

Econometric Modeling of Health Care Costs and Expenditures

A Survey of Analytical Issues and Related Policy Considerations

John Mullahy, PhD

Background: Econometric modeling of healthcare costs and expenditures has become an important component of decision-making across a wide array of real-world settings.

Objectives: The objective of this article is to provide a brief summary of important conceptual and analytical issues involved in econometric healthcare cost modeling. To this end, the article explores: outcome measures typically analyzed in such work; the decision maker's perspective in econometric cost modeling exercises; specific analytical issues in econometric model specification; statistical goodness-of-fit testing; empirical implications of "upper tail" (or "high cost") phenomena; and issues relating to the reporting of findings.

Data: Some of the concepts explored here are illustrated in light of samples drawn from the 2005 Medical Expenditure Panel Survey and the 2005 Nationwide Inpatient Sample.

Results and Conclusions: Analysts of healthcare cost data have at their disposal an increasingly sophisticated tool kit for analyzing such data that can in principle and in fact yield increasingly interesting insights into data structures. Yet for such analyses to usefully inform policy decisions, the manner in which such studies are designed, undertaken, and reported must accommodate considerations relevant to the decision-making community. The article concludes with some preliminary thoughts on how such bridges might be constructed.

Key Words: econometric models, health care costs, health care expenditures

(*Med Care* 2009;47: S104–S108)

This article provides a brief survey of issues in the econometric analysis of healthcare costs and expenditures. The main emphasis is on issues that policy makers who are users of such analysis may find important to consider. The bibliography referenced herein should offer readers direction to

pursue particular issues in greater depth (see also Jones, 2000, for an in-depth survey of health econometrics).¹

The article comprises 7 sections. Section 1 discusses leading applications for which econometric analysis of cost data is informative. Section 2 describes outcome measures typically analyzed in such exercises. Section 3 assesses decision makers' perspectives in econometric cost modeling exercises. Section 4 discusses specific analytical issues in econometric model specification. Section 5 addresses statistical goodness-of-fit (GOF) testing. Section 6 focuses on empirical implications of "upper tail" (or "high cost") phenomena. Section 7 discusses issues relating to the reporting of findings.

Two issues merit attention at the outset. First, many issues discussed here are equally applicable to discrete healthcare utilization outcomes (provider visits, inpatient days) as to outcomes measured in "continuous" currency units. Second, this article does not address issues relating to censored cost data; these are taken up in the article by Huang appearing in this volume.

APPLICATIONS

Econometric healthcare cost modeling is more than an academic exercise. Such empirical exercises inform fundamentally or peripherally decision-making over a wide array of domains: risk-adjusted provider payments; provider utilization review/profiling; cost-of-illness assessment; cost aspects of evaluation studies; and future projections of disease-specific healthcare cost burdens. Consequently, the care with which such econometric analysis is designed, conducted, and reported ultimately indeed matters.

DATA

To fix ideas, data series on healthcare expenditures and charges that will be familiar many analysts are described here. The expenditure data are the total healthcare expenditure series from the 2005 Medical Expenditure Panel Survey (MEPS) (the MEPS variable is *totexp05*). The charge data are the edited and unedited total charge series from the 2005 Nationwide Inpatient Sample (NIS) (*TOTCHG*; *TOTCHG_X*). Summary statistics are reported in Tables 1 and 2, respectively.

For many econometric explorations, not less than 4 prominent features of these data are typically important to accommodate. First, data on costs or expenditures are, for

From the Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin.

Reprints: John Mullahy, PhD, Department of Population Health Sciences, University of Wisconsin-Madison, 610 Walnut St., 787 WARF, Madison, WI 52726. E-mail: jmullahy@wisc.edu.

