Health, Econometrics and Data Group

THE UNIVERSITY of York

HEDG Working Paper 06/06

A Semiparametric Derivative Estimator in Log Transformation Models

Chunrong Ai
Edward Norton

July 2006

ISSN 1751-1976

A Semiparametric Derivative Estimator in Log Transformation Models

Chunrong Ai and Edward C. Norton
University of Florida, Gainesville, FL and the
University of North Carolina at Chapel Hill, Chapel Hill, NC

Abstract

This paper considers a regression model with a log-transformed dependent variable. The log transformed model is estimated by simple least squares, but computing the conditional mean of the dependent variable on the original scale given the explanatory variables analytically requires knowing the conditional distribution of the error term in the transformed model. We show how to obtain a consistent estimator for the conditional mean and its derivatives without specifying the conditional distribution of the error term. The asymptotic distribution of the estimator is derived. The proposed procedure is then illustrated with health expenditure data from the Medical Expenditure Panel Survey.

Key Words log transformation; conditional mean; series estimator; asymptotic distribution; derivative estimator

1 Introduction

Applied economists often estimate models with a log-transformed dependent variable. Common justifications for using the logarithmic transformation include to deal with a dependent variable that is badly skewed to the right and to compute elasticities (Manning, 1998). The conditional mean of the log-transformed dependent variable given the explanatory variables is usually estimated by simple least squares. The conditional mean of the dependent variable on the original scale, however, depends on the conditional distribution of the error term in the log transformed model. Consequently, the derivatives of the conditional mean on the original scale with respect to the explanatory variables also depend on the conditional distribution of the error term. Therefore, any estimates of the conditional mean and its derivatives must adjust for the error term distribution. Failure to account for the conditional distribution may lead to substantially biased estimates.

There are three approaches to account for the error term distribution. The first is the parametric approach, which specifies the conditional distribution of the error term parametrically and then computes the conditional mean and its derivatives either analytically (Manning, 1998; Mullahy, 1998; Manning and Mullahy, 2002) or numerically (Abrevaya, 2002). Such an approach, however, may yield misleading results if the functional form of the conditional distribution is misspecified. In practice, the functional form of the conditional distribution is rarely known. Hence, this approach is not robust.

The second approach decomposes the error term into a standard error multiplied by a standardized residual term. The standardized residual term is assumed to be independent of the explanatory variables and has an unknown distribution. The standard error is fully parameterized (Ai and Norton 2000; Abrevaya, 2002). This approach clearly imposes fewer restrictions on the conditional distribution of the error term than the first approach and hence is more robust. Still, this approach can yield biased estimates if either the independence condition on the standardized residual term is not satisfied or if the parameterization of the standard error is misspecified.

The third approach, which is adopted in this paper, assumes that the conditional distribution of the error term given the explanatory variables is completely unknown. This approach is semiparametric and therefore most robust. In particular, this approach allows for heteroskedasticity of any form, a problem that is deemed to be particularly difficult to deal with in practice (see Manning 1998). Under this specification, we show how to obtain consistent estimates for the conditional mean and its derivatives in Section 2. We also derive the asymptotic distribution of the proposed estimators and provide consistent estimates for the asymptotic variance in Section 3. Finally, we empirically implement our estimators with skewed health expenditure data from the Medical Expenditure Panel Survey in Section 4. Technical proofs are relegated to the Appendix.

2 Estimation

Before we introduce the model, we adopt the following notation conventions: bold lowercase letters denote random variables; standard lowercase letters denote the realizations of the random variables; and $E\{\cdot\}$ denotes the expectation taken with respect to the distribution of the bold letters in the bracket.

Assume that the original dependent variable $\mathbf{y}$ is possibly a nonlinear function of K explanatory variables $\mathbf{x}$, excluding a constant, and a random error term $\mathbf{u}$. The model is given by

$$\ln(\mathbf{y}) = h(\mathbf{x}, \beta_o) + \mathbf{u},$$

where $h(.,.)$ is a known measurable function, β_o is a vector of unknown parameters, and the error term $\mathbf{u}$ satisfies the conditional mean restriction $E\{\mathbf{u}|\mathbf{x} = x\} = 0$ for almost all x.

In most applications, the function $h(\mathbf{x}, \beta_o)$ is linear in parameters β_o. Here we allow $h(\mathbf{x}, \beta_o)$ to be nonlinear in both $\mathbf{x}$ and β_o. We assume that no element of $\mathbf{x}$ can be expressed as a function of the other elements, *e.g.*, there are no higher-order or interaction terms. This assumption is not as restrictive as it appears because it can always be satisfied by redefining the function h. For example, if $\mathbf{x}$ includes income and income squared, then the function $h(\mathbf{x}, \beta_o)$ can be redefined as a function of income.

Inverting the log function, we obtain the dependent variable on the untransformed scale

$$\mathbf{y} = \exp(h(\mathbf{x}, \beta_o) + \mathbf{u}) = \exp(h(\mathbf{x}, \beta_o)) \exp(\mathbf{u}).$$

The conditional mean of the original dependent variable $\mathbf{y}$ given the explanatory variables is

$$\begin{aligned} F(x, \beta_o) &= E\{\mathbf{y}|\mathbf{x} = x\} \\ &= \exp(h(x, \beta_o)) E\{\exp(\mathbf{u})|\mathbf{x} = x)\} \\ &= \exp(h(x, \beta_0)) D(x). \end{aligned}$$

The marginal effects of the explanantory variables are found by taking derivatives (or differences) of the conditional mean with respect to the continuous (discrete) regressors. Let $m = (m_1, m_2, ..., m_K)'$ denote a vector of non-negative integers and define $|m| = m_1 + m_2 + ... + m_K$. Define

$$\mu_m(x) = \partial^m F(x, \beta_o) \equiv \frac{\Delta^{|m|} F(x, \beta_o)}{\Delta x_1^{m_1} \Delta x_2^{m_2} ... \Delta x_K^{m_K}},$$

where Δ denotes either the derivative or the difference operator depending on whether x is continuous or discrete. For example, with $x = (x_1, x_2)$, x_1 continuous, x_2 a 0-1 dummy, and $m = (1, 1)'$, we have

$$\mu_m(x) = \partial^m F(x, \beta_0) = \frac{\partial F(x_1, 1, \beta_0)}{\partial x_1} - \frac{\partial F(x_1, 0, \beta_0)}{\partial x_1}.$$

