasticities for the data set used in this exercise using a variety systems of demand functions (both non-additive and additive), see Taylor (2004a, 2004b, 2004c).

[17] The elasticities in these tables are calculated according to the following formulae:

$$\eta_{tot.exp.} = \beta_j, \quad j = 1, ..., 6.$$

$$\eta_{ownprice} = (1 - w_j)\gamma_{jj} - \sum w_k \gamma_{jk}, \; k = 1, ..., 6, \; k \neq j$$

$$\eta_{cross\text{-}price} = - \sum w_k \gamma_{jk}, \quad i, k = 1, ..., 6, \; k \neq j,$$

where w_j is the budget weight of the j^{th} expenditure category. The elasticities, it should be noted, are aligned by column. Hence, the cross-elasticity for food with respect to the price of housing is 0.4058 (not -0.0750, which is the cross-elasticity of housing with respect to the price of food). The six β_j's are easily obtained from the five coefficients on lny, plus a sixth equation representing the constraint that the budget-share weighted income elasticities sum to 1. Calculation of the γ_{jk}'s, on the other hand, is a bit more complicated. Thirty-six equations are required to solve for them, 30 of which are obviously the equations connecting the γ's to the coefficients on $\ln p_j - \ln p_i$ in the five estimating equations. One would then think that the Hicks-Allen additivity conditions (i.e., that the income and own- and cross-price elasticities sum to 0 for each expenditure category) would provide the additional equations needed for identication. However, this is unfortunately not the case, for when linear relationships embodying these restrictions are included, 5 of the 6 price coefficients in the “left-out” equation turn out to be co-linear with the 31 other coefficients. Consequently, to achieve identification, I have assumed that the own-price elasticity for food is the negative of food’s total expenditure elasticity. This identifies γ_{11} (and therefore implicitly γ_{61}). For the remaining identifying restrictions, I have used discrepancy is in the own-price elasticities for health care -- -0.94 for the modal QR value vs. -0.58 for OLS.

(1). There is close agreement across the three tables in estimates of the total expenditure elasticities.

(2). There is also fairly close agreement, although not as strong as for the expenditure elasticities, in the own-price elasticities. The biggest discrepancy is in the own-price elasticities for health care -- -0.94 for the modal QR value vs. -0.58 for OLS.

(3). Estimates of cross-price elasticities are the most affected between OLS and QR, particularly in the estimates for transportation expenditures. This, in all likelihood, is a reflection of the fact that the

(4). As noted, the own-price elasticities for the QR median and modal models are close in magnitude, but this is less the case for the cross-elasticities. Not surprisingly, differences (for a given expenditure category) seem pretty much to be related to discrepancies between the median and the mode. For housing, the mode of distribution of the residuals (as calculated from the median QR regression) is virtually the same as the median (the mode is at quantile 0.4962), and there is in general little difference in the estimated elasticities. On the other hand, the differences are noticeably larger (again, as a general proposition) in the cross-price values for utilities and transportation, which would appear to be a consequence that, for both of these categories, the mode differs from the median by nearly five percentiles.[18]

V. Conclusions

The focus in this study has been on what may be an unappreciated problem in applied econometrics, namely, a situation in which the distribution of regression residuals is asymmetrical with fat tails, a circumstance that clearly has ominous implications for least-squares estimation. The "corrective" proposed in this exercise has been the use of quantile regressions, which is an increasingly used robust regression procedure that corresponds to estimation by minimizing the sum of absolute errors at particular quantiles on the distribution of a model's residuals. Two quantile regressions have been estimated, the first at the quantile corresponding to the median of the

---

the five apparent co-linear relationships between γ_{6k} (for k = 2, ..., 6), γ_{61} , and γ_{ji} for j = 1, ..., 5 and i = 1, ..., 6.

[18] This, of course, is obviously only an argument of *prima facie* plausibility.