Copyright © 2009 by Lippincott Williams & Wilkins
ISSN: 0025-7079/09/4700-0104

TABLE 1. Summary Statistics, Total Expenditures, 2005 MEPS

Observations	33,961
0.05 Percentile	0
0.25 Percentile	72
0.50 Percentile	522
0.75 Percentile	2431
0.95 Percentile	13,130
Mean	3083
Standard deviation	9998

TABLE 2. Summary Statistics, Total Charges, 2005 NIS

	Edited Data	Unedited Data
Observations	7,842,319	7,845,698
0.05 Percentile	1583	1580
0.25 Percentile	5709	5708
0.50 Percentile	11,634	11,635
0.75 Percentile	24,297	24,308
0.95 Percentile	73,946	74,137
Mean	22,360	22,708
Standard deviation	39,221	49,351

most practical purposes, nonnegative. Second, a sizable fraction of observations (20.2%, unweighted, for totexp05) are measured as 0 (this consideration is not relevant in the NIS, which is an inpatient sample). Third, the data exhibit “heavy” upper tails: in the MEPS data, 1.5% of the totexp05 observations exceed $10\times$ the sample mean; in the NIS data, 0.6% of the TOTCHG observations exceed $10\times$ the sample mean. Fourth, the data are right-skewed; note that skewness per se does not imply a heavy upper tail. An instructive way to characterize such data is to assess how much of the density of data clusters near the sample mean.

Since many datasets that would be the object of econometric healthcare cost analysis provide outcome measures sharing these characteristics, their accommodation in the design of the econometric analysis will likely be important. Yet it is equally essential to recognize that the degree to which such analyses must accommodate these features depends on the particular scientific, policy, or decision question(s) at hand. Analysis and decision-making ought to go hand in hand; the next section suggests the role of the decision maker’s perspective in such partnerships.

THE DECISION MAKER’S PERSPECTIVE

Most empirical analyses of healthcare cost data are regression based; that is, analysis considers various features of the distribution of cost outcomes (y) conditional on covariates (x). Although there are many features of such conditional distributions (denoted hereon as $f(y|x)$) that empirical work might explore, in most instances analysis focuses—often without explicit recognition—on the conditional mean $E[y|x]$ and the parameters describing it, eg, $E[y|x] = g(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k)$.

Whether the results of regression analyses of conditional means are informative to any given decision maker is a logically different matter. The specification of the policy- or decision-relevant question should be logically before the analysis. In some instances, such questions are best informed by estimation of conditional mean structures—eg, risk adjustment exercises to forecast total spending in a group heterogeneous in x ’s—but this is not always true. Other circumstances might dictate analysis of alternative conditional parameters (denoted hereon and generally as $g(x)$) that are functionals on $f(y|x)$, and/or their associated partial effects $\frac{\partial g(x)}{\partial x}$.

$g(x)$ as conditional probabilities $\text{Prob}(y \text{ in } S|x)$, where S is some subset of the nonnegative real numbers (eg, $S = \{0\}$, ie, the probability of 0 expenditures given x); or perhaps analysis of conditional quantiles, where $g(x)$ is a vector of conditional quantiles $Q(a|x)$.

These considerations notwithstanding, it is also noteworthy that decision makers are likely to put different weight on considerations like bias, variance, type I versus type II error probabilities, etc, than might analysts. $P < 0.05$ or asymptotic estimator consistency may be core considerations for academic publication but may not be so germane to the decision problem at hand. For research to be useful, consideration of the decision maker’s perspective on such matters ought not to be entirely disregarded in research design, execution, and reporting. This issue is addressed again in section 7.

ECONOMETRIC MODEL SPECIFICATION

General Specification Issues

A central issue in the econometric healthcare cost literature concerns functional form specification for $g(x)$. Once the analyst has determined a particular parameter to estimate (eg, a conditional mean), and assuming for now that $g(x)$ can be aptly described by a linear index functional form in parameters b , $g(x) = g(x; b) = g(b'x)$, the functional form of $g(\cdot)$ becomes the first-order consideration. Focusing for now on conditional means, ie, $g(x) = E[y|x]$, the literature suggests several key considerations.

The first is whether the function $g(\cdot)$ is linear or nonlinear in what is presumed to be a linear index function $b'x$ (a separate issue from whether nonlinear functions of and/or interactions among the basic measures of the x ’s are included as covariates). In some situations—particularly those utilizing very large samples—linear specifications of $g(\cdot)$ have been advocated and deployed usefully even though y is nonnegative.² In other instances, nonlinear specifications—typically log-link, ie, exponential conditional mean (ECM) specification with $E[y|x] = \exp(b'x)$ —are suggested.³ Such specifications enforce the restriction that $E[y|x]$ be positive in any interesting application.