For the same example with $m = (1, 0)'$, we have the marginal effect of a continuous variable x_1

$$\mu_m(x) = \partial^m F(x, \beta_0) = \frac{\partial F(x, \beta_0)}{\partial x_1};$$

and with $m = (0, 1)'$, we have the incremental effect for a discrete change in a dichotomous variable x_2

$$\mu_m(x) = \partial^m F(x, \beta_0) = F(x_1, 1, \beta_0) - F(x_1, 0, \beta_0).$$

For convenience, we define $\mu_m(x) = F(x, \beta_o)$ when $m = 0$. Thus, $\mu_m(x)$ encompasses the estimands of interest, such as the conditional mean, and the marginal and interaction effects of the explanatory variables.

The focus of this paper is to present a consistent estimator for the derivative $\mu_m(x)$ and for the average derivative $\mu_m = E\{\mu_m(\mathbf{x})\}$ and to derive the asymptotic distributions of these estimators.

To estimate the derivative $\mu_m(x)$, we need to estimate the conditional mean function $F(x, \beta_o)$ which depends on the unknown parameter β_o and the unknown function $D(x)$. The unknown parameter β_o can be estimated by standard regression techniques. Given a sample $\{y_i, x_i, i = 1, 2, ..., n\}$, let $\widehat{\beta}$ denote the least squares estimator of β_o. We shall not concern ourselves with the derivation of the asymptotic properties of $\widehat{\beta}$, which are well established. Instead we will assume that the model satisfies standard conditions so that the least squares estimator is $\sqrt{n}$ consistent.

Assumption 2.1. *The least squares estimator* $\widehat{\beta}$ *is* $\sqrt{n}$ *consistent.*

The unknown function $D(x)$ depends on the unknown conditional distribution of the error term $\mathbf{u}$ given the explanatory variables $\mathbf{x}$ and thus cannot be estimated by a simple parametric regression. We propose to use a parametric approximation and then estimate that parametric approximation by least squares. Specifically, for some integer J, let $p^J(x) = (p_1(x), ..., p_J(x))'$ denote the approximating functions so that there is a π such that the parametric function $p^J(x)'\pi$ approximates $D(x)$ well. Examples of the approximating functions include polynomials, splines, and Fourier series. Denote

$$\begin{aligned} P &= (p^J(x_1), p^J(x_2), ..., p^J(x_n))'; \\ \widehat{u}_i &= \ln(y_i) - h(x_i, \widehat{\beta}); \\ \widehat{Q} &= (\exp(\widehat{u}_1), \exp(\widehat{u}_2), ..., \exp(\widehat{u}_n))'. \end{aligned}$$

Then $D(x)$ is estimated by regressing $\exp(\widehat{u}_i)$ on $p^J(x_i)$:

$$\widehat{D}(x) = p^J(x)'(P'P)^{-1}P'\widehat{Q}.$$

The derivative of the conditional mean is now estimated by

$$\widehat{\mu}_m(x) = \partial^m[\exp(h(x,\widehat{\beta}))p^J(x)]'(P'P)^{-1}P'\widehat{Q}$$

and the average derivative is estimated by

$$\widehat{\mu}_m = \frac{1}{n}\sum_{i=1}^{n}\partial^m[\exp(h(x_i,\widehat{\beta}))p^J(x_i)]'(P'P)^{-1}P'\widehat{Q}.$$

In the following sections, we will derive the asymptotic distributions of $\widehat{\mu}_m(x)$ and $\widehat{\mu}_m$.

3 Asymptotic Results

We first derive the asymptotic properties of $\widehat{\mu}_m(x)$ and then derive the asymptotic properties of $\widehat{\mu}_m$. The derivation of the asymptotic properties of $\widehat{\mu}_m(x)$ draws heavily from Newey (1997). We begin by introducing some regularity conditions.

Assumption 3.1. $\{(y_i, x_i), i = 1, 2, ..., n\}$ *are drawn independently from the joint distribution of* $(\mathbf{y}, \mathbf{x})$.

This condition rules out dependent data and hence is restrictive. The main result however can be generalized to dependent data using the results of Ai and Sun (2005).

Let $\|B\| = \sqrt{trace(B'B)}$ be the Euclidean norm of matrix B. Also, let $\mathcal{X}$ be the support of x_i.

Assumption 3.2. *For every J there is a nonsingular constant matrix B such that: (i) the smallest eigenvalue of $E\{B \times p^J(\mathbf{x})p^J(\mathbf{x})' \times B\}$ is bounded away from zero uniformly in J; and (ii) there is a sequence of constants $\zeta_0(J)$ satisfying $\sup_{x\in\mathcal{X}} \left\|p^J(x)\right\| \leq \zeta_0(J)$ and $J = J(n)$ such that $\zeta_0(J)^2 J/n \to 0$ as $n \to \infty$.*

This condition imposes restrictions on the approximating functions. Conditions of this sort are common in the literature on series estimation; see Newey (1997) and Andrews (1991). When the density of $\mathbf{x}$ is bounded away from zero, the constant $\zeta_0(J)$ is computed for splines and power series as $c\sqrt{J}$ and cJ respectively for some constant c; and the restriction in part (ii) is satisfied by $J^2/n \to 0$ and $J^3/n \to 0$.

For any vector $\lambda = (\lambda_1, ..., \lambda_n)$, denote

$$\zeta_{|\lambda|}(J) = \max_{|\delta|\leq|\lambda|}\sup_{x\in\mathcal{X}}\left\|^{-1}\partial^\delta[p^J(x)]\right\|.$$

Denote $\boldsymbol{\varepsilon} = \exp(\mathbf{u}) - D(\mathbf{x})$. Because we approximate $D(x)$ by $p^J(x)'\pi$, the approximation error will cause bias in the proposed estimator. To control the bias, the approximation error must shrink to zero as more terms are added to the approximating functions. We now specify a rate of approximation for the approximating functions.

