

NBER WORKING PAPER SERIES

ESTIMATING THE EFFECTS OF COVARIATES
ON HEALTH EXPENDITURESDonna B. Gilleskie
Thomas A. MrozWorking Paper 7942
<http://www.nber.org/papers/w7942>NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2000

The authors thank Lara Bryant, Jae-Young Lim, and Bill Lu for valuable research assistance on this project. We are grateful to Will Manning and John Mullahy for providing data and programs used in our preliminary research. We also appreciate comments from participants at the Triangle Econometrics Conference and the Murray S. Johnson Memorial Conference at the University of Texas - Austin. The views expressed are those of the authors and not necessarily those of the National Bureau of Economic Research.

© 2000 by Donna B. Gilleskie and Thomas A. Mroz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating the Effects of Covariates on Health Expenditures
Donna B. Gilleskie and Thomas A. Mroz
NBER Working Paper No. 7942
October 2000
JEL No. I1

ABSTRACT

This paper addresses estimation of an outcome characterized by mass at zero, significant skewness, and heteroscedasticity. Unlike other approaches suggested recently that require retransformations or arbitrary assumptions about error distributions, our estimation strategy uses sequences of conditional probability functions, similar to those used in discrete time hazard rate analyses, to construct a discrete approximation to the density function of the outcome of interest conditional on exogenous explanatory variables. Once the conditional density function has been constructed, we can examine expectations of arbitrary functions of the outcome of interest and evaluate how these expectations vary with observed exogenous covariates. This removes a researcher's reliance on strong and often untested maintained assumptions. We demonstrate the features and precision of the conditional density estimation method through Monte Carlo experiments and an application to health expenditures using the RAND Health Insurance Experiment data. Overall, we find that the approximate conditional density estimator that we propose provides accurate and precise estimates of derivatives of expected outcomes for a wide range of types of explanatory variables. We find that two-part smearing models often used by health economists do not perform well. Our results, both in Monte Carlo experiments and in our real application, also indicate that simple one-part OLS models of level health expenditures can provide more accurate estimates than commonly used two-part models with smearing, provided one uses enough expansion terms in the one-part model to fit the data well.

Donna B. Gilleskie
Department of Economics
University of North Carolina at Chapel Hill
Chapel Hill, NC 25799-3305
and NBER
donna_gilleskie@unc.edu

Thomas A. Mroz
Department of Economics
University of North Carolina at Chapel Hill
Chapel Hill, NC 25799-3305
tom_mroz@unc.edu

Health economists and other empirical researchers often debate the advantages and disadvantages of various functional forms used in regression analyses. Researchers frequently use, for example, logarithmic transformations of the dependent variable when the variable exhibits significant skewness. Often it is the level of the dependent variable that is of interest in the analysis and retransformation of the estimated predictions is necessary. Many health outcomes are also characterized by a significant point mass at zero. It has become common among health economists to use a two-part model even though the outcome of interest is typically the unconditional outcome that includes the zero values. Examples of health outcomes exhibiting skewness and point mass at zero include health care expenditures, number of doctor visits, and duration of hospital stay. Several recent papers have addressed these modeling issues from a variety of points of view (e.g., Mullahy, 1998; Manning, 1998; Angrist, 2000; Manning and Mullahy, 2000).

Retransformation and two-part modeling require that the researcher make distributional assumptions. For example, in order to obtain predicted values of the unconditional level outcomes from a logarithmic transformed dependent variable, a researcher must retransform the predicted logged dependent variable. This retransformation requires assumptions about the dependence of the error term distribution on observable covariates. One also must specify assumptions about the relationship between the errors of the marginal and conditional outcomes represented in two-part models. To our knowledge there is no general empirical approach that simultaneously “makes irrelevant” the decision to transform or not and the choice of two-part versus one-part modeling while also allowing for possibly complex interactions of explanatory variables on the outcome of interest. Our goal in this paper is to provide such an approach.

In this paper we describe a relatively simple estimation approach that “solves” the transformation problem while incorporating explicitly the potential confounding effects of two-part models. Our estimation strategy uses sequences of conditional probability functions, similar to those used in discrete time hazard rate analyses, to construct a discrete approximation to the density function of the outcome of interest conditional on exogenous explanatory variables. Once we have constructed the conditional density function, it is straightforward to examine expectations of arbitrary functions of the outcome of interest and to evaluate how these expectations vary with observed exogenous covariates. Our implementations of the approach use flexible functional forms when defining the

sequences of conditional probabilities. This means that we have flexible representations of the conditional density functions, and consequently flexible representations for regression functions such as the expected value of the outcome conditional on exogenous covariates.

In our formulation we construct the sequences of conditional probabilities using standard binary outcome models. In our Monte Carlo experiments and health economics application we use a logit probability model because of its simplicity, but any binary outcome model could be used. We allow the arguments of these conditional probability functions to be loosely-specified polynomial functions of covariates. These models can then be estimated using standard computer packages such as SAS and Stata. In fact, we used Stata for all of the Monte Carlo experiments and real application estimates reported in this paper. It is simple for researchers to implement this approach in practice.

The approach we use naturally admits variations in covariate effects over particular ranges of the variables. It might be the case, for example, that particular variables have no impacts on an outcome of interest until the outcome exceeds some pre-specified cutoff level. Characteristics of one's health insurance contract (e.g., the deductible, the coinsurance rate after exceeding the deductible, and the maximum out-of-pocket expenditure) are obvious examples in the health insurance literature where the economic impacts of these covariates vary over the range of the dependent variable. One could, in principal, model directly how such features effect the budget constraint and then solve for the correct demand function to use in a least squares estimation. The resulting functional form for the regression model would almost always be quite nonlinear, and it would depend crucially upon arbitrary distributional and functional form assumptions. The approach we examine here incorporates such effects with almost no modification.

While the presence of unobserved individual heterogeneity could limit a researcher's ability to translate such economic restrictions directly to restrictions in a statistical model, these types of functional restrictions can provide key information to help identify nonparametrically both the behavioral relationship and the heterogeneity distribution (Mroz and Weir, 1990; Hahn, Todd, and Van der Klaauw, 2000); this is a clear case where a well-specified economic theory can help

provide identification without the imposition of arbitrary functional forms.¹ The discrete conditional density estimation approach we use can be applied directly to a wide range of situations where researchers have relied upon restrictive functional forms and distributional assumptions. It is straightforward to apply the approaches we describe to: discrete time hazard rate models (Allison, 1984); ordered data models, where researchers almost always assume either an ordered logit or an ordered probit model with covariate-invariant cutoff points (Maddala, 1983); count data models, where most researchers use a Poisson model or simple, parametrically restrictive modifications of the Poisson such as the negative binomial model or the zero-inflated Poisson model (Cameron and Trivedi, 1998); and multinomial models with mutually exclusive outcomes, where researchers have relied almost exclusively on multivariate probit or logit models (Maddala, 1983). In addition to this wide application of the conditional density estimation technique, the use of a maximum likelihood framework provides the foundation for more complex modeling of selection, endogeneity, and unobserved heterogeneity.

Using Monte Carlo experiments and our health economics application, we examine the performance of these discrete, conditional density approximations when the outcome of interest is a positively-skewed continuous variable with mass at zero. We find that the discrete approximation works quite well with these outcomes. This suggests that the approach may be useful when a researcher is interested in estimating the expected impact of exogenous covariates on particular functions of the outcome variable, regardless of whether the outcomes of interest are discrete, continuous, or mixed.

¹The approach we use yields estimates of how the entire distribution of health expenditures would change in response to variations in exogenous characteristics. Conceptually it is straightforward to add unobserved heterogeneity to the specifications we use along the lines suggested by Heckman and Singer (1984), Mroz (1999), and Mroz and Guilkey (1992). This would, in principal, allow one to examine how covariate effects vary within ranges of the outcome variable. For example, one might be interested in how increasing co-insurance rates would effect health care expenses for those spending more than \$500 per year. To do this well, however, the researcher would need to hold constant the distribution of the “heterogeneity” applicable for this range of the outcome and must specify precisely how the heterogeneity distribution would be identified. This would require strong assumptions or additional information. In general one would need multiple observations per agent in order to identify the heterogeneity distribution separately from the distribution of the outcome conditional on exogenous covariates and the heterogeneity (Heckman and Honore, 1990). Such extensions are beyond the scope of this paper. Hence, the approach we detail in this paper does not address the question of how one can obtain interesting and consistent estimates of how covariates influence outcomes over *intervals* of the support of the dependent variable conditional on the random variable falling within the interval.

1 Description of Estimation Procedure

We begin by describing in detail the estimation technique. The approach requires determination of intervals within the support of the dependent variable, approximation of the conditional expected value, and implementation of the empirical approximation in practice. We conclude this section by explaining how to calculate statistical derivations of interest.

The conditional density estimation approach we propose in this paper closely resembles the approaches used by Efron (1988) and Donald, Green, and Paarsch (2000). The Efron model, like that proposed here, approximates the distribution of a continuous outcome by a discrete distribution function. He proposes that one estimate the statistical approximation to the distribution function by a sequence of logit hazard rates, which is precisely the modeling approach we adopt. The main difference between our approach and Efron's is that we examine distributions conditional on observed covariates, while Efron models only the unconditional distribution. An important result from Efron's analysis concerns the fact that the efficiency loss due to discretization can be quite small. For a true underlying continuous Poisson process, for example, he finds that information loss quickly goes to zero as the number of discrete intervals gets large.

Donald, Green, and Paarsch (2000) propose an estimator quite similar to the one we describe below. The primary difference between their estimator and ours is that they use a continuous distribution with discrete structural shifts to approximate the underlying distribution, as in Meyer's (1990) approach to estimating hazard models. To allow for effects of covariates that vary over the support of the outcome, they rely on separate, discontinuous "proportional hazard" effects for various ranges of the the outcome variable. Our approach, on the other hand, allows the impacts of covariates to vary smoothly over the entire range of the support of the outcome of interest, except possibly at particular points or regions where the researcher has an a priori notion that behavior might be discontinuous. The complexity of the estimated distribution and its dependence upon covariates is pre-specified in the Donald, Green, and Paarsch approach, while the estimator we propose allows the data to determine the number of terms and breakpoints used to approximate the conditional distribution function. Eastwood and Gallant (1991), for example, find that data dependent rules for choosing the number of terms in an expansion often yields less bias than using fixed rules. A final advantage of our approach is that estimating the smoothed conditional density

function only requires one to use simple logit models. Donald, Green, and Paarsch’s estimator can be difficult to use because it requires that the researcher provide reasonable starting values for the parameter estimates. All of the differences we mention are quite minor, and the choice of which of these two approaches would be better depends upon whether one believes that the distribution of the outcome of interest has many interesting and important, discontinuous segments.

1.1 Discretizing the Support of the Dependent Variable

Figure 1 displays an arbitrary distribution function for a random variable Y conditional on a set of covariates x with density $f(y|x)$.² Suppose we break the range of the dependent variable into K intervals, where the k th interval is defined by $[y_{k-1}, y_k)$, for $y_{k-1} \leq y_k, y_0 = -\infty$ and $y_K = \infty$.