Table 2

Price and Total Expenditure Elasticities
Additive Double-Logarithmic Model
BLS-ACCRA Surveys 1996
OLS Regressions
(calculated at mean values)

	Food	Housing	Utilities	Trans.	Healthcare	Misc.	Total Expenditure
Food	-0.4275	-0.0750	-0.0811	-0.5258	-1.0742	0.8916	0.4275
Housing	0.4058	-0.5036	0.3766	0.2949	0.1743	0.1218	0.9102
Utilities	0.0927	0.2870	-0.7046	-0.0558	0.2125	-0.2207	0.4953
Trans.	0.0308	0.3583	0.1778	-1.1216	0.1763	0.1269	1.5085
Healthcare	0.0889	0.3100	-0.2428	-0.4912	-0.5806	0.0863	0.5721
Misc	-0.3261	-0.4678	0.3323	1.8754	2.4787	1.1682	1.3381

Table 3

Price and Total Expenditure Elasticities
Additive Double-Logarithmic Model
BLS-ACCRA Surveys 1996
Median QR Regressions
(calculated at mean values)

	Food	Housing	Utilities	Trans.	Healthcare	Misc.	Total Expenditure
Food	-0.4556	-0.2239	-0.2277	0.1017	-1.1729	0.6426	0.4556
Housing	0.2929	-0.4793	0.3395	0.2049	0.2559	0.1923	1.0112
Utilities	0.1008	0.2779	-0.7452	-0.1082	0.0407	-0.1219	0.5119
Trans.	-0.0013	0.3286	0.1041	-1.0105	0.1243	0.1471	1.2481
Healthcare	0.2821	0.2196	-0.0497	-0.3362	-0.8651	0.0258	0.5770
Misc	-0.5663	-0.2656	0.3709	1.1764	3.1780	-1.0647	1.3500

Table 4

Price and Total Expenditure Elasticities
Additive Double-Logarithmic Model
BLS-ACCRA Surveys 1996
Modal QR Regressions
(calculated at mean values)

	Food	Housing	Utilities	Trans.	Healthcare	Misc.	Total Expenditure
Food	-0.4658	-0.2591	-0.1332	0.2819	-1.1661	0.5903	0.4658
Housing	0.3047	-0.4692	0.3262	0.1994	0.2071	0.1973	0.9971
Utilities	0.1096	0.2910	-0.8583	-0.0747	-0.0254	-0.1136	0.5279
Trans.	0.0647	0.3842	-0.3936	-1.0582	0.1932	0.2115	1.2339
Healthcare	0.1790	0.2141	-0.1324	-0.1893	-0.9484	0.0110	0.5892
Misc	-0.4193	-0.2509	0.4402	0.8860	3.2860	-1.0202	1.3657

regression residuals and the second at the mode of the median regression's residuals. The procedure has been "illustrated" with a system of additive double-logarithmic demand functions applied to a cross-sectional data set consisting of 8056 household observations for six exhaustive expenditure categories from the four quarterly BLS consumer expenditure surveys (augmented with price data collected quarterly by the American Chambers of Commerce Researchers Association) for 1996.

The residuals from the estimating equations (whether from least squares or quantile regressions) all display the asymmetries in question. Empirically, in the face of these asymmetries, OLS appears reasonably robust in the estimation of income (or total-expenditure) effects, but much less so in estimation of price (especially cross-price) effects. Not surprisingly, differences appear particularly sensitive to low R^2's and strong asymmetry in the distribution of residuals. In terms of magnitude, the empirical estimates of total expenditure elasticities from the modal QR regressions (and even the OLS models, for that matter) make sense, the only mild surprise (for me) being the fairly high value (1.23) for transportation expenditures. The own-price elasticities also seem plausible.[19] The cross-price elasticities, though, are another matter, for accumulated knowledge of cross-price effects is in general meager, hence remarks concerning their plausibility pretty much lack a basis.[20]

---

[19] However, it must be kept in mind that in order to identify the parameters for the "left out" category (miscellaneous expenditures), I have assumed that the own-price elasticity for food is the negative of the total expenditure elasticity.