Assumption 3.3. *(i) $\mathcal{X}$ is compact; (ii) there are α and π such that*

$$\sup_{x\in\mathcal{X}} \left|\partial^\lambda (D(x) - p^J(x)'\pi)\right| = O(J^{-\frac{\alpha-|\lambda|}{K}}) \text{ as } J \to \infty$$

for any $\lambda \leq m$; (iii)$\sqrt{n}J^{-\frac{\alpha-|\lambda|}{K}} \to 0$; and (iv) $\frac{\zeta_{|\lambda|}(J)\sqrt{J}}{\sqrt{n}} \to 0$.

Assumption 3.4. *$E\{\boldsymbol{\varepsilon}^4|\mathbf{x} = x\}$ is bounded, and $E\{\boldsymbol{\varepsilon}^2|\mathbf{x} = x\}$ is bounded and bounded away from zero for all x.*

Assumption 3.5. *(i) For every x, $h(x, \beta)$ is twice continuously differentiable with respect to β in the neighborhood of β_0; (ii) $\exp(\ln(y) - h(x, \beta))$, $\frac{\partial \exp(\ln(y)-h(x,\beta))}{\partial\beta}$, and $\frac{\partial^2 \exp(\ln(y)-h(x,\beta))}{\partial\beta\partial\beta}$ satisfy the stochastic dominance condition in the neighborhood of β_0; and (iii) for each element β_r of β,*

$$Var(\frac{\partial \exp(\ln(\mathbf{y}) - h(\mathbf{x}, \beta_o))}{\partial\beta_r}|\mathbf{x} = x)$$

is bounded and

$$\left|\partial^\lambda E\{\frac{\partial \exp(\ln(\mathbf{y}) - h(\mathbf{x}, \beta_o))}{\partial\beta_r}|\mathbf{x} = x\} - \partial^\lambda p^J(x)'\pi_r\right| = O(J^{-(\alpha_r-|\lambda|)/K})$$

for some π_r and α_r.

Assumption 3.3(i) requires that the explanatory variables have bounded support. This condition can always be satisfied by discarding observations with large values. Assumption 3.3(ii) requires that the approximation error shrink at polynomial rate. This condition is satisfied by splines and power series approximations with α as the degree of smoothness of $D(x)$. Assumption 3.4 requires that the fourth conditional moment is bounded and the conditional variance is bounded both from above and below. This condition is common in the regression literature. Assumption 3.5 is needed so that we can replace the estimate $\widehat{\beta}$ by the true value. Assumption 3.5(ii) is a stochastic dominance condition that is commonly imposed in the nonlinear econometric literature.

Denote the regression residuals $\widehat{\varepsilon}_i = \exp(\widehat{u}_i) - \widehat{D}(x_i)$. Denote

$$\begin{aligned}
\sigma^2(x) &= E\{\boldsymbol{\varepsilon}^2|\mathbf{x}=x\}; \\
V_{nm}(x) &= \partial^m[\exp(h(x,\beta_o))p^J(x)]' \\
&\quad \times (P'P)^{-1}\sum_{i=1}^{n}\sigma^2(x_i)p^J(x_i)p^J(x_i)'(P'P)^{-1} \\
&\quad \times \partial^m[\exp(h(x,\beta_o))p^J(x)]; \\
\widehat{V}_{nm}(x) &= \partial^m[\exp(h(x,\widehat{\beta}))p^J(x)]' \\
&\quad \times (P'P)^{-1}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 p^J(x_i)p^J(x_i)'(P'P)^{-1} \\
&\quad \times \partial^m[\exp(h(x,\widehat{\beta}))p^J(x)].
\end{aligned}$$

The following theorem is proved in the Appendix.

Theorem 1. *Under Assumptions 2.1, and 3.1–3.5, we show: (1)*

$$V_{nm}(x)^{-1/2}\left(\widehat{\mu}_m(x) - \mu_m(x)\right)$$

has asymptotically standard normal distribution; and (2)

$$V_{nm}(x)^{-1/2}\widehat{V}_{nm}(x)^{1/2} \to 1 \text{ in probability.}$$

Part (1) of the theorem shows that the proposed estimator is consistent and asymptotically normally distributed. Part (2) provides a consistent estimator for the variance. These two results allow us to conduct the statistical inference on the estimands. For example, the t-ratio $\frac{\widehat{\mu}_m(x)}{\sqrt{v_{nm}(x)}}$ can be used for significance tests of the derivative $\mu_m(x)$. If $\left|\frac{\widehat{\mu}_m(x)}{\sqrt{v_{nm}(x)}}\right| > 1.96$, then the pointwise derivative $\mu_m(x)$ is statistically significant at the 5 percent level. The theorem also reveals that the estimate of the finite dimensional parameter has no effect on the asymptotic distribution of the estimated derivative. This result is not surprising because the estimator of the finite dimensional parameter converges to the true value at a faster rate.

The derivative estimator $\widehat{\mu}_m(x)$ and its asymptotic variance can be computed as follows:

(i) Estimate the log transformed model by least squares; save the predicted values on the log scale $h(x,\widehat{\beta})$ and the regression residuals $\widehat{u}_i = \ln(y_i) - h(x_i,\widehat{\beta})$;

(ii) regress $\exp(\widehat{u}_i)$ on $p^J(x_i)$ using robust standard errors; save the regression coefficients in $\widehat{\pi}$ and its heteroskedasticity-consistent covariance matrix in $\widehat{\Omega}$; and

(iii) compute $\widehat{\mu}_m(x) = \partial^m[\exp(h(x,\widehat{\beta}))p^J(x)]'\widehat{\pi}$ and $\widehat{V}_{nm}(x) = \partial^m[\exp(h(x,\widehat{\beta}))p^J(x)]'\widehat{\Omega}\partial^m[\exp(h(x,\widehat{\beta}))p^J(x)]$.