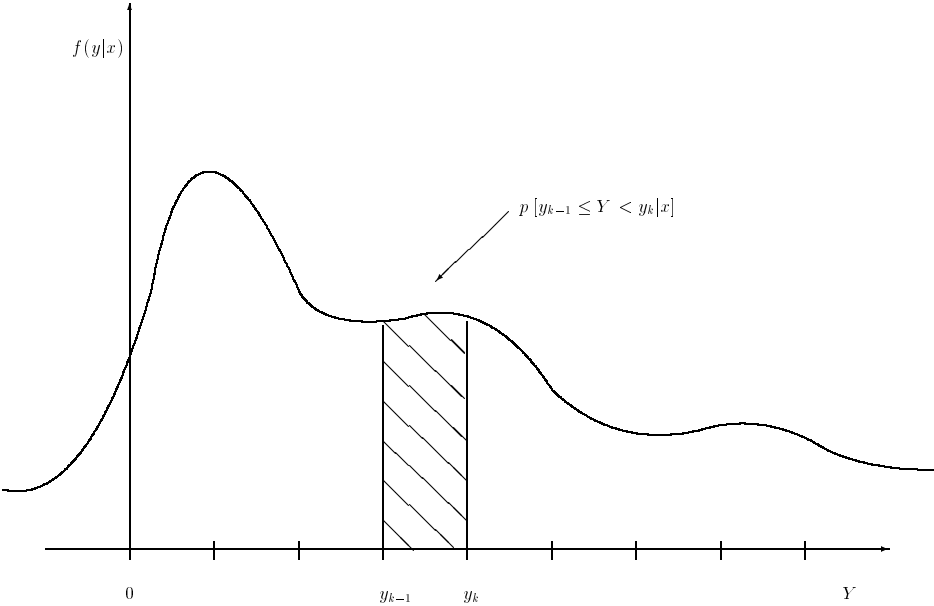


Figure 1: Arbitrary Distribution of Y

The probability that the random variable Y falls in the first interval is given by:

$$p[y_0 \leq Y < y_1 | x, Y \geq y_0] = \int_{y_0}^{y_1} f(y|x) dy = \lambda(1, x) \tag{1}$$

²If the random variable Y is discrete, then each partition may include one or more points in the support of Y . For most of the following discussion, we use notation for a continuous distribution for the outcome, but little would need to be changed to accommodate discrete or mixed outcomes.

where $\lambda(1, x)$ is that function of x that gives this probability. Note that $\lambda(1, x)$ is implicitly a function of the choice of partition for the support of Y , but this dependence on the partition is captured entirely through the values y_0 and y_1 . The probability that the random variable Y falls in the k th interval is given by

$$p[y_{k-1} \leq Y < y_k | x] = \int_{y_{k-1}}^{y_k} f(y|x) dy \quad (2)$$

and the conditional probability that the random variable falls in the k th interval given that it did not fall in one of the first $(k - 1)$ intervals is

$$\begin{aligned} \lambda(k, x) &= p[y_{k-1} \leq Y < y_k | x, Y \geq y_{k-1}] \\ &= \frac{\int_{y_{k-1}}^{y_k} f(y|x) dy}{1 - \int_{y_0}^{y_{k-1}} f(y|x) dy} \end{aligned} \quad (3)$$

The functions $\lambda(\cdot, \cdot)$ define the “discrete time hazard” function representation for the chosen partition of the support of Y . By the properties of hazard functions, the probability that the random variable Y falls in the k th interval is given by

$$p[y_{k-1} \leq Y < y_k | x] = \lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)] . \quad (4)$$

The hazard rate decomposition implies, by definition, a conditional independence between the events $\{y_{k-1} \leq Y < y_k | x, Y \geq y_{k-1}\}$ and $\{y_{j-1} \leq Y < y_j | x, Y \geq y_{j-1}\} \forall j \neq k$. If a researcher imposes a functional form for $f(y|x)$ or $\lambda(k, x)$, however, the potential for unobserved heterogeneity exists and the conditional independence can break down (Heckman and Singer, 1984). This happens because the same unobservable, say μ , influences all of the events represented in Equation (4). This, however, is only an issue for the hazard function that is constructed by conditioning on the unobserved heterogeneity. Suppose $\lambda_g(k, x|\mu)$ is the discrete hazard rate under density $g(y|x, \mu)$ that is conditional on the unobserved heterogeneity and that $Q(\mu)$ is the cumulative distribution of the unobserved heterogeneity. Conditional on the heterogeneity the independence properties hold. The conditional distribution function is given by

$$p[y_{k-1} \leq Y < y_k | x, \mu] = \lambda_g(k, x|\mu) \prod_{j=1}^{k-1} [1 - \lambda_g(j, x|\mu)] \quad (5)$$

and the unconditional distribution function is

$$p[y_{k-1} \leq Y < y_k | x] = \int \lambda_g(k, x|\mu) \prod_{j=1}^{k-1} [1 - \lambda_g(j, x|\mu)] dQ(\mu) . \quad (6)$$

This distribution in Equation (6) must, by definition, have an “independent” hazard rate decomposition like that described above in Equation (4).

The primary implication of unobserved individual heterogeneity is that one cannot interpret the distribution function $f(y|x)$ and the “hazard” function $\lambda(k, x)$ as holding the heterogeneity fixed, as would be the case if one were able to estimate the “hazard” functions $\lambda_g(k, x|\mu)$. Thus one cannot give a clear structural interpretation to effects of variations in x on the outcome y in sub-intervals of its support; to do this would require knowledge of the distribution of the heterogeneity conditional on being in the specified sub-intervals. The point here is similar to the interpretation of estimates of “duration dependence” of the hazard rate in waiting time models with unobserved heterogeneity. Unless one holds constant the unobserved heterogeneity in a waiting time model, the time shape of the hazard function does not have a structural interpretation; the time shape depends upon changes in the distribution of the unobserved heterogeneity over the support of the dependent variable. The decomposition we use, however, does permit one to make precise structural statements about the impact of the covariates x on the expectation of any function of y that does not condition on particular ranges for the random variable. It is straightforward, for example, to calculate how the mean and the variance of the random variable Y , or the mean and variance of functions of the random variable Y , vary with changes in the exogeneous covariates x .

1.2 Approximating Moments of the Distribution

In this discussion we focus on first conditional moments, but the discussion could easily be modified for any conditional moments. The true expectation of a function $h(\cdot)$ of a random variable Y , given x , is

$$E(h(Y)|x) = \int_{-\infty}^{\infty} h(y)f(y|x)dy \quad (7)$$

where $h(\cdot)$ is any smooth and continuous function of Y . For a partition of the support of Y with K intervals, we approximate the expectation of a function $h(Y)$ conditional on covariates x by

$$\tilde{E}(h(Y)|x) = \sum_{k=1}^K h^*(k|K)\lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)] \quad (8)$$

where each $h^*(k|K)$ is an approximation to $h(y)$ in the k th interval (corresponding to interval $[y_{k-1}, y_k)$). The approximation we use treats the $h^*(k|K)$ ’s as fixed for all values of x within the

k th interval. For smooth and continuous functions $h(Y)$, this approximation to the expectation will converge to the true expectation as the widths of the intervals shrink towards zero. The empirical question, then, is how well this approximation can work in practice.

1.3 Implementing the Approximation Empirically

Before one can apply this approximation in practice, several decisions need to be made regarding implementation. The five decisions at the discretion of the researcher are to:

1. choose the number of intervals to use (K),
2. specify boundaries of the intervals (i.e., the values of y_1, y_2, \dots, y_{K-1}),
3. pick a set of constants $h^*(1|K), h^*(2|K), \dots, h^*(K|K)$ to use in the approximation to the integral,
4. decide how to approximate the conditional density functions, and
5. calculate derivatives of the expectation of the function of the outcome of interest.

For each of the five decisions we use empirical analogues to guide our choices.

Choosing Widths of the Intervals

First, consider the choice of the boundaries of the intervals. Suppose that one has already chosen to have K intervals. For the most part, in our Monte Carlo experiments and empirical example we choose as boundary points values that place an equal number of observations in each interval (i.e., $1/K$ th of the sample of Y falls within each interval). If we chose 10 intervals, for example, then y_1 is the tenth percentile of the observed outcome Y , y_2 is the twentieth percentile of the observed outcome Y , and y_{K-1} is the ninetieth percentile of Y . Boundaries chosen in this fashion are equivalent for monotonic transformations of the random variable Y . In those instances where there are significant point masses in the observed distribution of Y , one can allow each mass point to define a single interval. In the work we report here we examine annual health care expenses. We allow zero expenditures to be a single interval and choose boundary points such that there are an equal number of observed positive expenditures in each of the remaining intervals.³

³When there are minor point masses, say heaping at \$100 in health expenditures for example, we typically do not allow for additional intervals that contain only one point of the distribution unless the number of

Selecting the Number of Intervals

Next, consider the question of how many intervals to use for the discrete distribution. In practice we address this question empirically. We choose as the number of points of support that number which maximizes the goodness of fit of the model given the four additional decisions being discussed. To do this, consider having already chosen K' intervals with each interval containing N/K' observations. Estimate the discrete distribution function and construct the value of the log likelihood of this choice of K' intervals as

$$\mathcal{L}(Y|K') = \sum_{i=1}^N \ln \left[\prod_{k=1}^{K'} \left\{ \lambda(k, x(i)) \prod_{j=1}^{k-1} [1 - \lambda(j, x(i))] \right\}^{1_{\{y(i) \in [y_{k-1}, y_k)\}}} \right] \quad (9)$$

Next, consider taking each of the K' intervals (each with an equal number of observed Y 's per interval) and breaking each one into R sub-intervals with an equal number of observed values of Y in each interval. Each of the new intervals contains $N/(K'R)$ observations. Let the estimated probability of an observation being in the k th of the original K' intervals be $\lambda(k, x(i), K')$. Now, allocate this probability equally among each of the R sub-intervals that comprise this k th interval. Under this allocation rule for distributing the estimated probability, the probability that an observation falls in one of these R sub-intervals is given by

$$\lambda^*(r, K', x(i), R \times K') = \frac{\lambda(k, x(i), K')}{R} \quad (10)$$

This is the adjusted probability that an observation falls in the r th sub-interval of the original k th interval. All we have done is distributed these probabilities equally over the finer partition.

Since $(1/R)$ of the observations that originally fell in the k th interval fall in each of these R new, smaller intervals, it is straightforward to calculate the sample log likelihood when one uses these equal allocation probabilities. This adjusted log likelihood is given by

$$\mathcal{L}^*(Y|K', R \times K') = \mathcal{L}(Y|K') - N \ln(R) \quad (11)$$

where N is the number of observations *not* at places of significant point mass. For each observation, the log-likelihood value is reduced by $\ln(R)$. This reflects the fact that it is “harder” to predict observations at that one value is large relative to the number of intervals. These heaped values, however, can make it impossible to distribute the mass equally among the remaining intervals. Do note that if such minor mass points are substantively interesting and important, for example at 26 weeks duration in an unemployment spell or 2000 hours in an annual hours worked distribution, one can and should allow these mass points to be separate intervals.

which interval an observation falls into when one allows for more intervals. This adjusted likelihood function value reflects changes in the log likelihood value associated with expanding the number of intervals and not re-estimating the model.

Now, consider estimating a model containing $R \times K'$ intervals. A reasonable criterion to use to decide whether one should use the estimates from the model with K' intervals or those from the model with $R \times K'$ intervals is whether or not the new log likelihood function value, $\mathcal{L}(Y|R \times K')$, exceeds the original log likelihood function value for K' intervals adjusted for the finer partition, $\mathcal{L}^*(Y|K', R \times K')$. If by choosing a finer partition we fit the data worse than we did by using a model with fewer intervals (and an equal allocation of probability within each interval), then we would choose the model with a fewer number of intervals. The usefulness of this criterion comes from the fact that our estimators of $\lambda(k, x)$ are smooth over regions of the support of y . If one were more nonparametric and allowed for completely separate functions $\lambda(k, x)$ across intervals, then following this criterion would always select the model with the most partitions.

To implement this in practice, consider comparing each choice of R to a model with only one partition.⁴ The adjusted log likelihood function value with one partition is

$$\mathcal{L}^{**}(Y|R, 1) = \mathcal{L}(Y|R) + N \ln(R) \tag{12}$$

This tells us how much better a model with R intervals fits the data than a model with only one interval and an equal allocation of probabilities across the R intervals. We choose as the number of intervals, K , that value of R which maximizes the above adjusted log likelihood function value. In our Monte Carlo experiments and real example, we examine sample sizes from 1,000 to 5,000. We found that one did not need to examine more than 50 partitions; typically 10 to 20 intervals were sufficient.

Evaluating Expectations of the Outcome

Next, consider how one should choose the evaluation point within each interval for the desired function of the outcome variable (i.e., the $h^*(k|K)$'s). For most applications, each interval will

⁴One partition, from $-\infty$ to $+\infty$, broken into R sub-intervals with equal probability is just a simple multinomial density with all intervals having the same probability. So, the approach we use chooses the number of intervals with the maximum gain in the likelihood function resulting from including covariates as determinants of the interval probabilities (after adjusting the likelihood function value for the number of intervals).

contain many observed values of the outcomes. In the Monte Carlo experiments and example reported here, we evaluate the function $h(\cdot)$ at each observed value within the interval and take a simple arithmetic average:

$$h^*(k|K) = \frac{\sum_{y \in [y_{k-1}, y_k]} h(y)}{\sum_{y \in [y_{k-1}, y_k]} 1} \quad (13)$$

It is important to note that this might be an extremely important assumption in small samples for the approach that we use. In particular, consider the derivative of the expected value of the function $h(Y)$

$$\frac{\partial \tilde{E}(h(Y)|x)}{\partial x} = \sum_{k=1}^K h^*(k|K) \frac{\partial \{\lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)]\}}{\partial x} \quad (14)$$

given our assumption that $h^*(k|K)$ does not vary with x .

A better approximation might be to recognize that the average of the function $h(Y)$ within the k th interval would vary with changes in the covariates; this would happen because the distribution of y *within* the interval would vary with changes in x . The actual derivative of the conditional expected value would be

$$\begin{aligned} \frac{\partial E(h(Y)|x)}{\partial x} &= \sum_{k=1}^K h^{**}(k|K, x) \frac{\partial \{\lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)]\}}{\partial x} \\ &\quad + \sum_{k=1}^K \frac{\partial h^{**}(k|K, x)}{\partial x} \lambda(k, x) \prod_{j=1}^{k-1} [1 - \lambda(j, x)] \end{aligned} \quad (15)$$

where $h^{**}(k|K, x)$ is the average value of Y in the k th interval as a function of the explanatory variables.