A related consideration is whether $g(x)$ should be specified as a single-index (single-part) or multi-index (multipart) model. If all the component expectations exist, $E[y|x]$ for nonnegative y can always be written as $E[y|x] = \sum_{j=1}^J \text{Prob}(y \in S_j|x) \times E[y|y \in S_j, x]$. The most promi-

nent such specification is the two-part model, $E[y|x] = \Pr \text{ob}(y > 0|x) \times E[y|y > 0, x]$ in which each component is estimated separately, eg, by logistic regression and by GLM, respectively.³ There is, of course, no a priori reason to base the analysis of conditional means on a 2-part or multipart structure even when 0s are present and prominent; after all, the correct functional form is the correct functional form. Consequently, if the main concern of the analysis is with the structure of $E[y|x]$ and/or its partial effects, there may be circumstances when these may be learned more straightforwardly and/or robustly by estimating $E[y|x]$ directly in a single regression.^{3,4}

Covariate Specification

Correct specification of the x vector in whatever regression model is the object of estimation is of obvious importance. In the healthcare cost modeling context, at least 2 considerations seem particularly prominent: interaction effects and endogenous covariates.

Interactions arise frequently in econometric models that describe characteristics of $f(y|x)$. A leading example is a class of risk adjustment models in which interactions among health status measures are featured (eg, CVD and diabetes).² Mullahy and Ai and Norton, explore interpretation of interaction effects—when cast as second-order cross-partial derivatives or differences—in the context of nonlinear $g(x)$ models. In such cases, the nature of interaction effects is more complex than would be indicated simply by consideration of a parameter b_{12} that multiplies an $x_1 \times x_2$ term in the vector.^{5,6} Mullahy, extends this inquiry to consider alternative characterizations of “interaction” in which interaction effects are defined by their economic implications.⁷

Situations where endogenous regressors arise in econometric analysis of healthcare costs are common. In virtually all such instances, the analytical challenges mount and the mutual reliance of theory, econometrics, and policy are unavoidable. For instance, the risk-adjustment literature discusses from a theoretical perspective the merits of risk adjusting reimbursements based on individual behaviors or behavior-related phenomena (eg, obesity, smoking) whose measures would not be defensibly exogenous in health care cost models.^{8,9} Yet whether it is conceptually defensible to risk-adjust payments based on such behaviors is distinct in practice from the econometric obstacles a framework like this might present to analysts charged with estimating such models consistently in the presence of endogeneity (for discussion of applying instrumental variable methods to ECM-type models).^{10,11}

Direct Versus Transformation or Retransformation Regression Strategies

For years, one of the most actively discussed topics in the econometric healthcare cost literature has been the transformation/retransformation problem: analysis based on some nonlinear transformation of the healthcare cost measures followed by linear regression of the transformed measures on x . In some domains, log-transformation of healthcare cost data seems an almost automatic response to observed skewness in the data (considerations of 0 observations on cost

notwithstanding). Veazie et al, consider square-root transformation which naturally accommodates 0 values.¹²

Although results of the transformation regression may sometimes be intrinsically of interest, estimation is typically followed by retransformation to map back to natural units of y . For nonlinear transformation $t(\cdot)$, it is obvious that $t(E[y|x])$ does not equal $E[t(y) | x]$, so that mapping back to natural units involves more than $t^{-1}(\cdot)$ and the estimated parameters.¹³ Moreover, when the distribution of u in the transformation regression $t(y) = b'x + u$ is heteroskedastic in x , retransformation exercises to consistently recover $E[y|x]$ become significantly more complicated.^{3,14} Manning and Mullahy, have suggested strategies to assess when transformation-retransformation is likely to create more problems than it solves.¹⁵ Since there is always available an alternative in which $E[y|x]$ is estimated directly—using either linear regression or nonlinear ECM-based regression methods like GLM—it may be instructive for analysts to undertake such assessments.