Next, we derive the asymptotic distribution of the estimated average derivative $\widehat{\mu}_m$. Denote

$$
\begin{aligned}
v_{nm} = & \frac{E\{(\partial^m[h(\mathbf{x},\beta_o)D(\mathbf{x})]-\mu_m)^2\}}{n} + \\
& \frac{\sum_{i=1}^n \partial^m[h(x_i,\beta_o)P^J(x_i)']}{n}(P'P)^{-1}P'\Sigma_m P(P'P)^{-1} \\
& \times \frac{\sum_{i=1}^n \partial^m[h(x_i,\beta_o)P^J(x_i)]}{n}
\end{aligned}
$$

where $\Sigma_m = diag(v^2(x_1),...,v^2(x_n))$ and $v^2(x) = E\{\boldsymbol{\varepsilon}^2|\mathbf{x}=x\}$. Denote

$$
\begin{aligned}
\widehat{v}_{nm} = & \frac{\sum_{i=1}^n(\partial^m[h(x_i,\widehat{\beta})\widehat{D}(x_i)]-\widehat{\mu}_m)^2}{n^2} + \\
& \frac{\sum_{i=1}^n \partial^m[h(x_i,\widehat{\beta})P^J(x_i)']}{n}(P'P)^{-1}P'\widehat{\Sigma}_m P(P'P)^{-1} \\
& \times \frac{\sum_{i=1}^n \partial^m[h(x_i,\widehat{\beta})P^J(x_i)]}{n}
\end{aligned}
$$

with $\widehat{\Sigma}_m = diag(\widehat{\varepsilon}_1^2,...,\widehat{\varepsilon}_n^2)$.

The following theorem is also proved in the Appendix.

Theorem 2. *Under Assumptions 2.1, and 3.1 - 3.5, we show that: (1) $v_{nm}^{-1/2}(\widehat{\mu}_m-\mu_m)$ has asymptotically standard normal distribution; and (2) $v_{nm}^{-1/2}\widehat{v}_{nm}^{1/2} \to 1$ in probability.*

Theorem 2 shows that the average derivative estimator is asymptotically normally distributed and provides a consistent estimator for the asymptotic variance. These results can be used for statistical inference on the average derivatives. For instance, if $\left|\frac{\widehat{\mu}_m}{\sqrt{v_{nm}}}\right| > 1.96$, then the average derivative μ_m is statistically significant at the 5 percent level. It is interesting to note that the estimated finite dimensional parameter has no effect on the asymptotic distribution of the average derivative estimator. This seems counterintuitive because both the parameter estimator $\widehat{\beta}$ and the average derivtive estimator converge at the same rate and, in a sequential estimation like ours, generally the estimator in the first step affects the asymptotic distribution of the estimator in the second step estimation. The estimated parameter $\widehat{\beta}$ affects the asymptotic distribtution of the average derivative estimator through two channels: one through $h(x_i, \widehat{\beta})$ and the other through the residual $\widehat{u}_i$. It so happens in this case that these two effects offset each other.

The average derivative estimator $\widehat{\mu}_m$ and its asymptotic variance can be computed as follows:

(i) Estimate the log transformed model by least squares; save the predicted values on the log scale $h(x, \widehat{\beta})$ and the regression residuals $\widehat{u}_i = \ln(y_i) - h(x_i, \widehat{\beta})$;

(ii) regress $\exp(\widehat{u}_i)$ on $p^J(x_i)$ using robust standard errors; save the regression coefficients in $\widehat{\pi}$ and its heteroskedasticity-consistent covariance matrix in $\widehat{\Omega}$;

(iii) compute both the sample mean and the sample variance of

$$\partial^m[\exp(h(x_i, \widehat{\beta}))p^J(x_i)]'\widehat{\pi},$$

denoted $\widehat{\mu}_m(x)$ and S; and

(iv) compute

$$\widehat{\nu}_{nm} = \frac{S}{n} + \frac{\sum_{i=1}^n \partial^m[h(x_i, \widehat{\beta})P^J(x_i)']}{n} \times \widehat{\Omega} \times \frac{\sum_{i=1}^n \partial^m[h(x_i, \widehat{\beta})P^J(x_i)]}{n}.$$

One potential criticism of our model specification is that the functional form of $h(\mathbf{x}, \beta)$ may be misspecified because the conditional mean of the log-transformed dependent variable given the explanatory variables is really unknown. The question then is whether our estimator is biased. It is interesting to note that our estimator is still consistent even if the function form of $h(\mathbf{x}, \beta)$ is misspecified. This is because $\ln(y) = h(x, \beta) + u$ and hence any bias resulting from misspecified $h(\mathbf{x}, \beta)$ will be corrected through the regression residuals u. However, if $h(\mathbf{x}, \beta)$ is correctly specified, then our estimator utilizes more information than the simple average derivative estimator proposed by Powell, Stock, and Stoker (1989) and hence is more efficient.

Another potential criticism of our approach is that the number of the approximating terms, J, is not uniquely determined by the sufficient conditions of Assumptions 2.1 and 3.1–3.5. In practice, these sufficient conditions are not very useful for choosing J. A feasible and practical way to determine J is to apply the cross-validation approach, which chooses J to minimize

$$\sum_{i=1}^n \left(\exp(\widehat{u}_i) - p^J(x_i)'(P'_{-i}P_{-i})^{-1}P'_{-i}\widehat{Q}_{-i}\right)^2$$

where P_{-i} and $\widehat{Q}_{-i}$ denote P and $\widehat{Q}$ with the i-th row deleted.

4 Empirical Example

4.1 Data and Model

We illustrate the methods by analyzing data from the Medical Expenditure Panel Survey 2000 Full Year Consolidated Data File, collected by the Agency for Healthcare Research and Quality. The MEPS contain data on health care services and expenditures, as well as insurance, demographics, and employment, for the general American population. We limited the sample to the 22,095 persons (out of the original 25,096 persons) who had complete responses and were not in active military duty in 2000. Of these, we analyzed the 12,222 who were aged 18 and older and had positive health care expenditures, to abstract from the concerns of modeling zero expenditures.

The dependent variable is the logarithm of total expenditures, which includes inpatient, outpatient, dental, and emergency room expenditures. The mean expenditure was $1,922, and ranged from $3 to $213,023 (see Table 1). Taking the logarithm of the dependent variable removed much of the skewness (from 12.27 down to .26).

We also controlled for the standard demographic characteristics and health status. The sample was 42.1 percent male, 12.7 percent African-American, and 17.8 percent Hispanic. The mean age was 48 years, and ranged from 1 to 90. Nearly one-half had more than a high school education. We also controlled for income in five broad categories.