A comparison of Equations (14) and (15) reveals that the approach we use ignores the second term in the formulation of the derivative. Note that one could run a regression of the outcomes in each interval on the covariates and use that regression to calculate the average of the derivative within each interval. For smooth and continuous functions $h(\cdot)$ with finite expected value, the second term in the above derivative expression could be asymptotically negligible compared to the first term in the sum (i.e., as the number of partitions grows large, the limit of the ratio of the first term in the sum to the total sum is one), but we have not yet shown this for general cases. This is clearly a topic that deserves additional attention. But note that the Monte Carlo experiments presented here do indicate that ignoring the second term in this sum appears to introduce little bias in the estimated derivatives.

Estimating the Conditional Multinomial Probabilities

Our final specification of the approximation to the expected conditional value concerns the approximation to the density function $p[y_{k-1} \leq Y < y_k | x]$. We approximate this density using the hazard rate decomposition discussed above. In particular, we specify a logit function for the probability of an outcome falling in the k th interval given that it was not in one of the lower $(k-1)$ intervals. In practice, one could estimate a separate logit model for each “hazard” of falling within each interval.⁵ This, however, would introduce a large number of parameters in most realistic-sized problems. Instead, we estimate one logit probability using polynomials in functions of the covariates and the interval number.

Suppose one chooses to use K partitions. By using partitions containing an equal number of observations, the unconditional probability (not conditional on the x 's) of an observation being in the k th interval given that it was not in one of the lower $(k-1)$ intervals is $\frac{1}{K-(k-1)}$. Let $\alpha_k = -\ln(K-k)$ for $k < K$. Then,

$$\text{logit}(\alpha_k) = \frac{e^{\alpha_k}}{1 + e^{\alpha_k}} = \frac{1}{K - (k - 1)}. \quad (16)$$

If one estimated a single logit function for all of the $(K-1)$ hazard rates with α_k as the only covariate in the logit function, then the fit of the unconditional discrete distribution function would be “perfect;” the predicted probabilities of the outcome falling in each of the intervals is identical to what one would obtain by estimating a logit hazard model with dummy variables for each interval, or a separate logit model for each event of an observation falling into each of the intervals. By using α_k as the only covariate in a logit formulation of the hazard function, we are guaranteed to fit exactly the discretized marginal distribution of y given the choice of K intervals.

In our estimation of the “hazard” functions that condition on covariates we follow a similar strategy and estimate a single logit model for all hazard rates. We include polynomials in α_k in addition to polynomials in the observed covariates as linear arguments to the logit function. We also include interactions among the covariate polynomials and the α_k 's. This provides a flexible

⁵The fact that each outcome can contribute an “observation” to more than one “logit” and that outcomes vary in the number of “observations” they contribute is a non-issue. One can always apply a “hazard” decomposition to any conditional or unconditional distribution function and end up with exactly this type of formulation. Every maximum likelihood estimation problem with continuous outcomes, then, can be thought of as one where each observed outcome can contribute an “observation” to more than one interval.

way to smooth the hazard rate decomposition of the conditional density function. When we have an outcome with substantively and quantitatively interesting point masses, as we do at zero expenditures in our analysis of health care expenses, we estimate a separate logit model with polynomials in the covariates and without the α_k polynomials.⁶

We use downwards testing to guide the selection of the degree of the polynomial to use in the logit model of the hazard functions. In particular, the most complex model we consider includes a fourth degree polynomial in α_k and all fourth order terms in the covariates, their interactions, and the covariate polynomials (including their interactions) interacted with the α_k polynomial. We then reduce the order by one for the covariate polynomials, their interactions, and their interactions with the shape parameter α_k , but retain fourth order polynomials in α_k . We test whether the additional coefficients as a group are significant with a Wald test at the five percent level.⁷ If the higher order terms are significant, then we keep the unrestricted specification. If not significant, we reduce the polynomials and interactions by an additional order. We then use another Wald test at the five percent level to test whether the more detailed specification provides a better fit than the restricted model. If we do not find a significant improvement with the higher order terms, we reduce the specification further by eliminating all third order polynomials, third order interactions of the covariates with themselves and with α_k , and fourth order polynomials in α_k . Again, we test at the five percent level whether the additional terms improve the goodness of fit. The simplest model we consider includes α_k , the square of α_k , and first order terms in the covariates. In our Monte Carlo work we find that this procedure almost always selected the most complex set of interactions in the logit hazard function.⁸

⁶We found that polynomials in the logarithms of the covariates seemed to provide somewhat more stable estimates than polynomials in the levels of the variables. Except for the dummy variables we use as explanatory variables, we add a constant to each covariate to ensure that the minimum value is positive and then construct polynomials in the logarithms of the normalized covariates.

⁷While the hazard formulation implies theoretically that there is independence between the events defined by the hazard rates, we recognize that the approximation we use may not be perfect. We use standard error estimators (Eicher-Huber-White) that allow for correlation among the conditional events for each observed outcome in the Wald tests.

⁸Our estimation programs automatically drop linearly dependent variables and we take this into account when testing. See Appendix Tables A3, A4, and A8 for a precise description of the polynomial terms we include in the conditional density estimation for the Monte Carlo experiments and the Rand Health Insurance Experiment data.

1.4 Calculating Derivatives

When calculating derivatives of the expectation of the function of the outcome of interest, we hold constant the approximation to the function within each interval (the $h^*(k|K)$'s) as mentioned above, differentiate the approximating density function for each interval with respect to a set of covariates, and then sum the products of the $h^*(k|K)$'s and the derivatives of the density function, as described in Equation (14). To do this in practice we evaluate the conditional expected value at various values of the explanatory variables and calculate the “arc” derivatives. Suppose we are interested in calculating the derivatives of the expectation of health care expenses with respect to the level of deductible in the health insurance plan, which we report in Table 1. One way of calculating the average derivative (Average in the tables), is first to calculate the expected value of expenditures using the estimated distribution function for all observations. Next, deviate each observation’s deductible level and recalculate the expected values for all observations. Given the polynomial approximations we use, it is advisable to ensure that the deductible levels are set to values that are actually observed in the data; each observation could have a different deviation. Then, take the difference between these two expectations observation by observation and divide the difference by the finite difference chosen for the covariate for each observation to obtain the “derivative” for each observation. The population derivative is the average of these derivatives across observations. This most closely corresponds to the average impact in the population of a one unit change in a covariate⁹ In other instances, we evaluate the expected value of the function of the outcome of interest for each observation with one particular covariate set to the same value for all observations. We then deviate the covariate for each observation, recalculate the expectation, take differences, divide through by the change in the covariate, and average across observations. This corresponds to the average derivative at a particular value of the chosen covariate, when all of the other characteristics of individuals follow the joint (marginal) distribution observed in the data. We also experimented with aggregating these values to construct different measures of the overall “average” derivative of a covariate. These are reported as Min Variance, Equal Weights, Weight 1, and Weight 2 in Table 1; these are defined in detail in Appendix Table A2. Clearly, one

⁹In our Monte Carlo experiments we randomly draw sets of explanatory variables from sets of covariates for 1219 observations observed in the NMES data set. Rather than calculating derivatives for the estimation sample, we calculate the derivatives for these 1219 sets of explanatory variables.

could chose a wide variety of ways to calculate derivatives of the expected value with respect to particular covariates and how these derivatives vary with values of the other covariates.

1.5 Summary of the Estimation Procedure

First, split the sample into K intervals. Estimate the approximation to the conditional density function using all of the possible polynomial specifications, and test downwards from the most complex model to simpler models. Select that model that first indicates that the additional coefficients are statistically significant. Repeat this procedure for a wide range of K values for the number of intervals. Next select the value of K that maximizes the likelihood function value adjusted to reflect the differences in the number of intervals. This procedure yields the estimator of the density function. We then calculate the derivatives using the finite difference procedure described above. When we examine the data from the Rand Health Insurance Experiment, we bootstrap this entire procedure to obtain estimators of the standard errors. This means that the standard errors we report control for all of the pre-testing we do with respect to the degree of the polynomial and the selection of the number of intervals.

2 Specification of the Monte Carlo Experiments

In the Monte Carlo experiments we focus on five specifications of the data generating process. For each model we use data from the National Medical Expenditure Survey (NMES) from 1987 to set coefficients and to define the joint distribution of the explanatory variables. For the most part, we ran simple regressions of health care expenditures on age, household income, coinsurance rate, deductible amount, and demographic controls to define the parameters determining the continuous outcome in each data generating process (DGP). We used probit models, in which the dependent variable is whether an individual had any medical expenditures in the NMES data, to calibrate those DGPs that use a “two-part” approach. We used the same explanatory variables in the probit equation as in the conditional expenditures equation along with an indicator of whether an individual had a regular health care provider. For each replication within each DGP we draw samples of the explanatory variables, with replacement, from a set of 1219 individuals in the NMES

data set. We examined sample sizes of 1000, 3000, and 5000, but we focus on sample size 5000 in our discussion.

The first DGP we examine is a single equation OLS model where the outcome is a simple linear function of the covariates. This model has normal disturbances, and OLS is the best unbiased estimator (efficient). We use this DGP to examine whether the complex estimation model we use can replicate well a simple process and to uncover how much efficiency might be lost because we allow for extremely flexible functional forms and distributions.

The second DGP we examine is a two-part model. In the first part a probit model with normal errors, calibrated with the NMES data, determines whether or not an individual has positive expenditures. If the outcome generated by the DGP indicates that the individual had positive expenditures, then a simple linear regression function with an independent normal error determines the natural logarithm of the expenditure for the individual. With this DGP and the following three DGPs, we use a “two-part” version of the conditional density estimation approach where we specify a simple, separate logit for the first stage (any expenditures) and a sequence of related logits to define the conditional density function for positive expenditures.

The third DGP is quite similar to the second DGP. It only differs by allowing the error variance in the second part to be proportional to the expected value of the logarithm of the dependent variable. The fourth DGP adds additional heterogeneity and heteroscedasticity. In particular, it specifies a random parameter model in the second stage regression function for the logarithmic expenditures. The final DGP we consider is a mixture model. Here, there is a distribution of individuals, and each type has a different propensity to have positive expenditures and a different expected level of logged expenditures given that they have some positive expenditures. The error terms in the two parts are related, so a standard two-part model would suffer from selection bias.

We compare the conditional density approximation estimators (CDE) of the derivatives of expected expenditures to a variety of OLS estimators. Each OLS estimator regresses observed expenditures (including the zero values) on a set of covariates; we do not use two-part models.¹⁰ Because we use level expenditures as the dependent variable in the OLS regressions, we need not

¹⁰Since $E(Y|x) = p(Y = 0|x) \cdot 0 + p(Y > 0|x) \cdot E(Y|Y > 0, x) = R(x)$, $E(Y|x)$ is a function of x only. The one-part models that we estimate are approximations to the expected value function $R(x)$.

consider issues related to how one can translate from regressions of the outcome estimated in logarithms to levels.¹¹

The OLS estimators we consider are: 1) a simple OLS model with only linear covariates [labeled OLS (levels, 1st order) in the tables];¹² 2) an OLS model that uses (up to) fourth order polynomials in the levels of the explanatory variables where the forms of the polynomial interactions are the same as those used in the conditional density estimation procedure except there is no need to specify the α_k 's as this linear form is the conditional expectation [labeled OLS (levels, 4th order)];¹³ and 3) an OLS estimator identical to the second one, except that it uses polynomials in the logs of the transformed explanatory variables rather than polynomials in the levels of the explanatory variables [labeled OLS (logs, 4th order)]. A list of the full set of expansion terms used in estimation are provided in Appendix Table A3. The frequency of selection of order of the polynomial expansion is provided for each model for each data generating process in Appendix Table A4.

In Tables 1a through 1e we report the derivative of expected expenditures with respect to the insurance deductible amount from several estimation models for each DGP in the Monte Carlo experiments. The deductible is that amount of health expenditure dollars that a consumer pays before the insurance company begins sharing the costs. Table 1a contains Monte Carlo evidence when the DGP is a simple OLS model, and the DGPs for Tables 1b through 1e are described in the table titles. We calculate the derivative by numerically differentiating with respect to the explanatory variable. When calculating this derivative, we hold constant the values of the error terms for each observation and rely upon the averaging across observations to integrate out the heterogeneity. Note that the simplest OLS model is the efficient estimator for the first DGP. For all other DGPs, all the empirical models are approximations to the true conditional expectations

¹¹We considered two-part models with logged expenditures conditional on any expenditures and the smearing retransformation in our comparisons of different models. We found that these models, when estimated with higher levels of polynomials whose coefficients could not be deemed insignificant, produced unreliable estimates and extremely large standard errors. Given this finding, we did not explore modifications of the smearing/retransformation approach in our Monte Carlo experiments. Examples of this type of result are found in our application to real data in the next section.

¹²For the first DGP this OLS estimator is the efficient estimator. In all others, the true DGP is considerably more complex than a simple linear model.