GOODNESS OF FIT TESTING

Regardless of what parameter is the object of estimation, an overriding consideration is how well the estimated $g(x)$ “fits” the data. Some prominent tools in the portfolio of GOF tests or diagnostics that may be useful in various healthcare cost modeling contexts include: goodness-of-link tests for conditional mean models;^{16,17} tests for conditional mean or conditional probability models; χ^2 and Kolmogorov-type tests;^{18,19} and p-p or q-q plots²⁰ for estimates of full probability models; and—data permitting—out of sample prediction criteria like MAE and MSE (which criteria relate to different decision-making criteria or loss functions).^{21,22}

Whether the “fit” is defined with respect to in-estimation-sample or out-of-estimation-sample data, there arise ultimately fundamental issues regarding what GOF criteria are germane in any particular application. With modern statistical software, the marginal cost of executing any single GOF test is approximately 0, so GOF tests whose results may confer even small marginal benefit are frequently conducted and reported. Moreover, it is worth noting that with sufficiently large sample sizes it becomes likely that any null parametric model will be rejected by the data.

Recalling section 3, useful GOF testing should be oriented both toward informing analysts about possibly remediable shortcomings of important aspects of estimated models as well as informing decision makers about possible vulnerabilities of decision-relevant features of the estimated econometric models on which decisions are based. The key points are that estimated models that perform acceptably in fitting 1 feature of $f(y|x)$ may not perform well in fitting other features and that not all features of $f(y|x)$ are equally important from the perspective of research consumers.

UPPER TAIL CONSIDERATIONS

As noted in section 2, data on healthcare costs often exhibit “heavy” upper tails. This phenomenon presents several important considerations for analysts.³² First is whether the data are coded accurately. This is obviously a consider-

TABLE 3. Reporting Regression Results

Totexp05	Observed Coefficient	Bootstrap SE	z	P > z	Normal-Based (95% CI)	
Age	0.0338074	0.0007035	48.06	0.000	0.0324287	0.0351862
Sex	0.3150216	0.0382786	8.23	0.000	0.239997	0.3900463
White	0.0999562	0.0358611	2.79	0.005	0.0296697	0.1702426
Schooling	9.71e-06	0.0055323	0.00	0.999	-0.0108335	0.0108529
_Cons	5.978755	0.0746245	80.12	0.000	5.832493	6.125016

glm totexp age sex white schooling, link(log) fam(gam)

ation throughout the distribution of cost outcomes, but its importance regarding potential influence on estimates of parameters like $E[y|x]$ is most obvious at the upper end of the distribution. For instance, the \$43,530,739 observation in the unedited TOTCHG_X series from the 2005 NIS data (Table 2) probably arises from miscoding. Yet without an independent audit, it is not clear how to determine whether the “real” value is one-tenth, one-hundredth, or even 10 times as large. The best one might hope for is that the econometric analyst is also a savvy data anthropologist who will conduct, as appropriate, observation-specific case studies to understand the nature of the data so presented.

Second, heavy upper tails may influence the “robustness” with which some parameters are estimated. Indeed, in worlds described by heavy-tailed Pareto or Burr-Singh-Maddala distributions^{23,24} some traditionally interesting parameters (means, variances) may not even be finite, a situation never encountered in, eg, a normal or log-normal world. Such concerns should translate into empirical strategies that target the high-end parameters of particular interest, eg, models for $\text{Prob}(y \geq k|x)$ or quantile regression models, as well as GOF testing procedures that focus on model fit at the upper end of the distribution (eg, χ^2 tests of fit in upper-end cells for competing models—eg, lognormal vs. Pareto—estimated by ML).

DISCUSSION: REPORTING ANALYSIS

A central theme of the AHRQ-NCI conference that begat the articles in this volume was knowledge transfer: How to deploy the results of quantitative studies of healthcare costs in such a way that they are ultimately useful to decision makers and even to broader communities of research consumers. In addition to topics covered above, a few provisional suggestions are offered here.