We use the 9×1 column vector $\mathbf{x}$ to denote the main explanatory variables, not including age. Specifically, $\mathbf{x}_1$ is the constant term; $\mathbf{x}_2 - \mathbf{x}_5$ are dummy variables for male, African American, Hispanic, and education greater than 12 years; $\mathbf{x}_6 - \mathbf{x}_9$ are the dummy variables for the four income groups, with the lowest income group as the reference. We specify that the regression function is linear in parameters. The model also controls for age, age squared, and interactions between age and the other explanatory variables. The log-transformed model is given by:

$$\ln(\mathbf{y}) = \mathbf{x}'\beta_0 + (\mathbf{x} \times \mathbf{age})'\beta_1 + (\mathbf{x} \times \mathbf{age}^2)'\beta_2 + \mathbf{u},$$

which will be estimated by simple least squares. We subtracted 40 from age so that the referent person is 40 years old, then scaled age by dividing by 100. The cross-validation for the MEPS data suggests the following approximation in the second stage regression

$$\exp(u) = \mathbf{x}'\pi_0 + (\mathbf{x} \times \mathbf{age})'\pi_1 + (\mathbf{x} \times \mathbf{age}^2)'\pi_2 + \boldsymbol{\varepsilon}.$$

Hence,

$$p(x) = (x', age \times x', age^2 \times x')'$$

is a column vector with 27 elements. We estimate the unknown coefficients β_0 by simple least squares. Let $\widehat{\beta}$ denote the least squares estimator obtained by regressing $\ln(y_i)$ on x_i. Denote $\widehat{u}_i = \ln(y_i) - x_i'\widehat{\beta}$. Let $\widehat{D}(x)$ denote the predicted value obtained from regressing $\exp(\widehat{u}_i)$ on $p(x_i)$. The estimated conditional mean and its derivatives are

$$\begin{aligned}\widehat{\mu}_0(x) &= \exp(x'\widehat{\beta}_0 + (x \times age)'\widehat{\beta}_1 + (x \times age^2)'\widehat{\beta}_2)\widehat{D}(x)\\ &= \exp(x'\widehat{\beta}_0 + (x \times age)'\widehat{\beta}_1 + (x \times age^2)'\widehat{\beta}_2)\\ &\quad \times(x'\widehat{\pi}_0 + (x \times age)'\widehat{\pi}_1 + (x \times age^2)'\widehat{\pi}_2);\\ \widehat{\mu}_m(x) &= \partial^m[\widehat{\mu}_0(x)].\end{aligned}$$

We compute three estimates—mean expenditures, and the incremental effect of sex and of race—both for a representative person and for the entire sample. The representative person is a 40-year-old non-Hispanic white female with a high school education in the lowest income group (*i.e.*, $x_0 = (0,0,0,0,0,0,0,0,0)'$). For the representative person calculations are simplified because interactions with age (really age less 40) are all zero. For the representative person, mean expenditures are

$$\widehat{\mu}_0(x_0) = \exp(\widehat{\beta}_{01})(\widehat{\pi}_{01});$$

the incremental effect of changing sex from female to male for a referent person (*i.e.*, $\Delta x_{sex} = (0,1,0,0,0,0,0,0,0)'$);

$$\begin{aligned}\frac{\Delta\widehat{\mu}_0(x_0)}{\Delta x_{sex}} &= \widehat{\mu}_0(x_0 + \Delta x_{sex}) - \widehat{\mu}_0(x_0)\\ &= \exp(\widehat{\beta}_{01} + \widehat{\beta}_{02}) \times (\widehat{\pi}_{01} + \widehat{\pi}_{02}) - \widehat{\mu}_0(x_0)\end{aligned}$$

and the incremental effect of changing race from white to African-American (*i.e.*, $\Delta x_{race} = (0,0,1,0,0,0,0,0,0)'$);

$$\begin{aligned}\frac{\Delta\widehat{\mu}_0(x_0)}{\Delta x_{race}} &= \widehat{\mu}_0(x_0 + \Delta x_{race}) - \widehat{\mu}_0(x_0)\\ &= \exp(\widehat{\beta}_{01} + \widehat{\beta}_{03}) \times (\widehat{\pi}_{01} + \widehat{\pi}_{03}) - \widehat{\mu}_0(x_0).\end{aligned}$$

The estimates $\widehat{\beta}_{01}, \widehat{\beta}_{02}, \widehat{\beta}_{03}$ are the coefficients on the constant, male, and African-American in the main regression, and $\widehat{\pi}_{01}, \widehat{\pi}_{02}, \widehat{\pi}_{03}$ are the coefficients on the constant, male, and African-American in the regression to predict the error squared.

To compute the four estimates for the sample, similar calculations are required for each observation. However, in general the formulas will have more non-zero terms, including interactions. For example, for a person aged 50 (the age variable equals .1) changes in sex or race requires also changing the iteractions between those variables and age and age squared.

4.2 Results

Most of the demographic variables have a statistically significant main effect on logged annual health care expenditures (see the first column of Table 2). Log expenditures are lower for men. They are also lower African-Americans and Hispanics, which may reflect differences in access or insurance. The relationship between income and expenditures has the familiar U-shape; log expenditures are highest for those in the lowest income group, but increase with income for the upper four income groups. Log expenditures increase with age. Only a few of the interactions between age and other variables are statistically signficant (results not shown).

Although the heteroskedasticity regression, which predicts the squared residuals, has a low R^2, several variables are statistically significant (see the second column of Table 2). This will lead to differences in predicted expenditures on the raw scale from those calculated with a simple scalar Duan smearing factor. The predicted error variance is higher for both men and African-Americans. Therefore, predictions of mean expenditures for men will differ from women for two reasons. One is the difference between men and women found in the main expenditure regression, and the other is the difference between men and women due to heteroskedasticity. These differences go in the opposite direction. Likewise for race. Therefore, even though the coefficients on male and African-American are negative in the expenditure equation, predicted expenditures might end up being higher after controlling for heteroskedasticity.