¹³We select the degree of the polynomial by a downwards testing approach using a similar Wald test at the five percent level for each Monte Carlo replication. We also calculate the derivatives of the expected values in a fashion similar to that used with the conditional density estimator. We take finite differences of expectations by deviating one explanatory variable from the value it took on in the data set exactly as we do for the conditional density estimator.

function. The tables report the numerical derivative at each chosen evaluation point and the average derivative calculated by numerically differentiating at every observed value in the data. The tables report the “true” derivative, where we similarly numerically differentiate the known DGP with respect to each explanatory variable, and the derivatives from the OLS and CDE models. We begin by demonstrating the ability of our quite non-linear conditional density estimation (CDE) technique to uncover the true derivative according to the Monte Carlo experiment.

In Table 1a, using OLS to estimate derivatives provides the most efficient estimator. We recover the true mean of -0.399 almost exactly and fairly accurately with a standard deviation is 0.057. Higher orders of level X 's and OLS also reproduce the true effects because the first order model is always chosen by the Wald tests. However, higher orders of logged X 's do a relatively poor job of fitting the truth. The mean value of the average derivative is almost two (OLS 1st order) standard deviations from the truth. The CDE technique with polynomials and interactions in the logged X 's produces estimates with little bias (from the truth) but with somewhat larger standard deviations. This larger variance is expected since OLS is efficient. However, the CDE procedure does very well when the data are generated from a simple model. It has the ability, generally, to capture the true constant derivative at different evaluation points.

The different DGPs examined in Tables 1b through 1e are intended to generate increasingly more heterogeneity in the data generating process. Immediately we see that the simple first order OLS model does not do a good job of fitting the data as the DGP becomes more complicated. Allowing for higher order level polynomials in the OLS model provides a better fit, but at a loss of efficiency. The best OLS model is one with polynomials and interactions in the logged values of the transformed explanatory variables. The CDE technique produces little bias and has standard deviations that are close to those provided by the one-part OLS models with fourth order interactions of all the explanatory variables.

We also examined the impacts of age, income, and coinsurance rate in these Monte Carlo experiments. The assessments of the various estimators for these alternative effects are similar to those for the deductible level examined in Table 1. The simple OLS model yields derivative estimates that are quite far from the true average derivatives in models with any complexity. The OLS estimators with fourth order terms and the CDE estimators all perform quite well in recovering the average derivative as well as the non-constant derivative at specific evaluation points. They

Table 1a: Average Deductible Derivatives
DGP: $Y_i = \beta' X_i + \epsilon_i$, iid normal errors, OLS

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.014 (0.883)	-0.190 (1.050)
50	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.807 (0.261)	-0.488 (0.567)
100	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.591 (0.200)	-0.404 (0.183)
150	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.473 (0.165)	-0.396 (0.123)
200	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.398 (0.142)	-0.398 (0.106)
250	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.346 (0.126)	-0.400 (0.097)
300	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.267 (0.101)	-0.397 (0.095)
Average*	-0.399 (0.000)	-0.396 (0.057)	-0.396 (0.057)	-0.493 (0.113)	-0.409 (0.172)
Min Variance*					-0.372 (0.061)
Equal Weights*					-0.382 (0.130)
Weight 1*					-0.382 (0.121)
Weight 2*					-0.391 (0.110)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table 1b: Average Deductible Derivatives
DGP: $\ln(Y_i) = \beta'X_i + \epsilon_i$, iid normal errors, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-2.632 (0.000)	-0.279 (0.064)	-1.173 (0.532)	-2.752 (0.947)	-2.543 (1.288)
50	-0.412 (0.000)	-0.279 (0.064)	-0.823 (0.290)	-0.435 (0.296)	-0.368 (0.584)
100	-0.219 (0.000)	-0.279 (0.064)	-0.536 (0.164)	-0.262 (0.146)	-0.227 (0.243)
150	-0.166 (0.000)	-0.279 (0.064)	-0.307 (0.158)	-0.188 (0.112)	-0.152 (0.180)
200	-0.131 (0.000)	-0.279 (0.064)	-0.133 (0.193)	-0.146 (0.098)	-0.112 (0.144)
250	-0.094 (0.000)	-0.279 (0.064)	-0.008 (0.223)	-0.119 (0.090)	-0.088 (0.123)
300	-0.070 (0.000)	-0.279 (0.064)	0.102 (0.311)	-0.081 (0.076)	-0.060 (0.118)
Average*	-0.513 (0.000)	-0.279 (0.064)	-0.518 (0.141)	-0.531 (0.117)	-0.533 (0.184)
Min Variance*					-0.472 (0.128)
Equal Weights*					-0.507 (0.163)
Weight 1*					-0.429 (0.164)
Weight 2*					-0.394 (0.145)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table 1c: Average Deductible Derivatives
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, $\text{var}(\epsilon) \sim E[\ln(Y)]$, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-2.577 (0.000)	-0.292 (0.077)	-1.257 (0.695)	-3.033 (1.394)	-2.965 (1.490)
50	-0.491 (0.000)	-0.292 (0.077)	-0.863 (0.360)	-0.360 (0.411)	-0.396 (0.725)
100	-0.241 (0.000)	-0.292 (0.077)	-0.547 (0.207)	-0.237 (0.188)	-0.229 (0.274)
150	-0.193 (0.000)	-0.292 (0.077)	-0.304 (0.212)	-0.183 (0.118)	-0.147 (0.181)
200	-0.121 (0.000)	-0.292 (0.077)	-0.126 (0.239)	-0.152 (0.096)	-0.106 (0.148)
250	-0.098 (0.000)	-0.292 (0.077)	-0.005 (0.250)	-0.130 (0.089)	-0.082 (0.130)
300	-0.082 (0.000)	-0.292 (0.077)	0.073 (0.393)	-0.100 (0.089)	-0.055 (0.120)
Average*	-0.583 (0.000)	-0.292 (0.077)	-0.537 (0.170)	-0.531 (0.134)	-0.527 (0.219)
Min Variance*					-0.522 (0.141)
Equal Weights*					-0.568 (0.169)
Weight 1*					-0.473 (0.152)
Weight 2*					-0.432 (0.145)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table 1d: Average Deductible Derivatives
DGP: $\ln(Y_i) = \beta_i'X_i + \epsilon_i = \bar{\beta}X_i + (\beta_i - \bar{\beta})X_i + \epsilon_i$, random coefficients model

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-2.477 (0.000)	-0.267 (0.100)	-1.004 (0.959)	-2.670 (1.591)	-2.695 (1.836)
50	-0.415 (0.000)	-0.267 (0.100)	-0.701 (0.503)	-0.458 (0.397)	-0.288 (0.842)
100	-0.211 (0.000)	-0.267 (0.100)	-0.466 (0.258)	-0.259 (0.228)	-0.193 (0.281)
150	-0.158 (0.000)	-0.267 (0.100)	-0.292 (0.263)	-0.184 (0.171)	-0.147 (0.236)
200	-0.125 (0.000)	-0.267 (0.100)	-0.172 (0.327)	-0.145 (0.140)	-0.124 (0.203)
250	-0.121 (0.000)	-0.267 (0.100)	-0.100 (0.348)	-0.121 (0.119)	-0.112 (0.171)
300	-0.060 (0.000)	-0.267 (0.100)	-0.098 (0.384)	-0.091 (0.090)	-0.101 (0.134)
Average*	-0.481 (0.000)	-0.267 (0.100)	-0.464 (0.235)	-0.531 (0.184)	-0.509 (0.213)
Min Variance*					-0.458 (0.173)
Equal Weights*					-0.523 (0.256)
Weight 1*					-0.433 (0.256)
Weight 2*					-0.394 (0.214)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table 1e: Average Deductible Derivatives
DGP: Mixture model where type depends on unobserved health state, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-5.814 (0.000)	-0.551 (0.090)	-2.982 (1.058)	-5.487 (1.865)	-4.149 (1.581)
50	-1.117 (0.000)	-0.551 (0.090)	-1.805 (0.518)	-0.908 (0.809)	-1.272 (0.845)
100	-0.714 (0.000)	-0.551 (0.090)	-0.977 (0.260)	-0.604 (0.264)	-0.618 (0.275)
150	-0.407 (0.000)	-0.551 (0.090)	-0.454 (0.260)	-0.461 (0.142)	-0.398 (0.201)
200	-0.538 (0.000)	-0.551 (0.090)	-0.188 (0.274)	-0.374 (0.112)	-0.304 (0.167)
250	-0.388 (0.000)	-0.551 (0.090)	-0.135 (0.239)	-0.315 (0.109)	-0.251 (0.153)
300	-0.221 (0.000)	-0.551 (0.090)	-0.665 (0.483)	-0.230 (0.125)	-0.176 (0.167)
Average*	-1.155 (0.000)	-0.551 (0.090)	-1.016 (0.233)	-0.848 (0.233)	-0.990 (0.236)
Min Variance*					-0.985 (0.182)
Equal Weights*					-1.024 (0.220)
Weight 1*					-0.884 (0.219)
Weight 2*					-0.851 (0.192)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

exhibit little bias and have roughly comparable standard errors. Comparisons of these estimated effects across the various estimation approaches for the other variables are provided in Appendix Tables A5 through A7.

3 An Application: Annual Medical Expenses

To demonstrate the flexibility and comparative advantages of the conditional density estimation procedure we use data from the Rand Health Insurance Experiment (RHIE) to explain annual medical expenses. The RHIE randomly assigned health insurance to participating individuals with the hope of overcoming adverse selection issues associated with the purchase of health insurance when examining the impacts of health insurance on medical care expenditures. While the data were collected some time ago (between 1974 and 1981), this data set is useful for our purposes because 1) these data have recently become available for public use, 2) results from the Experiment are widely known among health economists, and 3) the data exhibit features that presumably require special econometric treatment in estimation or prediction.

Individuals from six sites in different regions of the country participated in the RHIE which randomly assigned households to one of 19 different health insurance plans. Because dental and mental health expenses were compensated differently by many plans we restrict our attention to physician and hospital expenses. Similarly, we analyze expenses of individuals randomized to plans with free care or a coinsurance rate and a maximum deductible amount only. We do not consider Health Maintenance Organization (HMO) plans or plans with a deductible. We restrict our attention to individuals between the ages of 14 and 62. We drop individuals in Dayton, Ohio because it was a site that was atypical with regard to the Experiment. We consider the expenses of each individual in every full year of participation (for a total of 1 to 5 years) unless attrition was the result of death. In this case, we retain the part year observation because expenditures prior to death are likely to be influenced by health insurance. In order to keep the sample size as large as possible we impute values of missing variables such as income, general health index, and number of diseases using data on individuals with complete records. All dollar values are in 1999 dollars.

Table 2 lists summary statistics and definitions for variables used in our analysis.¹⁴ A list of the full set of expansion terms used in estimation are provided in Appendix Table A8.

Table 2: Summary Statistics from Rand HIE — All person years (8039)

Continuous Variables	Mean	Std.Dev.	Min	Max
Annual medical expenses (\$1999)	947.25	3052.48	0	79455.11
Annual family income in thousands (\$1999)	31.75	13.13	0	98.93
Annual participation incentive (\$1999)	994.04	866.28	0	2955.65
Index of general health at enrollment	70.34	14.15	5.7	100.00
Number of disease conditions at enrollment	10.33	8.04	0	58.60
Maximum deductible amount (\$1999)	876.02	943.09	0	2437.95
Coinsurance rate (%)	30.14	37.13	0	95
Age in years	32.76	13.13	14	61
Annual family size	3.53	1.85	1	14
Dummy Variables	Percent			
Seattle (omitted category)	27.37			
Fitchburg	16.15			
Franklin County	18.97			
Charleston	16.73			
Georgetown County	20.79			
Female	53.14			
Non-white	18.36			
Distribution of person-year observations by year	Percent			
1976	5.22			
1977	20.32			
1978	20.35			
1979	24.54			
1980	15.66			
1981	13.89			

The conditional density estimation procedure, as the Monte Carlo experiments demonstrate, produces reliable estimates of the effects of covariates regardless of the functional form of the dependent variable or the true, underlying regression framework. Thus, there is no need to concern oneself with appropriate treatment of the error term when retransforming the predictions to levels.

¹⁴A considerable amount of work went into preparation of the data set used in estimation in order to obtain consistent observations on each eligible person-year within a family and to produce a set of explanatory variables that most closely resembles that used in the published literature (Manning et al., 1987). SAS code to replicate our research data file is available upon request.

Additionally, the procedure explores high order polynomials and interactions of the explanatory variables and chooses the specification with the best fit. Finally, the estimation technique allows for variation in covariate effects over particular ranges of the dependent variable. For example, the effect of a one thousand dollar increase in income might have very different effects on expenditures at low income levels than at high income levels. These differential effects are not simply a trivial implication of a dependent variable transformation.