[20] The cross-price elasticity in Table 4 that most obviously stands out is the value of 3.29 for the elasticity of health care with respect to the price of miscellaneous expenditures. It will be interesting to see whether this strong substitution effect holds up in future research, or is simply an artefact of this particular data set.

In closing, I would like to emphasize that the asymmetries that have been the focus of this exercise are a real phenomenon. I say this on the basis of estimating numerous (indeed, hundreds!) of Engel curves and demand functions with data for 1996 through 1999 from the quarterly BLS consumer expenditure surveys. Virtually without exception, residuals from the equations that have been examined have an asymmetrical distribution. To use estimation that is more robust than least squares in this situation seems mandatory. Whether quantile regression at the mode of the distribution of the residuals is the best for the circumstances, I do not know. However, as mentioned earlier, the procedure has intuitive appeal, and this seems sufficient reason, for now, for investigating its use.

REFERENCES

Bassett, G. W. and Cen, H. L. (2000), "Quantile Style: Return-Based Attribution Using Quantile Regression," Department of Finance, University of Illinois at Chicago, December 2000.

Bassett, G. W. and Koenker, R. (1978), "The Asymptotic Distribution of the Least Absolute Error Estimator," *Journal of the American Statistical Association*, Vol. 73, pp. 618-622.

Bassett, G. W. and Koenker, R. (1982), "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, Vol. 77, No. 378, June 1982, pp. 407-415.

Buchinsky, M. (1994), "Changes in the U. S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, Vol. 62, pp. 405-458.

Buchinsky, M. (1994a), "Quantile Regression, Box-Cox Transformation Model, and the U. S. Wage Structure," Journal of Economics, Vol. 65, pp. 109-154.

Buchinsky, M. (1995b), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models," *Journal of Econometrics*, Vol. 68, pp. 303-308.

Buchinsky, M. and Hahn, J. (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, Vol. 66, No. 3, May 1998, pp. 653-671.

Hahn, J. (1995), "Bootstrapping Quantile Regression Models," *Econometric Theory*, Vol. 11, pp. 307-333.

Horowitz, J. (1998), "Bootstrap Methods for Median Regression Models," *Econometrica*, Vol. 66, pp. 1327-1352.

Houthakker, H.S. (1960), "Additive Preferences," Econometrica, Vol. 28, pp. 244-257.

Ker, A.P. and Goodwin, B.K. (2000), "Nonparametric Estimation of Crop Insurance and Rates Revisited," *American Journal of Agricultural Economics*, Vol. 83, May 2000, pp. 463-478.

Koenker, R. and Bassett, G. W. (1978), "Regression Quantiles," Econometrica, Vol. 46, No. 1, January 1978, pp. 33-50.

Mello, M.and Perrelli, R. (2003), "Growth Equations: A Quantile Regression Exploration," *The Quarterly Review of Economics and Finance*, Vol. 43, pp. 643-667.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability 26, Chapman and Hall, London.

Taylor, L. D. (2003), "Stability of U.S. Consumption Expenditure Patterns: 1996 - 1999," Department of Agricultural and Resource Economics, University of Arizona, Tucson, Arizona.

Taylor, L. D. (2004a), "Price and Income Elasticities Estimated From BLS Consumer Expenditure Surveys and ACCRA Price Data," Department of Agricultural and Resource Economics, University of Arizona, Tucson, Arizona.

Taylor, L. D. (2004b), "Estimation of Theoretically Plausible Demand Functions From U. S. From U. S. Consumer Expenditure Survey Data, Department of Agricultural and Resource Economics, University of Arizona, Tucson, Arizona.

Taylor, L. D. (2004c), "An Additive Double-Logarithmic Demand System," Department of Agricultural and Resource Economics, University of Arizona, Tucson, Arizona.

Wand, M.P. and Jones, M.C. (1995), *Kernel Smoothing*, Monographs on Statistics and Applied Probability 60, Chapman and Hall, London.