Predicted expenditures for a 40-year-old non-Hispanic white female with a high school education in the lowest income group are $2,946 (see Table 3). This is higher than the overall mean largely because of being in the lowest income group. This is also nearly 50 percent higher than the estimate made with a single scalar Duan smearing factor. Accounting for the heteroskedasticity in the retransformation makes an enormous difference. Men are predicted to spend about $500 less than women, a smaller gender difference than when assuming homoskedasticity.

The incremental effect of race demonstrates the importance of controlling for heteroskedasticity in the retransformation. African-Americans spend more than whites. Although this result is not statistically significant, it shows that the overall effect can be positive even when a variable has a negative coefficient in the expenditure equation. The positive coefficient on race in the error variance equation dominates the negative coefficient on race in the expenditure equation.

When averaged over the sample, the predicted mean expenditures is the same as the overall sample mean (to within a dollar). This is somewhat closer than assuming homoskedasticity.

The incremental effects averaged over the sample assuming heteroskedasticity are the opposite sign than when assuming homoskedasticity. Using the semiparametric derivative estimator, men spend on average slightly more than women ($61), although the difference is not statistically significant. This is in stark contrast to using a scalar Duan smearing factor, which finds that men spend $467 less than women. The difference, again, is explained by how heteroskedasticity is handled during the retransformation. The incremental effect of gender ranged over the whole sample from –$1,257 to $826, using the semiparametric derivative estimator, but was always negative for the scalar Duan smearing factor. The difference between African-Americans and whites over the sample is $140, again opposite in sign to the effect found under homoskedasticity.

5 Conclusion

The log transformation is commonly used to deal with skewed data, and the conditional mean of the original dependent variable, marginal effects and interaction effects of explanatory variables on the original dependent variables are often the variables of interest in applied econometrics. In this paper, we present estimators for those variables for log transformed dependent variable models where the error term is possibly heteroskedastic and has an unknown distribution. We show that the estimators are consistent and asymptotically normally distributed. We provide consistent estimators for the asymptotic variances. The ratio of the estimate divided by the estimated standard error has a standard normal distribution and can be used for statistical inference. To illustrate the importance of calculating the correct interaction effect and standard errors, we consider a model to predict health care expenditures for adults using a nationally representative sample. As is commonly done, the dependent variable was transformed by taking the logarithm. We find that the incremental effects of explanatory variables are quite sensitive to assumptions about homoskedasticity.

6 Appendix

Proof of Theorem 1. Denote

$$\begin{aligned}
Q &= (\exp(u_1), \exp(u_2), ..., \exp(u_n))'; \\
R &= (D(x_1) - p^J(x_1)'\pi, ..., D(x_n) - p^J(x_n)'\pi)'; \\
E &= (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'; \\
\widetilde{D}(x) &= p^J(x)' \times (P'P)^{-1}P'Q.
\end{aligned}$$

For any vector $\lambda \leq m$

$$\begin{aligned}
\partial^\lambda \widehat{D}(x) - \partial^\lambda \widetilde{D}(x) = & \sum_{r=1}^{\dim(\beta_0)} \partial^\lambda p^J(x)' \times (P'P)^{-1}P' \frac{\partial Q}{\partial \beta_r} \times (\widehat{\beta}_r - \beta_{r0}) + \\
& \frac{1}{2} \sum_{r=1}^{\dim(\beta_0)} \sum_{s=1}^{\dim(\beta_0)} \partial^\lambda p^J(x)' \times (P'P)^{-1}P' \frac{\partial^2 \overline{Q}}{\partial \beta_r \partial \beta_s} \times \\
& (\widehat{\beta}_s - \beta_{s0})(\widehat{\beta}_r - \beta_{r0}) \\
= & A1 + A2,
\end{aligned}$$

where $\overline{\beta}$ is between β_0 and $\widehat{\beta}$, and

$$\overline{Q} = (\exp(\ln(y_1) - h(x_1, \overline{\beta})), ..., \exp(\ln(y_n) - h(x_n, \overline{\beta})).$$

By Theorem 1 of Newey (1997),

$$\partial^\lambda p^J(x)' \times (P'P)^{-1}P' \frac{\partial Q}{\partial \beta_r} \to \partial^\lambda \left[E\left\{ \frac{\partial \exp(\ln(\mathbf{y}) - h(\mathbf{x}, \beta_0))}{\partial \beta_r} | \mathbf{x} = x \right\} \right]$$

in probability. Assumption 2.1 and 3.5 imply $A1 = O_p(n^{-1/2})$.

Note that

$$\begin{aligned}
& \left| \partial^\lambda p^J(x)' \times (P'P)^{-1}P' \frac{\partial^2 \overline{Q}}{\partial \beta_r \partial \beta_s} \right|^2 \\
\leq & \partial^\lambda p^J(x)' \times (P'P)^{-1} \times \partial^\lambda p^J(x) \times \sum_{i=1}^{n} \left(\frac{\partial^2 \exp(\ln(y_n) - h(x_n, \overline{\beta}))}{\partial \beta_r \partial \beta_s} \right)^2 \\
= & \zeta_{|\lambda|}(J)^2 \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial^2 \exp(\ln(y_n) - h(x_n, \overline{\beta}))}{\partial \beta_r \partial \beta_s} \right)^2 = O_p(\zeta_{|\lambda|}(J)^2)
\end{aligned}$$

where the last equality follows from Assumption 3.5(ii). By Assumption 3.5(iii), $A2 = o_p(n^{-1/2})$. Hence, $\partial^\lambda \widehat{D}(x) - \partial^\lambda \widetilde{D}(x) = O_p(n^{-1/2})$.

Denote

$$V_{\lambda n}(x) = \partial^\lambda p^J(x)' \times (P'P)^{-1} \sum_{i=1}^{n} \sigma(x_i)^2 p^J(x_i) p^J(x_i)'(P'P)^{-1} \times \partial^\lambda p^J(x).$$

By Theorem 2 of Newey (1997), $V_{\lambda n}(x)^{-1/2}(\partial^\lambda \widetilde{D}(x) - \partial^\lambda D(x)) \to N(0,1)$ in distribution. Furthermore, it is straightforward to show

$$V_{\lambda n}(x) = O_p(\zeta_{|\lambda|}(J)^2/n).$$

Hence, $V_{\lambda n}(x)^{-1/2}(\partial^\lambda \widehat{D}(x) - \partial^\lambda D(x)) \to N(0,1)$ in distribution.