Table 3 reveals the varying derivatives for different values of selected explanatory variables. We begin by discussing the derivative of expenditures with respect to income. Standard errors for all estimation procedures come from the standard deviation of 50 bootstrap replications. All derivatives are calculated by taking finite differences as discussed above. For all estimation models, upwards testing for the appropriate level of expansion overwhelmingly selected the fourth order with these real data. In general, the CDE procedure reveals that \$23 of an additional \$1000 dollars at low income levels (\$10,000) will be spent on medical care. As income initially rises, individuals spend more on medical care, but at a decreasing rate. As income continues to rise to higher levels, there is little significant change in the amount spent on medical care. The second column of Table 3 reports results using the traditional two-part estimation technique and smearing to account for the predicted errors in retransformation of the predicted log expenditures to produce level outcomes. Here we use fourth order polynomials in the explanatory variables as we rejected the restrictions implied by a third order polynomial model at the 5% level. The two-part model with smearing yields qualitatively similar point estimates for the income effects, but the bootstrapped standard errors for this commonly used procedure are two to forty times larger than those of the CDE model.¹⁵

The last two columns of Table 3 contain estimates from one-part models where those individuals with zero expenditures are included in the same OLS regression as those with positive expenditures. Expenditures are measured in level real dollars, so there is no need to make ad hoc

¹⁵Much of the original Rand work analyzing the HIE data attempted to account for potential heteroscedasticity by using plan-specific smearing retransformations, while we have used a single retransformation correction using all observations with positive outcomes. Neither approach, however, can account for the full range of variation in the x 's that might be the source of heteroscedasticity. When we tried to estimate more complex forms of heteroscedasticity, the excessively large standard errors often became even more absurd.

Table 3: Selected Covariate Derivatives Using Alternative Models

Evaluation Point	CDE		Two-part, $\ln(Y)$, Smearing, 4th order		One-part, OLS Y , log X 's, 4th order		One-part, OLS Y , level X 's, 4th order	
	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error
Household Income (1000s 1999\$, mean: 32.7, sd: 16.9)								
10	22.923	(13.866)	13.727	(57.120)	21.932	(17.631)	18.039	(15.390)
15	5.687	(6.942)	0.055	(14.448)	6.199	(10.004)	9.568	(9.254)
20	-1.861	(5.062)	-4.491	(13.771)	-0.925	(7.243)	3.407	(7.581)
25	-4.458	(4.323)	-4.978	(20.910)	-4.009	(5.546)	-0.859	(7.134)
30	-4.671	(3.616)	-3.825	(33.364)	-5.102	(4.283)	-3.645	(6.186)
35	-3.770	(3.024)	-2.064	(52.397)	-5.170	(3.584)	-5.367	(5.138)
40	-2.380	(2.771)	-0.115	(80.257)	-4.703	(3.629)	-6.439	(5.387)
45	-0.811	(2.943)	1.869	(120.163)	-3.958	(4.259)	-7.275	(7.128)
50	0.790	(3.430)	3.846	(176.511)	-3.079	(5.158)	-8.291	(9.128)
avg	-0.008	(0.008)	-0.010	(0.013)	-0.005	(0.010)	0.003	(0.004)
Coinsurance Rate								
0	-10.076	(3.450)	111.435	(7.65 $\times 10^7$)	-6.347	(5.320)	-10.783	(5.254)
25	4.021	(6.845)	-114.300	(7.09 $\times 10^7$)	-0.428	(9.452)	2.950	(10.251)
50	-4.698	(4.058)	-7.038	(3.23 $\times 10^{12}$)	-4.846	(5.004)	-4.709	(4.900)
avg	-6.202	(2.085)	83.067	(5.55 $\times 10^7$)	-4.918	(2.267)	-7.445	(2.296)

Table 3: Selected Covariate Derivatives Using Alternative Models — continued

Evaluation Point	CDE		Two-part, $\ln(Y)$, Smearing, 4th order		One-part, OLS Y , $\log X$'s, 4th order		One-part, OLS Y , level X 's, 4th order	
	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error
General Health Index (higher is better health, mean: 70.3, sd: 14.2)								
30	-12.847	(30.583)	-58.472	(83.355)	-18.766	(41.129)	-36.987	(30.129)
40	-20.213	(12.935)	-39.185	(16.188)	-27.653	(16.839)	-35.839	(16.387)
50	-19.080	(5.325)	-25.090	(6.226)	-26.683	(8.988)	-26.234	(9.793)
60	-14.504	(3.800)	-15.186	(3.700)	-20.540	(7.531)	-14.337	(6.572)
70	-9.333	(3.180)	-8.245	(2.319)	-12.008	(4.319)	-6.309	(6.619)
80	-4.623	(4.629)	-3.084	(4.028)	-2.574	(5.937)	-8.316	(6.531)
avg	-11.354	(2.445)	-17.366	(30.022)	-14.244	(3.361)	-15.537	(3.286)
Female Indicator	388.258	(63.397)	435.579	(63.504)	361.629	(93.917)	337.893	(89.007)
Non-white Indicator	-193.209	(123.959)	-273.867	(2753.129)	-307.955	(157.995)	-288.725	(159.226)

assumptions to transform the estimated effects to level expenditures. The third column uses polynomials in the levels of the explanatory variables, while the last column uses polynomials in the logarithms of the adjusted explanatory variables. As in the Monte Carlo experiments, these one-part models perform almost as well as the CDE model. They yield point estimates quite similar to those of the CDE model, but their bootstrapped standard errors are almost always larger than those from the CDE approach.

Similar comparisons are found when we examine derivatives with respect to other variables. Consider the effects of the coinsurance rate, or the percent paid out-of-pocket by the individual, on health expenditures. In the RHIE data this variable is not continuous, as it takes on only the values 0, 25, 50, and 95%. The free plan (0% coinsurance rate) also has associated with it a \$0 maximum deductible amount (MDE); the MDE is that level of out-of-pocket expenditures after which additional medical care is free. In calculation of the impact on expenditures of changes in the coinsurance rate, we have to move individuals from the free plan to a plan with some out-of-pocket responsibility. Such a move requires that we assign an MDE at the evaluation of a non-zero coinsurance rate. We set the new MDE to 10% of family income or \$2200 (1999 dollars), whichever is less.¹⁶ This reassignment of originally free plans to paying plans is considerably different from simply increasing the percentage paid by the individual who faced some out-of-pocket responsibility originally. The CDE technique significantly predicts that movement from a free plan to a 25% plan results in a \$10 reduction in total health care expenditures per coinsurance percentage point increase (i.e., a \$250 decrease in expenditures with movement from a free plan to the 25% plan). Changes in behavior associated with movement from a 25% plan to a 50% plan are less precise (across all models) because few individuals in the RHIE were randomized to the 50% plan (about 5% of our sample). On average, across all coinsurance rates, a one percentage point increase in out-of-pocket responsibility results in a \$6 decrease in expenditures.

The estimates of the coinsurance effects from the two-part smearing model are disturbing. In several of the bootstrap replications, a small fraction of the sample had absurdly large predicted medical expenditures. This appears to happen because, for some combinations of the explanatory variables, changing the coinsurance rate to evaluate the effect yields a set of explanatory variables

¹⁶The RHIE assigned MDE's as 5, 10 and 15% of income or \$1000 (current year dollars), whichever was less. Note that this way of setting the MDE in the RHIE is inconsistent with the true experimental design.

that is not well represented in the estimation sample. While the predictions of log expenditures in these instances are not too extreme, once one antiloggs the predicted level expenditures become huge. This sensitivity to slightly out of sample predictions only seems to affect substantially the two-part smearing model. Both of the one-part level outcome models, also with fourth degree polynomials in the explanatory variables, yield estimates of the coinsurance effects close to those from the CDE model. As for the income effects discussed above, these two OLS models do have larger bootstrapped standard errors than the CDE approach.

The effect on health expenditures, as measured by the derivative with respect to the general health index, reveals that better health leads to lower medical expenditures. In terms of this health index, a one unit increase in health has a larger reduction in expenditures at low levels of health than at levels at or above the mean level of the health index. Again, the two-part smearing model provides imprecise estimates. The two OLS models predict somewhat larger responses than the CDE model, and, as above, they uniformly have larger bootstrapped standard errors. The differences in mean expenditures between men and women and between whites and non-whites are quite similar across estimation procedures. As above, the CDE model has the smallest bootstrapped standard errors of all procedures.

We next explore how some of the average effects described in Table 3 vary across subsets of individuals. To do this, we take each person-year in the RHIE dataset, change two or three characteristics of the explanatory variables at a time, and examine how predicted health expenditures vary for each characteristic. The changes correspond directly to some of those in Table 3, except that the impacts can vary by health and demographic characteristics. Because all the evidence from the Monte Carlo experiments and from Table 3 indicate that the CDE model provides accurate estimates, we only present these multidimensional derivatives as calculated from the CDE model. Bootstrapped standard errors are after the derivatives at each point and are in parentheses.

The first panel in Table 4 displays how age effects vary by gender. At age 20, for example, men appear to increase health expenditures by \$4.37 for each year they age. By age 30, health expenditures of men increase by \$12.06 for every year they get older, and at age 45 and later expenditures are increasing, on average, by more than \$16.00 per year. Expenditures of women follow a much different pattern. During their teens and twenties, young womens' health expenditures, on average, rise rapidly. This is due to pregnancy and childbirth costs. Linear interpolation of the age

Table 4: Derivative Effects by Group Using the Conditional Density Estimation Technique

Evaluation Point	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error								
By Gender																
	Male				Female											
Age = 15	0.579	(21.791)	71.456	(20.878)												
Age = 20	4.368	(10.076)	40.661	(15.009)												
Age = 25	8.667	(7.906)	14.646	(12.091)												
Age = 30	12.057	(7.570)	-1.499	(10.133)												
Age = 35	14.485	(7.201)	-10.803	(8.560)												
Age = 40	15.979	(6.585)	-15.706	(7.584)												
Age = 45	16.563	(7.518)	-17.689	(9.029)												
Age = 50	16.312	(11.087)	-17.778	(11.684)												
Average	12.388	(4.401)	15.055	(4.967)												
By Gender and Race																
	White Male				Non-white Male				White Female				Non-white Female			
Income = 10K	18.300	(14.953)	11.447	(10.523)	29.914	(17.294)	18.372	(14.230)								
Income = 15K	1.674	(8.316)	6.084	(5.593)	8.006	(9.229)	10.970	(7.912)								
Income = 20K	-5.246	(6.859)	3.795	(5.594)	-1.724	(6.706)	7.704	(7.665)								
Income = 25K	-7.296	(6.033)	3.237	(6.304)	-5.296	(5.560)	6.880	(8.408)								
Income = 30K	-7.107	(5.038)	3.588	(6.867)	-5.891	(4.581)	7.372	(9.052)								
Income = 35K	-5.962	(4.074)	4.423	(7.386)	-5.070	(3.800)	8.575	(9.625)								
Income = 40K	-4.475	(3.361)	5.527	(7.958)	-3.609	(3.412)	10.169	(10.233)								
Income = 45K	-2.928	(3.026)	6.789	(8.626)	-1.897	(3.482)	11.981	(10.926)								
Income = 50K	-1.440	(3.074)	8.150	(9.405)	-0.128	(3.903)	13.909	(11.720)								
Average	-0.006	(0.007)	0.006	(0.008)	-0.005	(0.010)	0.007	(0.009)								
Coinsurance = 0%	-4.618	(4.248)	-1.490	(7.157)	-15.850	(5.154)	-10.769	(10.641)								
Coinsurance = 25%	-2.485	(7.720)	-12.127	(5.734)	12.402	(8.702)	-8.189	(8.408)								
Coinsurance = 50%	-5.171	(4.284)	8.436	(4.219)	-9.436	(5.848)	12.868	(6.082)								
Average	-4.425	(2.392)	-0.840	(3.639)	-9.836	(2.831)	-4.198	(5.996)								

Table 4: Derivative Effects by Group Using the Conditional Density Estimation Technique — continued

Evaluation Point	Mean	Std.Error	Mean	Std.Error	Mean	Std.Error
By Income						
	Income = 10K		Income = 20K		Income = 30K	
Coinsurance = 0%	-4.455	(9.675)	-14.294	(6.087)	-13.515	(4.651)
Coinsurance = 25%	-6.546	(10.483)	-1.423	(8.174)	3.986	(7.369)
Coinsurance = 50%	-3.625	(5.053)	-2.973	(3.985)	-4.132	(4.178)
Average	-2.792	(5.093)	-9.695	(3.326)	-8.770	(2.806)
GHI = 30	-25.290	(33.713)	-25.001	(34.047)	-17.386	(32.687)
GHI = 40	-27.572	(15.113)	-30.776	(14.959)	-25.116	(13.846)
GHI = 50	-22.704	(7.108)	-26.732	(7.133)	-23.042	(6.154)
GHI = 60	-16.007	(6.176)	-19.660	(5.443)	-17.420	(4.490)
GHI = 70	-9.806	(6.435)	-12.745	(4.848)	-11.460	(3.706)
GHI = 80	-4.649	(7.591)	-6.938	(5.821)	-6.248	(4.825)
Average	-13.382	(5.070)	-16.078	(4.331)	-13.876	(3.249)
By General Health Index						
	GHI = 50		GHI = 60		GHI = 70	
Coinsurance = 0%	-10.992	(6.045)	-9.083	(4.647)	-8.981	(3.929)
Coinsurance = 25%	11.726	(10.938)	7.193	(8.084)	3.408	(6.733)
Coinsurance = 50%	-8.182	(7.198)	-5.879	(5.199)	-4.142	(3.919)
Average	-6.484	(3.189)	-5.440	(2.546)	-5.544	(2.187)
By Race						
	White		Non-white			
Female Indicator	393.183	(73.822)	384.401	(173.283)		

effects implies that womens' average health expenditures do not return to their average age 20 level until they are 45 years old.