Appendix

A. Consumption Expenditure Categories Included in ACCRA Price Surveys.

Groceries	Housing	Utilities	Transportation	Health Care	Miscellaneous
t-bone stk.	apt. rent	all electric	bus fare	hosp. room	hamburger sand.
gd. beef	home price	part electric	tire bal.	Dr. appt.	pizza
sausage	mortgage rate	other energy	gasoline	dentist	2-pc. chicken
fry chicken	home P+I	telephone		aspirin	hair cut
tuna					beauty salon
gal. milk					tooth paste
dz. eggs					shampoo
margarine					dry clean
parmesan cheese					men's shirt
potatoes					underwear
bananas					slacks
lettuce					washer repair
bread					newspaper
cigarettes					movie
coffee					bowling
sugar					tennis balls
cereal					monopoly set
sweet peas					liquor
tomatoes					beer
peaches					wine
Kleenex					
Cascade					
Crisco					
orange juice					
frozen corn					
baby food					
Coke					

B. Preparation of Data.

The CES quarterly data sets employed in the analysis have been developed from the Public Use Interview Microdata sets for 1996 that are available on CD-ROM from the U.S. Bureau of Labor Statistics.[21] "Cleansing" of the CES files included elimination of households with reported income of less than $5000 and then of households with zero (or negative)

[21] See http://www.bls.gov/cex/home.htm.

expenditures for the commodity category in question. The CES surveys do not include price data. The price data for the analysis are taken from the on-going price surveys of the 62 items of consumer expenditure listed in Table A1 above in more than 300 cities in the U.S. that are conducted quarterly by ACCRA[22]. From the 62 items of expenditure, ACCRA constructs six price indices (food, housing, etc.), and then from these an all-items index (which in principle are comparable, on a city basis, to BLS city CPI's). The ACCRA city indices in a state for each quarter are aggregated to the state level using city populations from the US Census of 2000 as weights.[23] The resulting ACCRA prices are then attached to CES households according to state of residence.[24]

C. Definitions of Variables.

lnfood	logarithm of expenditures for food consumed at home
lnhous	logarithm of housing expenditures
lnutil	logarithm of expenditures for household utilities
lntrans	logarithm of transportation expenditures
lnhealth	logarithm of health care expenditures
lnmisc	logarithm of miscellaneous consumption expenditures
lnincome	logarithm of household income
lntotexp	logarithm of total consumption expenditure
lnpfood	logarithm of price index for food consumed at home
lnphous	logarithm of price index for housing
lnputil	logarithm of price index for utility expenditures
lnptrans	logarithm of price index for transportation expenditures

[22] See http://www.ACCRA.com.

[23] See http://www.census.gov/Press-Release/www/2003/SF4.html

[24] In instances in which CES does not code state-of-residence for reasons of non-disclosure, the households in question are dropped.

lnphealth	logarithm of price index for health care expenditures
lnpmisc	logarithm of price index for miscellaneous expenditures
lnpall	logarithm of all-items price index
no_earnr	number of income earners in household
fam_size	size of household
age_ref	age of head of household
dsinglehh	dummy variable for single household
drural	dummy variable for rural area of residence
dnochild	dummy variable for no children in household
dchild1	dummy variable for children in household under age 4
dchild4	dummy variable for oldest child in household between 12 and 17 and at least one child less than 12
ded10	dummy variable for education of head of household: grades 1 through 8
dedless12	dummy variable for education of head of household: some high-school, but no diploma
ded12	dummy variable for education of head of household: high-school diploma
dedsomecoll	dummy variable for education of head of household: some college, but did not graduate
ded15	dummy variable for education of head of household: Bachelor's degree
dedgradschool	dummy variable for education of head of household: post-graduate degree
dnortheast	dummy variable for residence in northeast

dmidwest	dummy variable for residence in midwest
dsouth	dummy variable for residence in south
dwest	dummy variable for residence in west (excluded)
dwhite	dummy variable for white head of household
dblack	dummy variable for black head of household
dmale	dummy variable for male head of household
down	dummy variable for owned home
dfdstmps	dummy variable for household receiving food stamps
D1, D2, D3, D4	seasonal quarterly dummy variables.