First, well-executed graphical depiction of key findings is almost always welcome by readers. Integrated graphics software like Stata’s²⁵ provides straightforward means to translate numerical results into pictures. Koenker and Hallock, is a good example of how well-designed graphs elucidate otherwise-difficult-to-summarize econometric findings.²⁶

Second, most studies in the applied health econometrics literature report results in multi-row tables listing, inter alia, parameter-by-parameter P values against null hypotheses of (typically) 0 restrictions. Reporting results in this manner implies a primary concern about type-I errors. Yet the picture is more complicated. First, until recently, the econometric literature has been largely silent about multiple comparisons

TABLE 4. Reporting Regression Results: An Alternative

Totexp05	Estimated 95% CI	
Age	0.0324287	0.0351862
Sex	0.239997	0.3900463
White	0.0296697	0.1702426
Schooling	-0.0108335	0.0108529
_Cons	5.832493	6.125016

issues although these are obvious in published tables of k-variate regression results. Tools less draconian than Bonferroni adjustments that can handle elegantly multiple comparisons testing situations (many of which have grown out of statistical genetics work studying associations between gene features and disease phenotypes) are readily available; for instance, the literature on false discovery rates^{27,28} provides useful middle-ground strategies for addressing such considerations. An overarching consideration, however, is whether P value reporting is per se a useful way to summarize empirical research²⁹; for an important discussion of this topic.

A related consideration on which the econometric literature is almost equally silent concerns type-II errors. Multiple comparisons issues notwithstanding, just because a single P value does not recommend rejection of a particular H_0 at a standard α -level like 0.05 does not mean that there is no evidence against the null hypothesis in the data. In an exceptional article that, unfortunately, seems largely unknown or at least unappreciated in applied econometrics circles, Andrews, offers a sensible and straightforward approach to treating symmetrically type-I and type-II errors in applied econometric exercises where low power may be a concern.³⁰ If one is serious about using the results of econometric analysis to inform decisions, then Andrews’s approach merits serious attention; it is difficult to imagine a decision maker whose loss function does not place at least some weight on both type-I and type-II errors.

Finally, an even more radical approach would be to rethink entirely the manner in which econometric results are tabulated and reported in published work. For instance, Table 4 displays the results as reported by Stata 9.0 of a simple GLM (log link, gamma family) regression of totexp05 on age, sex, race, and schooling using the 2005 MEPS sample. Current publication standards notwithstanding, how would decisions made on the basis of empirical research ultimately differ if the only results displayed were as in panel B? While much information is discarded, the issue at hand is whether

there is much decision-relevant information that is discarded. The reader is referred to Ziliak and McCloskey's discussion of the work of Gosset (student) in related contexts.²⁹

A less radical proposal is for analysts and decision makers to engage in discussion of whether confidence intervals instead of (or alongside) point estimates and *P* values—or any other different mode of the presentation of research results—might ultimately be more informative for consumers of research (notwithstanding the consideration that *P*-dimension confidence ellipsoids may be more informative for some questions than unidimension confidence intervals). One goal of such an approach would be to force focus on the inherent uncertainty of the empirical findings and on the correspondingly appropriate range of answers to the “what if” questions driving the research in the first place. It would be more than a modest proposal to suggest that we eat our point estimates and report only confidence intervals, but some creative thinking—informed by engagement with research consumers—about how analysts can present research usefully would be timely. In this light, it is useful to close with this reflection.

Social scientists and policymakers alike seem driven to draw sharp conclusions, even when these can be generated only by imposing much stronger assumptions than can be defended. We need to develop a greater tolerance for ambiguity. We must face up to the fact that we cannot answer all of the questions that we ask.³¹

ACKNOWLEDGMENTS

The author thanks 2 anonymous reviewers, Stephanie Robert, and participants at the December 2007 AHRQ-NCI Conference Health Care Costs: Standardized Methods for Research and Policy Applications for helpful comments.