The second panel of Table 4 explores how the income effects vary by race and gender. For each of the four groups, additional income has the largest impact on health expenditures at the lowest income levels. At income level \$10,000, for example, an additional thousand dollars in income increases health expenditures by \$11.45 for non-white males and by \$29.91 for white females. For non-whites, the point estimates indicate that additional income always increases health expenditures at higher income levels. But these point estimates for both whites and non-whites are relatively small and not significantly different from zero.

The second panel in Table 4 also contains how the effects of changing the coinsurance rate vary by race and gender. Except for non-white males, the largest reduction in health expenditures occurs when the coinsurance rate is raised from zero. Presumably, even a small copayment can reduce substantively average health expenditures.

The third panel of Table 4 examines the coinsurance and health effects on expenditures as a function on income level. Somewhat surprisingly, a rise in the coinsurance rate from zero reduces health expenditures more for those with average incomes (\$30K) than for those with quite low incomes (\$10K). Given the low levels of expenditures at very low incomes, this might simply reflect the fact that the poor have very little in the way of "discretionary" health expenditures.

The final three sets of results indicate that the CDE models can easily estimate constant effects across levels of characteristics. The bottom of the third panel in Table 4 indicates that the impact of health status, as measured by the general health index, does not vary by income level. Likewise, the fourth panel in Table 4 indicates that the effect of the coinsurance rate on expenditures does not vary across health levels. The final panel of this table shows that there is almost no difference by race in the male-female expenditure differential.

4 Conclusion

This paper explores the performance of a new approach for estimating how expected derivatives of an outcome vary with covariate values when the distribution of the outcome is characterized by a point mass at zero and large positive skews. Such types of outcomes are often encountered

in health economics, where a significant fraction of people have no health expenditures and a few have extremely high expenditures. In Monte Carlo work we calibrated our experimental design to National Medical Expenditure Survey data and compared the performance of the approximate conditional density estimator to a simple ordinary least squares model and to more complex ordinary least squares models that use as explanatory variables polynomials in the explanatory variables used in the simple OLS models. Overall, we found that the approximate conditional density estimator that we propose provided accurate and precise estimates of derivatives of expected outcomes for a wide range of types of explanatory variables. The simple OLS models performed quite poorly, while the OLS models including the higher order polynomials in the explanatory variables performed nearly as well as the conditional density estimation approach.

Using the CDE model we reexamine the Rand Health Insurance Experiment data and uncover several new empirical results. We find that the largest increases in health expenditures due to increases in income happen at the lowest levels of the income distribution. We also find that increases in the coinsurance rate from free health care to low levels of copayment reduce health care expenditures more for those with average incomes than for those with well-below average incomes. It might be the case that those with very low incomes have few discretionary health expenditures. The third most important result from our examination of the Rand Health Insurance Experiment data is that increases in health expenditures due to declines in health do not vary by income level. On average, the poor increase their health expenditures by the same amount as those with average incomes when their health, as measured by Rand's general health index, declines. We also reconfirm other researchers' results that the largest impacts of increases in coinsurance rates on health expenditures take place when one raises the coinsurance rate from zero (Newhouse, 1993). It is important to note that these effects are based on estimates from the Rand Health Insurance Experiment where even the poorest individuals in the analysis sample do have health insurance.

We also examined briefly the performance of the commonly used two-part model, where one estimates whether or not there are positive outcomes with a logit model in the first part and a simple OLS model relating the logarithm of the positive outcomes to explanatory variables in a second part of the statistical model. In order to calculate how the expected outcome changes with covariates in this framework, it is necessary to exponentiate the predicted log-outcome and adjust multiplicatively for the expectation of the antilog of the disturbance term. In our preliminary Monte

Carlo work we found that this approach yielded imprecise and often absurd estimates. The worst performances happened in those models where the true data generating process differed from the exact statistical specification used to define the two-part model with logarithmic outcomes. When we tested for whether one could use a simple specification of the explanatory variables in the linear-in-logarithm part of the model instead of including higher order polynomial terms, we often rejected the simpler model because it did not fit the data well. Once higher order terms were introduced into this continuous, logarithmic part of the model, the estimation and approximation errors interacted with the exponentiation of the expected outcome to yield quite inaccurate predictions of the level outcomes. The corrections for the expectation of the antilog of the disturbance did not introduce anywhere near as much noise into the estimates, unless we attempted to model and estimate the heteroscedasticity of the error terms when applying a “smearing” type correction. We ended up not using the two-part models with predicted logarithmic expenditures in our Monte Carlo experiments, as the preliminary evidence suggested that this approach was dominated by the other estimators we examined. Such extreme inaccuracies were not important for the conditional density estimators or for the OLS models with *level* dependent variables and polynomials in the explanatory variables. Our examination of the Rand Health Insurance Experiment data does provide an illustration of the poor performance of the two-part model approach used with logarithmic positive expenditures and smearing.

The implications from our analysis, however, are quite encouraging. They indicate that researchers who wish to estimate the impact of exogenous changes in covariates such as income, coinsurance rate, insurance deductible, and health status on expected health expenditures need only estimate standard ordinary least squares models that combine zero expenditures along with the level of positive expenditures. The Monte Carlo work also indicates that it is imperative to use generously specified functions of and interactions among the explanatory variables in these OLS models. The simplest OLS models we explore in the Monte Carlo experiments, i.e., those with level outcomes and regressors and no higher order terms or interactions, almost always fail to measure anything resembling a population average derivative in even moderately complex models.

The only drawback to using the OLS and CDE models with polynomials and interactions of the explanatory variables is the seemingly low precision of the estimates. The standard deviation for the CDE estimator of the deductible derivative in Table 1a, for example, is three times larger

than the standard deviation of the OLS estimator (0.172 vs. 0.057); these standard deviations indicate the magnitude of the standard errors of the estimates from these estimators. By using the flexible CDE model with the simple linear model DGP in Table 1a (and also Tables A5a, A6a, and A7a, all of which have the same DGP), calculated t-statistics will be about three times smaller for the CDE than for the simple OLS estimator. This “lack of precision,” however, is due to the fact that the CDE model allows for the possibility that the effect of insurance deductible can vary with the level of the deductible and with the levels of the other covariates. The OLS model imposes the true restriction, for this DGP, that the deductible effect is constant across all dimensions.

To put the CDE model onto a more level playing field with the OLS estimator for this DGP, consider imposing the restriction that the CDE estimated deductible derivatives are estimating the same quantity at each level of the deductible that we examined in Table 1a. We do not impose the restriction that these derivative levels are also constant across all values of all of the other covariates, as is imposed by the simple OLS estimator. We impose these restrictions ex post by calculating the covariance matrix for the CDE estimators of the derivative at the seven deductible levels displayed in this table and solving for that weighted average of the seven point derivatives that yields the smallest variance. This result is the evaluation point labeled “Min Variance” in Table 1a and the other tables of Monte Carlo results. We find that by imposing this restriction ex post, the standard error of the CDE estimator of the deductible effect is only 7% higher than that of the OLS estimator. If one is willing to impose this additional information on the empirical model, the CDE approach can do nearly as well as the efficient OLS estimator.

Looking across the tables for all of the other DGPs, i.e., those where the classical regression model is not a true representation of the DGP (Tables 1b-1e, A5b-A5e, A6b-A6e, A7b-A7e), it is clear that in almost all cases that the simple OLS estimator has the largest bias of all of the four approaches we examine. It rarely has an average effect that is closer to the true effect than any of the other approaches, and often the bias exceeds 50% of the true average derivative. Simple OLS still retains the feature that it has the smallest standard deviation of all of the estimators, but for these DGPs this “precision gain” is more than offset by the misleading information conveyed by the OLS point estimate of the derivative. This result highlights the fact that relying upon simple regression specifications can lead to quite misleading implications. These Monte Carlo experiments demonstrate that is not the case that a mis-specified model is correct on average. Simple and

mis-specified models can exhibit significant biases, and tests based upon estimates from these approaches will provide misleading information. The flexible approaches that we use in the Monte Carlo experiments and in the examination of the RHIE data do not have this drawback. They do exhibit larger standard errors than the simple OLS models, but these larger standard errors merely reflect the fact that the researcher is, truthfully, often unsure about the precise form of the true regression specification. The apparent precision for the simple OLS estimator is fictional and only comes from the fact that one estimates many fewer parameters than are needed to fit the data well.

These CDE models allow one to model more directly the impact of covariates on outcomes. One could, for example, model exactly how the budget constraint and structure of coinsurance rates and deductible levels impact health expenditures at various points in the health expenditure distribution. To do this well would require that one address issues of unobserved endogeneity, and the CDE approach allows one to use the economic structure of the budget constraint to provide important restrictions to help achieve identification. A final advantage of the CDE approach is that in most real world situations researchers will not have experimental data with random assignment of the features of the health insurance programs that individuals face. Examination of such data requires that one model endogeneity and address sample selection issues. Accurate estimation will often entail the use of maximum likelihood models that fit the data well, and that is precisely what the conditional density estimation models can do.

References

- Allison, P. (1984). Event History Analysis : Regression for Longitudinal Event Data. Beverly Hills, Calif. : Sage Publications.
- Angrist, J. (2000). "Estimation of Limited-Dependent Variable Models with Binary Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*, forthcoming.
- Cameron, A. and P. Trivedi (1998). Regression Analysis of Count Data, Cambridge: Cambridge University Press.
- Donald, S., D. Green, and H. Paarsch (2000). "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates," *Review of Economic Studies*, forthcoming.
- Eastwood, B. and R. Gallant (1991). "Adaptive Rules for Semiparametric Estimators that Achieve Asymptotic Normality," *Econometric Theory* 3, 307-340.
- Efron, B. (1988). "Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve," *Journal of the American Statistical Association* 83, 414-425.
- Hahn, J., P. Todd, and W. Van der Klaauw (2000). "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, forthcoming.
- Heckman, J. and B. Honore (1990). "The Empirical Content of the Roy Model," *Econometrica* 58, 1121-49.
- Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52, 271-320.
- Manning, W. (1998). "The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem," *Journal of Health Economics* 17, 283-295.
- Manning, W. and J. Mullahy. (2000). "Estimating Log Models: To Transform or Not to Transform," working paper.

- Manning, W., J. Newhouse, N. Duan, E. Keeler, A. Leibowitz, and M. Marquis (1987). "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review* 77, 251-277.
- Maddala, G. (1983). "A Survey of the literature on Selectivity Bias as it Pertains to Health Care Markets," in Scheffler, R. and L. Rossiter (eds.) Advances in Health Economics and Health Services Research 6, London: JAI Press.
- Meyer, B. (1990). "Unemployment Insurance and Unemployment Spells," *Econometrica* 58, 757-782.
- Mullahy, J. (1998). "Much Ado About Two: Reconsidering the Retransformation and the Two-part Model in Health Economics," *Journal of Health Economics* 17, 247-281.
- Mroz, T. (1999). "Discrete Factor Approximation in Simultaneous Equation Models: Estimating the Impact of a Dummy Endogenous Variable on a Continuous Outcome," *Journal of Econometrics* 92, 233-274.
- Mroz, T. and D. Guilkey (1992). "Discrete Factor Approximations for Use in Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables," manuscript, University of North Carolina at Chapel Hill.
- Mroz, T. and D. Weir (1990). "Structural Change in Life Cycle Fertility during the Fertility Transition: France before and after the Revolution of 1789," *Population Studies* 44, 61-87.
- Newhouse, J. (1993). Free for All? Lessons from the RAND Health Insurance Experiment, Cambridge: Harvard University Press.