Note that$\widehat{\mu}_m(x) - \mu_m(x) =$

$$\begin{aligned}
&\sum_{\lambda \le m} \partial^{m-\lambda} \exp(h(x,\beta_o)) \times (\partial^\lambda \widehat{D}(x) - \partial^\lambda D(x)) + \\
&\sum_{\lambda \le m} (\partial^{m-\lambda} \exp(h(x,\widehat{\beta})) - \partial^{m-\lambda} \exp(h(x,\beta_0))) \times (\partial^\lambda \widehat{D}(x) - \partial^\lambda D(x)) \\
&\sum_{\lambda \le m} (\partial^{m-\lambda} \exp(h(x,\widehat{\beta})) - \partial^{m-\lambda} \exp(h(x,\beta_0))) \times \partial^\lambda D(x) \\
=& \sum_{\lambda \le m} \partial^{m-\lambda} \exp(h(x,\beta_0)) \times (\partial^\lambda \widehat{D}(x) - \partial^\lambda D(x)) + O_p(n^{-1/2}) \\
=& \sum_{\lambda \le m} \partial^{m-\lambda} \exp(h(x,\beta_0)) \times (\partial^\lambda \widetilde{D}(x) - \partial^\lambda D(x)) + O_p(n^{-1/2}).
\end{aligned}$$

Denote

$$\begin{aligned}
V_n(x) =& \partial^m \left[\exp(h(x,\beta_0)) \times p^J(x)\right]' (P'P)^{-1} \sum_{i=1}^n \sigma(x_i)^2 p^J(x_i) p^J(x_i)' \\
&\times (P'P)^{-1} \times \partial^m \left[\exp(h(x,\beta_0)) \times p^J(x)\right].
\end{aligned}$$

By Theorem 2 of Newey (1997),

$$V_n(x)^{-1/2} \sum_{\lambda \le m} \partial^{m-\lambda} \exp(h(x,\beta_0)) \times (\partial^\lambda \widetilde{D}(x) - \partial^\lambda D(x)) \to N(0,1)$$

in distribution. This proves $V_n(x)^{-1/2}(\widehat{\mu}_m(x) - \mu_m(x)) \to N(0,1)$ in distribution.

Denote $\widetilde{\varepsilon}_i = \exp(u_i) - p^J(x_i)'(P'P)^{-1}P'Q$ and

$$\begin{aligned}
\widetilde{V}_n(x) =& \partial^m \left[\exp(h(x,\beta_0)) \times p^J(x)\right]' (P'P)^{-1} \sum_{i=1}^n \widetilde{\varepsilon}_i^2 p^J(x_i) p^J(x_i)' (P'P)^{-1} \\
&\times \partial^m \left[\exp(h(x,\beta_0)) \times p^J(x)\right].
\end{aligned}$$

By Theorem 2 of Newey (1997), $\widetilde{V}_n(x)^{1/2} V_n(x)^{-1/2} \to 1$ in probability. Note that the difference between $\widetilde{V}_n(x)$ and $V_n(x)$ is that β_0 is replaced by a $\sqrt{n}$ consistent estimator $\widehat{\beta}$. It is easy to show that $\widehat{V}_n(x)^{1/2} V_n(x)^{-1/2} \to 1$ in probability. This completes the proof of the theorem.

Proof of Theorem 2. The proof is similar to the proof of Theorem 1 except that the $O_p(n^{-1/2})$ terms are no longer ignored. First, from the proof of Theorem 1, we immediately have

$$\partial^\lambda \widehat{D}(x) - \partial^\lambda \widetilde{D}(x) = \sum_{r=1}^{\dim(\beta_0)} \partial^\lambda p^J(x)'(P'P)^{-1}P'\frac{\partial Q}{\partial \beta_r}(\widehat{\beta}_r - \beta_{r0}) + o_p(n^{-1/2})$$

holds uniformly for all x and all $\lambda \leq m$. Denote

$$\widetilde{\mu}_m(x) = \partial^m[\exp(h(x,\beta_0)) \times \widetilde{D}(x)]$$

Write $\sum_{i=1}^n \widehat{\mu}_m(x_i) - \sum_{i=1}^n \widetilde{\mu}_m(x_i) =$

$$\begin{aligned}
&\sum_{\lambda\leq m} \partial^{m-\lambda}\exp(h(x_i,\beta_0)) \times (\partial^\lambda \widehat{D}(x_i) - \partial^\lambda \widetilde{D}(x_i)) + \\
&\sum_{\lambda\leq m} (\partial^{m-\lambda}\exp(h(x_i,\widehat{\beta})) - \partial^{m-\lambda}\exp(h(x_i,\beta_0)))(\partial^\lambda \widehat{D}(x_i) - \partial^\lambda \widetilde{D}(x_i)) \\
&\sum_{\lambda\leq m} (\partial^{m-\lambda}\exp(h(x_i,\widehat{\beta})) - \partial^{m-\lambda}\exp(h(x_i,\beta_0))) \times \partial^\lambda \widetilde{D}(x_i) \\
=&\sum_{i=1}^n \partial^m[\exp(h(x_i,\beta_0))p^J(x_i)'] \times (P'P)^{-1}P'\frac{\partial Q}{\partial \beta'} \times (\widehat{\beta} - \beta_0) + \\
&\sum_{i=1}^n (\widehat{\beta} - \beta_0)' \times \partial^m[\frac{\partial \exp(h(x_i,\beta_0))}{\partial \beta} \times p^J(x_i)'](P'P)^{-1}P'Q + o_p(n^{1/2}) \\
=&\sum_{i=1}^n \partial^m[\frac{\partial[\exp(h(x_i,\beta_o))E\{\exp(\ln(\mathbf{y}) - h(\mathbf{x},\beta_o))|\mathbf{x} = x_i\}]}{\partial \beta'}](\widehat{\beta} - \beta_0) \\
&+o_p(n^{1/2}) = o_p(n^{1/2})
\end{aligned}$$