REFERENCES

- Jones AM. “Health econometrics.” In: Culyer AJ, Newhouse JP, eds. *Handbook of Health Economics*. Amsterdam, The Netherlands: Elsevier; 2000.
- Zhao Y, Ellis RP, Ash A, et al. “Measuring population health risks using inpatient diagnoses and outpatient pharmacy data.” *Health Serv Res*. 2001;36(part II):180–193.
- Mullahy J. “Much ado about two: reconsidering retransformation and the two-part model in health econometrics.” *J Health Econ*. 1998;17:247–281.
- Angrist JD. “Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice: reply.” *J Bus Econ Stat*. 2001;19:27–28.
- Mullahy J. “Interaction Effects and Difference-in-Difference Estimation in Loglinear Models.” *NBER TWP 245*. 1999.
- Ai C, Norton EC. “Interaction Terms in Logit and Probit Models.” *Econ Lett*. 2003;80:123–129.
- Mullahy J. “Interaction Effects in Econometric Models.” Working Paper. 2008.
- Schokkaert E, Van de Voorde C. “Risk selection and the specification of the conventional risk adjustment formula.” *J Health Econ*. 2004;23:1237–1259.
- Mullahy J. “Econometric risk adjustment, endogeneity, and extrapolation bias.” *NBER WP 12236*. 2006.
- Windmeijer FAG, Santos Silva JMC. “Endogeneity in count data models: an application to the demand for health care.” *J Appl Econ*. 1997;12:281–294.
- Mullahy J. “Instrumental-variable estimation of count data models: applications to models of cigarette smoking.” *Rev Econ Stat*. 1997;79:586–593.
- Veazie PJ, Manning WG, Kane RL. “Improving risk adjustment for medicare capitated reimbursement using nonlinear models.” *Med Care*. 2003;41:741–752.
- Duan N. “Smearing estimate: a nonparametric retransformation method.” *J Am Stat Assoc*. 1983;78:605–610.
- Manning WG. “The logged dependent variable, heteroskedasticity, and the retransformation problem.” *J Health Econ*. 1998;17:283–295.
- Manning WG, Mullahy J. “Estimating log models: to transform or not to transform?” *J Health Econ*. 2001;20:461–494.
- Pregibon D. “Goodness of link tests for generalized linear models.” *Appl Stat*. 1980;29:15–24.
- Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: Wiley; 1989.
- Andrews DWK. College Station, Tx. “Chi-square diagnostic tests for econometric models: theory.” *Econometrica*. 1988;56:1419–1453.
- Andrews DWK. “A conditional Kolmogorov test.” *Econometrica*. 1997;65:1097–1128.
- Silverberg G, Verspagen B. “The size distribution of innovations revisited: an application of extreme value statistics to citation and value measures of patent significance.” *J Econom*. 2007;139:318–339.
- Basu A, Manning WG, Mullahy J. “Comparing alternative models: log vs. Cox proportional hazard?” *Health Econ*. 2004;13:749–765.
- Manning W, Basu A, Mullahy J. “Generalized modeling approaches to risk adjustment of skewed outcomes data.” *J Health Econ*. 2005;24:465–488.
- Mandelbrot B. “New methods in statistical economics.” *J Polit Econ*. 1963;71:421–440.
- Singh SK, Maddala GS. “A function for size distribution of incomes.” *Econometrica*. 1976;44:963–970.
- Mitchell MN. *A Visual Guide to Stata Graphics*. 2nd ed. Stata Press; 2008.
- Koenker R, Hallock KF. “Quantile regression.” *J Econ Perspect*. 2001;15:143–156.
- Benjamini Y, Hochberg Y. “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *JRSS-B*. 1995;57:289–300.
- Benjamini Y, Yekutieli D. “The control of the false discovery rate in multiple testing under dependency.” *Ann Stat*. 2001;29:1165–1188.
- Ziliak ST, McCloskey DN. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor, MI: University of Michigan Press; 2008.
- Andrews DWK. “Power in econometric applications.” *Econometrica*. 1989;57:1059–1090.
- Manski CF. *Identification Problems in the Social Sciences*. Cambridge, United Kingdom: Harvard University Press; 1995.
- Zhou Y, Ash AS, Haughton J, et al. “Identifying future high cost cases through predictive modeling.” *Dis Manag Health Outcomes*. 2003;11:389–397.