Table A1: Summary Statistics from Rand HIE — First year of each person (2479)

Continuous Variables	Mean	Std.Dev.	Min	Max
Annual medical expenses (\$1999)	783.52	2592.95	0	79455.11
Annual family income in thousands (\$1999)	31.79	13.56	0	98.93
Annual participation incentive (\$1999)	995.46	864.89	0	2955.65
Index of general health at enrollment	70.34	13.93	5.7	100.00
Number of disease conditions at enrollment	10.31	7.97	0	58.60
Maximum deductible amount (\$1999)	876.84	937.53	0	2437.95
Coinsurance rate (%)	30.87	37.50	0	95
Age in years	31.10	13.56	14	61
Annual family size	3.71	1.91	1	13
Dummy Variables	Percent			
Seattle (omitted category)	27.43			
Fitchburg	15.69			
Franklin County	19.00			
Charleston	16.54			
Georgetown County	21.34			
Female	52.84			
Non-white	18.96			
Distribution of First Year Observed	Percent			
1976	16.94			
1977	49.70			
1978	1.98			
1979	27.71			
1980	2.38			
1981	1.29			

Table A2: Average Derivative Definitions

Derivative	Definition
	Let the evaluation points for calculation of the numerical derivative with respect to variable k be denoted $e_{jk}, j = 1, \dots, J(k)$. We approximate the derivative by finite differences.
	Let the sample size at each evaluation point for variable k be denoted n_{jk} and the cutoff points be denoted c_{kj} where n_{jk} is the number of observations observed to have the value of variable k fall in the interval $[c_{jk}, c_{j+1,k})$.
Average	For each observed value of X_k , evaluate the function at $X_k \pm \frac{\delta}{2}$ where δ is either the exact distance between adjacent evaluation points (i.e., $\delta = e_{j+1,k} - e_{jk}$) or the average of the distances between the evaluation points (i.e., $\delta = (\sum_j e_{jk})/J(k)$). The choice of δ depends on the nature of the observed data (continuous or discrete). Calculate the difference between values of the function at each deviated point.
Min Variance	In the calculations of the derivatives for variable k , find the covariance matrix of the $J(k)$ derivatives. Find the weighted sum that minimizes the sum of the squared errors.
Equal Weights	Calculate a simple average of the $J(k)$ derivatives at each evaluation point. This measure accounts for neither the sample size at different evaluation points nor the variances or covariances of the $J(k)$ derivatives.
Weight 1	Calculate the weighted average of the derivatives where the weight is defined as the sample size about the evaluation point (n_{jk}). This measure accounts for the sample size at different evaluation points but not the variances or covariances of the $J(k)$ derivatives.
Weight 2	Calculate the weighted average of the derivatives where the weight is defined as $w_{jk} = 1/((1/n_{jk}) + (1/n_{j+1,k}))$ for each derivative evaluated at a specific evaluation point. This measure accounts for sample size at different evaluation points. In part, it accounts for the variances because it accounts for the sample sizes about the points at where the derivatives are calculated.

Table A3: List of Explanatory Variables used in Monte Carlo Experiments

Order of Polynomial Expansion	List of Variables								
First Order	ded	≡	insurance deductible level						
	age	≡	age						
	inc	≡	income						
	coi	≡	coinsurance rate (%)						
	fem	≡	female indicator						
	nwh	≡	non-white indicator						
	hsg	≡	good health status indicator						
	hsf	≡	fair/poor health status indicator						
	α_k	≡	function of cell indicator used only in CDE model						
	α_k^2	≡	squared α used only in CDE model						
Second Order	ded*ded	ded*coi	ded*inc	ded*age	ded*fem	ded*nwh	ded*hsg	ded*hsf	
		coi*coi	coi*inc	coi*age	coi*fem	coi*nwh	coi*hsg	coi*hsf	
			inc*inc	inc*age	inc*fem	inc*nwh	inc*hsg	inc*hsf	
				age*age	age*fem	age*nwh	age*hsg	age*hsf	
						fem*nwh	fem*hsg	fem*hsf	
							nwh*hsg	nwh*hsf	
	α_k *(all 1st order variables)								
	α_k^3								
Third Order	ded ² *ded	ded ² *coi	ded ² *inc	ded ² *age	ded ² *fem	ded ² *nwh	ded ² *hsg	ded ² *hsf	
	coi ² *ded	coi ² *coi	coi ² *inc	coi ² *age	coi ² *fem	coi ² *nwh	coi ² *hsg	coi ² *hsf	
	inc ² *ded	inc ² *coi	inc ² *inc	inc ² *age	inc ² *fem	inc ² *nwh	inc ² *hsg	inc ² *hsf	
	age ² *ded	age ² *coi	age ² *inc	age ² *age	age ² *fem	age ² *nwh	age ² *hsg	age ² *hsf	
	α_k *(all 2nd order variables)								
	α_k^2 *(all 1st order variables)								
	α_k^4								
Fourth Order	ded ³ *ded	ded ³ *coi	ded ³ *inc	ded ³ *age	ded ³ *fem	ded ³ *nwh	ded ³ *hsg	ded ³ *hsf	
	coi ³ *ded	coi ³ *coi	coi ³ *inc	coi ³ *age	coi ³ *fem	coi ³ *nwh	coi ³ *hsg	coi ³ *hsf	
	age ³ *ded	age ³ *coi	age ³ *inc	age ³ *age	age ³ *fem	age ³ *nwh	age ³ *hsg	age ³ *hsf	
	inc ³ *ded	inc ³ *coi	inc ³ *inc	inc ³ *age	inc ³ *fem	inc ³ *nwh	inc ³ *hsg	inc ³ *hsf	
	ded ² *coi ²	ded ² *inc ²	ded ² *age ²	coi ² *inc ²	coi ² *age ²	inc ² *age ²			
	α_k *(all 3rd order variables)								
	α_k^3 *(all 1st order variables)								

The first part of all two-part specifications allows for different effects of the explanatory variables and does not include the shape parameters or its interactions in the CDE model.

Table A4: Frequency of Selection of Order of Polynomial Expansion
for each Data Generating Process

Model	Order of Polynomial Expansion				
	First	Second	Third	Fourth	Total
DGP: $Y_i = \beta' X_i + \epsilon_i$, iid normal errors, OLS					
OLS levels	50	0	0	0	50
OLS logs	7	42	1	0	50
CDE	0	0	0	50	50
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, iid normal errors, 2-part					
OLS levels	0	8	31	11	50
OLS logs	0	44	4	2	50
CDE	0	0	0	50	50
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, $\text{var}(\epsilon) \sim E[\ln(Y)]$, 2-part					
OLS levels	0	11	31	8	50
OLS logs	0	41	8	1	50
CDE	0	0	0	50	50
DGP : $\ln(Y_i) = \beta_i' X_i + \epsilon_i = \bar{\beta} X_i + (\beta_i - \bar{\beta}) X_i + \epsilon_i$, random coefficients model					
OLS levels	0	30	10	10	50
OLS logs	0	41	7	2	50
CDE	0	0	0	50	50
DGP : Mixture model where type depends on unobserved health state, 2-part					
OLS levels	0	0	4	46	50
OLS logs	0	10	23	17	50
CDE	0	0	0	50	50

Table A5a: Average Age Derivatives
DGP: $Y_i = \beta' X_i + \epsilon_i$, iid normal errors, OLS

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
30	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	7.841 (1.733)	7.369 (2.272)
35	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	7.549 (0.748)	7.149 (1.720)
40	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	7.253 (0.752)	7.230 (1.733)
45	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	6.967 (1.172)	7.323 (1.613)
50	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	6.697 (1.539)	7.375 (1.763)
55	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	6.446 (1.822)	7.423 (2.904)
Average*	7.504 (0.000)	7.356 (0.597)	7.356 (0.597)	7.400 (0.949)	7.586 (1.481)
Min Variance*					7.371 (0.672)
Equal Weights*					7.311 (0.780)
Weight 1*					7.314 (0.899)
Weight 2*					7.313 (0.749)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A5b: Average Age Derivatives
DGP: $\ln(Y_i) = \beta'X_i + \epsilon_i$, iid normal errors, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
30	11.274 (0.000)	15.324 (1.031)	13.010 (2.152)	12.809 (1.742)	11.472 (2.267)
35	11.650 (0.000)	15.324 (1.031)	13.537 (2.209)	14.528 (1.411)	12.265 (2.067)
40	13.939 (0.000)	15.324 (1.031)	14.148 (2.354)	15.496 (1.408)	13.588 (2.356)
45	13.594 (0.000)	15.324 (1.031)	15.144 (2.489)	16.051 (1.603)	15.006 (2.301)
50	15.749 (0.000)	15.324 (1.031)	16.821 (2.859)	16.364 (1.917)	16.179 (2.522)
55	15.782 (0.000)	15.324 (1.031)	19.480 (5.542)	16.534 (2.350)	16.654 (4.044)
Average*	12.759 (0.000)	15.324 (1.031)	15.042 (1.908)	13.504 (1.165)	13.477 (1.713)
Min Variance*					13.871 (1.047)
Equal Weights*					14.194 (1.195)
Weight 1*					13.160 (1.024)
Weight 2*					13.591 (0.998)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A5c: Average Age Derivatives
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, $\text{var}(\epsilon) \sim E[\ln(Y)]$, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
30	13.339 (0.000)	16.165 (0.956)	14.457 (2.595)	14.165 (2.238)	12.912 (2.634)
35	12.580 (0.000)	16.165 (0.956)	13.978 (2.641)	15.070 (1.817)	13.333 (2.502)
40	13.872 (0.000)	16.165 (0.956)	14.449 (2.443)	15.833 (1.785)	14.000 (2.383)
45	15.288 (0.000)	16.165 (0.956)	15.686 (2.756)	16.372 (1.860)	14.904 (2.164)
50	16.130 (0.000)	16.165 (0.956)	17.504 (3.063)	16.696 (2.087)	15.738 (2.565)
55	16.573 (0.000)	16.165 (0.956)	19.717 (5.237)	16.837 (2.733)	15.986 (4.537)
Average*	14.389 (0.000)	16.165 (0.956)	16.182 (1.815)	14.732 (1.466)	13.823 (1.819)
Min Variance*					14.354 (1.056)
Equal Weights*					14.479 (1.190)
Weight 1*					13.876 (1.239)
Weight 2*					14.128 (1.107)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A5d: Average Age Derivatives
DGP: $\ln(Y_i) = \beta_i'X_i + \epsilon_i = \bar{\beta}X_i + (\beta_i - \bar{\beta})X_i + \epsilon_i$, random coefficients model

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
30	12.963 (0.000)	18.157 (1.646)	13.669 (3.171)	13.881 (2.519)	14.528 (2.537)
35	13.464 (0.000)	18.157 (1.646)	14.794 (2.871)	16.352 (2.446)	16.231 (2.389)
40	16.153 (0.000)	18.157 (1.646)	16.733 (2.832)	17.996 (2.542)	17.347 (2.860)
45	18.340 (0.000)	18.157 (1.646)	19.103 (3.947)	19.217 (2.589)	17.795 (2.947)
50	20.627 (0.000)	18.157 (1.646)	21.526 (4.730)	20.199 (3.540)	17.318 (3.932)
55	22.756 (0.000)	18.157 (1.646)	23.618 (8.008)	21.036 (5.866)	15.277 (7.665)
Average*	16.067 (0.000)	18.157 (1.646)	17.162 (2.031)	15.524 (1.730)	14.422 (1.905)
Min Variance*					16.196 (1.547)
Equal Weights*					16.416 (2.081)
Weight 1*					15.837 (1.602)
Weight 2*					16.042 (1.739)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A5e: Average Age Derivatives
DGP: Mixture model where type depends on unobserved health state, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
30	12.291 (0.000)	16.489 (1.257)	15.396 (3.626)	14.635 (3.469)	13.706 (3.367)
35	13.134 (0.000)	16.489 (1.257)	15.691 (3.797)	15.140 (2.836)	14.108 (2.539)
40	13.041 (0.000)	16.489 (1.257)	14.968 (2.566)	15.570 (2.409)	14.933 (2.759)
45	15.615 (0.000)	16.489 (1.257)	14.868 (3.314)	16.229 (2.430)	15.790 (2.423)
50	16.841 (0.000)	16.489 (1.257)	17.035 (3.786)	17.149 (2.724)	16.585 (2.406)
55	17.258 (0.000)	16.489 (1.257)	23.110 (5.742)	18.295 (5.061)	17.389 (4.410)
Average*	14.135 (0.000)	16.489 (1.257)	15.801 (2.249)	14.644 (2.145)	14.169 (2.332)
Min Variance*					15.357 (1.249)
Equal Weights*					15.419 (1.281)
Weight 1*					14.757 (1.532)
Weight 2*					15.040 (1.323)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A6a: Average Income Derivatives
DGP: $Y_i = \beta' X_i + \epsilon_i$, iid normal errors, OLS