where the second equality follows from substituting for $\partial^\lambda \widehat{D}(x_i) - \partial^\lambda \widetilde{D}(x_i)$ and $\partial^\lambda \widetilde{D}(x_i)$ and linearizing $\partial^{m-\lambda}\exp(h(x_i,\widehat{\beta}))$, the third equality follows from applying Theorem 1 of Newey to obtain

$$\begin{aligned}
&\partial^m[\exp(h(x,\beta_0)) \times p^J(x)'] \times (P'P)^{-1}P'\frac{\partial Q}{\partial \beta'} \\
&\rightarrow \partial^m[\exp(h(x,\beta_0))\frac{\partial E\{\exp(\ln(\mathbf{y}) - h(\mathbf{x},\beta_0))|\mathbf{x} = x\}}{\partial \beta}]; \\
&\partial^m[\frac{\partial \exp(h(x_i,\beta_0))}{\partial \beta} \times p^J(x_i)'] \times (P'P)^{-1}P'Q \\
&\rightarrow \partial^m[\frac{\partial \exp(h(x_i,\beta_0))}{\partial \beta}E\{\exp(\ln(\mathbf{y}) - h(\mathbf{x},\beta_0))|\mathbf{x} = x\}]
\end{aligned}$$

in probability uniformly over x, and the last equality follows from the fact

$$\partial^m[\frac{\partial[\exp(h(x_i,\beta_0))E\{\exp(\ln(\mathbf{y})-h(\mathbf{x},\beta_0))|\mathbf{x}=x_i\}]}{\partial\beta'}]=0.$$

Hence, $\sum_{i=1}^n(\widehat{\mu}_m(x_i)-\mu_m)$

$$\begin{aligned}
&=\sum_{i=1}^n(\widetilde{\mu}_m(x_i)-\mu_m)+o_p(n^{1/2})\\
&=\sum_{i=1}^n(\partial^m[\exp(h(x_i,\beta_0))p^J(x_i)']\times(P'P)^{-1}P'D-\mu_m)\\
&\quad+\sum_{i=1}^n\partial^m[\exp(h(x_i,\beta_0))p^J(x_i)']\times(P'P)^{-1}P'E+o_p(n^{1/2})\\
&=\sum_{i=1}^n(\partial^m[\exp(h(x_i,\beta_0))D(x_i)]-\mu_m)+\\
&\quad\sum_{i=1}^n\partial^m[\exp(h(x_i,\beta_0))p^J(x_i)']\times(P'P)^{-1}P'E+o_p(n^{1/2}).
\end{aligned}$$

The theorem now follows from applying a central limit theorem.

References

Abrevaya J. Computing marginal effects in the Box-Cox model. *Econometric Reviews* 2002; 21(3): 383–394.

Ai C, Norton EC. Standard errors for the retransformation problem with heteroscedasticity. *Journal of Health Economics* 2000; 19(5): 697–718.

Ai C, Norton EC. Interaction terms in logit and probit models. *Economics Letters* 2003; 80(1): 123–129.

Ai C, Sun. Asymptotic Normality of Functionals of Series estimators with dependent data. 2005; University of Florida working paper.

Andrews DWK. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* 1991; 59(2): 307–345.

Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 1998; 17(3): 283–295.

Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 1998; 17(3): 247–281.

Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of Health Economics* 2001; 20(4): 461–494.

Newey WK. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 1997; 79(1): 147–168.

Powell JL, Stock JH, Stoker TM. Semiparametric estimation of index coefficients. *Econometrica* 1989; 57(6): 1403–1430.

White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980; 48(4): 817–838.

Table 1
Summary statistics for the 2000 MEPS data.

Variable	**Min.**	**Mean**	**Max.**	**Std. Dev.**
Dependent Variables				
Expenditures	3	1,922	213,023	5,749
ln(expenditures)	1.10	6.25	12.27	1.56
Explanatory Variables				
Age	18	48.0	90	17.4
Male	0	.421	1	.494
African-American	0	.127	1	.333
Hispanic	0	.178	1	.383
Education > 12 years	0	.449	1	.497
Income group 1 (lowest)	0	.106	1	.308
Income group 2	0	.044	1	.204
Income group 3	0	.135	1	.342
Income group 4	0	.313	1	.464
Income group 5 (highest)	0	.402	1	.490

N=12,222

Table 2

Regression results for linear model to predict logged health care expenditures and the squared error.

	OLS			
Variable	**ln(*y*)**		$\exp(\hat{\varepsilon})$	
Constant	6.388	**	4.97	**
	(.070)		(.75)	
Male	–.400	**	1.16	**
	(.034)		(.33)	
African-American	–.184	**	1.54	*
	(.057)		(.62)	
Hispanic	–.372	**	.35	
	(.047)		(.38)	
Education $>$ 12 years	.095	**	–.13	
	(.036)		(.31)	
Income group 2	–.175		–2.60	**
	(.111)		(.92)	
Income group 3	–.172	*	–1.78	
	(.084)		(.93)	
Income group 4	–.108		–2.10	*
	(.074)		(.88)	
Income group 5 (highest)	–.0659		–2.69	**
	(.074)		(.89)	
$(Age-40)/100$	1.09	**	–2.26	
	(.41)		(3.06)	
$((Age-40)/100)^2$	.13		–4.02	
	(1.36)		(9.79)	
Additional interactions	16 included		16 included	
N	12,222		12,222	
R^2	.08		.008	

N=12,222. Robust standard errors are estimated using Huber-White robust standard errors. * Statistically significant at the 5 percent level; ** statistically significant at the 1 percent level.

Table 3
The three estimates of interest ($), both for a specific person and for the entire sample.

	Semiparametric Derivative Estimator	Scalar Duan Smearing Estimator
For the reference person		
Mean	2,946 **	2,078
	(443)	
Incremental effects		
Male compared to female	–510 **	–685
	(155)	
Black compared to white	266	–349
	(298)	
Averaged over sample		
Mean	1,923 **	1,998
	(312)	
Incremental effects		
Male compared to female	61	–467
	(237)	
Black compared to white	140	–513
	(353)	

The specific person is a 40-year old non-Hispanic white female, with high school education in the lowest income group. * Statistically significant at the 5 percent level; ** statistically significant at the 1 percent level.