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
5	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	5.400 (2.908)	4.964 (3.562)
10	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	5.596 (0.915)	4.356 (1.539)
15	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	5.156 (0.649)	4.292 (1.239)
20	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	4.701 (0.714)	4.286 (1.079)
25	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	4.305 (0.770)	4.266 (0.931)
30	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	3.969 (0.798)	4.226 (0.887)
35	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	3.684 (0.806)	4.169 (0.974)
40	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	3.441 (0.804)	4.099 (1.143)
45	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	3.230 (0.795)	4.018 (1.341)
Average*	4.391 (0.000)	4.408 (0.453)	4.408 (0.453)	4.310 (1.592)	4.411 (1.849)
Min Variance*					4.217 (0.623)
Equal Weights*					4.297 (0.708)
Weight 1*					4.388 (0.877)
Weight 2*					4.420 (0.984)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A6b: Average Income Derivatives
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, iid normal errors, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
5	32.353 (0.000)	13.406 (0.857)	32.388 (5.360)	36.985 (3.233)	33.601 (4.648)
10	22.130 (0.000)	13.406 (0.857)	26.066 (2.241)	23.892 (1.485)	23.326 (1.881)
15	16.872 (0.000)	13.406 (0.857)	20.709 (1.717)	17.770 (1.252)	17.802 (2.023)
20	14.227 (0.000)	13.406 (0.857)	16.229 (2.580)	14.205 (1.288)	14.252 (2.078)
25	11.406 (0.000)	13.406 (0.857)	12.537 (3.087)	11.868 (1.319)	11.710 (1.938)
30	11.673 (0.000)	13.406 (0.857)	9.545 (3.271)	10.215 (1.326)	9.763 (1.825)
35	8.512 (0.000)	13.406 (0.857)	7.166 (3.497)	8.984 (1.322)	8.210 (1.867)
40	8.840 (0.000)	13.406 (0.857)	5.310 (4.118)	8.030 (1.312)	6.933 (2.084)
45	7.353 (0.000)	13.406 (0.857)	3.889 (5.198)	7.270 (1.302)	5.854 (2.426)
Average*	20.164 (0.000)	13.406 (0.857)	22.683 (1.609)	25.220 (1.322)	22.529 (1.811)
Min Variance*					14.653 (0.785)
Equal Weights*					14.606 (0.904)
Weight 1*					18.861 (0.871)
Weight 2*					20.004 (0.892)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A6c: Average Income Derivatives
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, $\text{var}(\epsilon) \sim E[\ln(Y)]$, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
5	30.522 (0.000)	13.862 (0.912)	31.142 (4.823)	36.797 (2.315)	34.075 (3.872)
10	24.562 (0.000)	13.862 (0.912)	25.751 (2.258)	24.294 (1.294)	23.400 (2.046)
15	17.096 (0.000)	13.862 (0.912)	21.095 (1.635)	18.359 (1.213)	17.883 (1.889)
20	14.553 (0.000)	13.862 (0.912)	17.121 (2.189)	14.860 (1.184)	14.445 (1.696)
25	12.383 (0.000)	13.862 (0.912)	13.774 (2.577)	12.540 (1.190)	12.059 (1.459)
30	10.243 (0.000)	13.862 (0.912)	11.002 (2.773)	10.885 (1.230)	10.290 (1.472)
35	10.686 (0.000)	13.862 (0.912)	8.750 (3.164)	9.642 (1.293)	8.921 (1.811)
40	8.625 (0.000)	13.862 (0.912)	6.965 (4.121)	8.672 (1.367)	7.827 (2.318)
45	7.520 (0.000)	13.862 (0.912)	5.592 (5.688)	7.894 (1.444)	6.929 (2.868)
Average*	23.489 (0.000)	13.862 (0.912)	22.700 (1.833)	25.628 (1.536)	22.871 (2.205)
Min Variance*					15.102 (0.990)
Equal Weights*					15.092 (1.209)
Weight 1*					19.135 (1.177)
Weight 2*					20.283 (1.244)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A6d: Average Income Derivatives
DGP: $\ln(Y_i) = \beta_i' X_i + \epsilon_i = \bar{\beta} X_i + (\beta_i - \bar{\beta}) X_i + \epsilon_i$, random coefficients model

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
5	28.395 (0.000)	11.573 (1.236)	22.747 (5.416)	28.427 (3.934)	27.908 (4.824)
10	19.444 (0.000)	11.573 (1.236)	19.666 (2.767)	19.763 (1.494)	18.553 (2.142)
15	14.883 (0.000)	11.573 (1.236)	17.002 (1.693)	15.392 (1.424)	14.285 (1.888)
20	12.681 (0.000)	11.573 (1.236)	14.725 (2.045)	12.729 (1.558)	11.942 (2.003)
25	11.092 (0.000)	11.573 (1.236)	12.803 (2.539)	10.925 (1.760)	10.456 (1.898)
30	10.219 (0.000)	11.573 (1.236)	11.208 (2.893)	9.616 (2.026)	9.382 (1.783)
35	8.816 (0.000)	11.573 (1.236)	9.908 (3.329)	8.619 (2.324)	8.518 (1.845)
40	9.702 (0.000)	11.573 (1.236)	8.873 (4.144)	7.833 (2.621)	7.765 (2.122)
45	7.666 (0.000)	11.573 (1.236)	8.074 (5.477)	7.196 (2.904)	7.076 (2.538)
Average*	21.139 (0.000)	11.573 (1.236)	17.709 (2.211)	20.200 (1.881)	18.865 (2.244)
Min Variance*					12.914 (1.171)
Equal Weights*					12.876 (1.254)
Weight 1*					15.703 (1.323)
Weight 2*					16.605 (1.423)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A6e: Average Income Derivatives
DGP: Mixture model where type depends on unobserved health state, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
5	37.375 (0.000)	15.628 (0.984)	39.450 (4.452)	39.784 (4.527)	40.267 (4.931)
10	27.317 (0.000)	15.628 (0.984)	30.468 (1.975)	27.726 (2.024)	28.573 (2.656)
15	21.204 (0.000)	15.628 (0.984)	23.459 (2.082)	22.122 (1.624)	22.068 (1.832)
20	19.742 (0.000)	15.628 (0.984)	18.292 (2.755)	18.894 (2.004)	17.901 (1.903)
25	16.285 (0.000)	15.628 (0.984)	14.834 (2.967)	16.784 (2.491)	15.048 (2.214)
30	13.597 (0.000)	15.628 (0.984)	12.954 (3.225)	15.289 (2.960)	13.000 (2.530)
35	12.681 (0.000)	15.628 (0.984)	12.521 (4.598)	14.166 (3.388)	11.474 (2.843)
40	12.177 (0.000)	15.628 (0.984)	13.401 (7.461)	13.286 (3.771)	10.310 (3.155)
45	10.595 (0.000)	15.628 (0.984)	15.465 (11.597)	12.572 (4.111)	9.409 (3.491)
Average*	29.559 (0.000)	15.628 (0.984)	28.424 (2.083)	29.753 (2.292)	27.355 (2.595)
Min Variance*					18.736 (1.180)
Equal Weights*					18.672 (1.477)
Weight 1*					23.371 (1.313)
Weight 2*					24.694 (1.450)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A7a: Average Coinsurance Derivatives
DGP: $Y_i = \beta' X_i + \epsilon_i$, iid normal errors, OLS

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-5.000 (0.000)	-4.985 (0.955)	-4.985 (0.955)	-4.801 (4.017)	-4.644 (5.101)
10	-5.000 (0.000)	-4.985 (0.955)	-4.985 (0.955)	-4.585 (2.881)	-4.621 (3.247)
20	-5.000 (0.000)	-4.985 (0.955)	-4.985 (0.955)	-3.590 (2.350)	-4.695 (6.294)
Average*	-5.000 (0.000)	-4.985 (0.955)	-4.985 (0.955)	-4.184 (1.871)	-4.621 (3.054)
Min Variance*					-4.633 (1.484)
Equal Weights*					-4.653 (2.390)
Weight 1*					-4.683 (5.119)
Weight 2*					-4.634 (1.500)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A7b: Average Coinsurance Derivatives
DGP: $\ln(Y_i) = \beta'X_i + \epsilon_i$, iid normal errors, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-5.565 (0.000)	-9.393 (1.530)	-5.977 (4.899)	-4.528 (4.566)	-5.206 (5.429)
10	-7.270 (0.000)	-9.393 (1.530)	-7.810 (3.228)	-8.155 (3.316)	-7.082 (3.386)
20	-6.637 (0.000)	-9.393 (1.530)	-8.103 (10.263)	-7.027 (4.088)	-8.057 (9.418)
Average*	-6.212 (0.000)	-9.393 (1.530)	-7.668 (4.501)	-7.235 (2.338)	-7.395 (4.043)
Min Variance*					-6.626 (2.120)
Equal Weights*					-6.781 (3.663)
Weight 1*					-7.708 (7.696)
Weight 2*					-6.558 (2.136)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A7c: Average Coinsurance Derivatives
DGP: $\ln(Y_i) = \beta' X_i + \epsilon_i$, $\text{var}(\epsilon) \sim E[\ln(Y)]$, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-7.450 (0.000)	-9.791 (1.592)	-6.754 (6.809)	-4.172 (6.586)	-5.698 (6.733)
10	-7.468 (0.000)	-9.791 (1.592)	-8.145 (3.789)	-8.984 (3.500)	-7.472 (3.826)
20	-7.101 (0.000)	-9.791 (1.592)	-8.547 (8.204)	-7.496 (3.792)	-8.243 (7.828)
Average*	-7.122 (0.000)	-9.791 (1.592)	-8.198 (3.756)	-7.755 (2.473)	-7.645 (3.349)
Min Variance*					-7.064 (1.892)
Equal Weights*					-7.138 (2.912)
Weight 1*					-7.942 (6.298)
Weight 2*					-6.965 (1.934)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A7d: Average Coinsurance Derivatives
DGP: $\ln(Y_i) = \beta_i'X_i + \epsilon_i = \bar{\beta}X_i + (\beta_i - \bar{\beta})X_i + \epsilon_i$, random coefficients model

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-5.791 (0.000)	-9.430 (1.961)	-8.905 (5.742)	-3.995 (8.117)	-3.988 (7.898)
10	-7.656 (0.000)	-9.430 (1.961)	-7.420 (4.146)	-9.230 (4.465)	-8.573 (4.046)
20	-8.829 (0.000)	-9.430 (1.961)	-8.459 (12.778)	-7.374 (5.252)	-9.017 (11.922)
Average*	-8.092 (0.000)	-9.430 (1.961)	-8.012 (5.527)	-7.771 (3.423)	-8.048 (5.444)
Min Variance*					-7.186 (2.088)
Equal Weights*					-7.192 (4.190)
Weight 1*					-8.527 (9.641)
Weight 2*					-7.140 (2.157)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A7e: Average Coinsurance Derivatives
DGP: Mixture model where type depends on unobserved health state, 2-part

Derivative Evaluation point	Truth	OLS <i>Levels</i> <i>1st order</i>	OLS <i>Levels</i> <i>4th order</i>	OLS <i>Logs</i> <i>4th order</i>	CDE
0	-16.186 (0.000)	-16.352 (1.904)	-13.445 (8.236)	-12.729 (7.427)	-10.509 (8.116)
10	-12.461 (0.000)	-16.352 (1.904)	-12.808 (5.133)	-11.821 (4.542)	-16.380 (4.764)
20	-8.644 (0.000)	-16.352 (1.904)	-12.567 (21.886)	-12.757 (8.804)	-14.362 (9.012)
Average*	-11.443 (0.000)	-16.352 (1.904)	-13.514 (8.118)	-13.011 (4.436)	-15.211 (4.403)
Min Variance*					-14.250 (2.704)
Equal Weights*					-13.750 (3.701)
Weight 1*					-14.217 (7.336)
Weight 2*					-14.341 (2.695)

Standard deviations from the Monte Carlo experiments are in parentheses.
See Appendix Table A2 for measures of the average derivatives.

Table A8: List of Explanatory Variables used in HIE Application — continued

Order of Polynomial Expansion		List of Variables									
Third Order	α_k^* α_k^{2*} α_k^4 α_k^*	$mde^2 * mde$ $coi^2 * mde$ $inc^2 * mde$ $age^2 * mde$ $ghi^2 * mde$ $dis^2 * mde$ $fam^2 * mde$	$mde^2 * coi$ $coi^2 * coi$ $inc^2 * coi$ $age^2 * coi$ $ghi^2 * coi$ $dis^2 * coi$ $fam^2 * coi$	$mde^2 * inc$ $coi^2 * inc$ $inc^2 * inc$ $age^2 * inc$ $ghi^2 * inc$ $dis^2 * inc$ $fam^2 * inc$	$mde^2 * age$ $coi^2 * age$ $inc^2 * age$ $age^2 * age$ $ghi^2 * age$ $dis^2 * age$ $fam^2 * age$	$mde^2 * ghi$ $coi^2 * ghi$ $inc^2 * ghi$ $age^2 * ghi$ $ghi^2 * ghi$ $dis^2 * ghi$ $fam^2 * ghi$	$mde^2 * dis$ $coi^2 * dis$ $inc^2 * dis$ $age^2 * dis$ $ghi^2 * dis$ $dis^2 * dis$ $fam^2 * dis$	$mde^2 * fam$ $coi^2 * fam$ $inc^2 * fam$ $age^2 * fam$ $ghi^2 * fam$ $dis^2 * fam$ $fam^2 * fam$	$mde^2 * fem$ $coi^2 * fem$ $inc^2 * fem$ $age^2 * fem$ $ghi^2 * fem$ $dis^2 * fem$ $fam^2 * fem$	$mde^2 * nwh$ $coi^2 * nwh$ $inc^2 * nwh$ $age^2 * nwh$ $ghi^2 * nwh$ $dis^2 * nwh$ $fam^2 * nwh$	
Fourth Order	α_k^* α_k^{2*} α_k^3 α_k^*	$mde^3 * mde$	$coi^3 * coi$	$inc^3 * inc$	$age^3 * age$	$ghi^3 * ghi$	$dis^3 * dis$	$fam^3 * fam$			

The first part of all two-part specifications allows for different effects of the explanatory variables and does not include the shape parameters or its interactions in the CDE model.