ANNUAL REVIEWS

*Annual Review of Public Health*

# Modeling Health Care Expenditures and Use

Partha Deb[1] and Edward C. Norton[2]

[1]Department of Economics, Hunter College, City University of New York, New York, NY 10065, USA; and National Bureau of Economic Research; email: partha.deb@hunter.cuny.edu

[2]Departments of Health Management and Policy and Economics, University of Michigan, Ann Arbor, Michigan 48109, USA; and National Bureau of Economic Research; email: ecnorton@umich.edu



Annu. Rev. Public Health 2018. 39:489–505

First published as a Review in Advance on January 12, 2018

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

https://doi.org/10.1146/annurev-publhealth-040617-013517



## Keywords

health econometrics, Affordable Care Act, health insurance, expenditures, skewness, two-part models, count models, difference-in-differences, interaction terms, treatment effect

## Abstract

Health care expenditures and use are challenging to model because these dependent variables typically have distributions that are skewed with a large mass at zero. In this article, we describe estimation and interpretation of the effects of a natural experiment using two classes of nonlinear statistical models: one for health care expenditures and the other for counts of health care use. We extend prior analyses to test the effect of the ACA's young adult expansion on three different outcomes: total health care expenditures, office-based visits, and emergency department visits. Modeling the outcomes with a two-part or hurdle model, instead of a single-equation model, reveals that the ACA policy increased the number of office-based visits but decreased emergency department visits and overall spending.

# 1. INTRODUCTION

Health care expenditure and use data typically have two key statistical features. First, their distributions display substantial skewness, manifesting in empirical densities with long, thin right tails. Second, their distributions have a substantial point mass at zero. In modeling such outcomes, especially in the context of natural experiments, one of which we use as the application in this article, it is tempting to ignore the skewness and mass at zero and estimate linear regression models using ordinary least squares (OLS) or weighted least squares (when the data include sampling weights). But the last few decades have seen a proliferation of sophisticated statistical methods that are better suited for outcomes such as health care expenditures and use. Advances in computing power mean that researchers can estimate such complex statistical models faster than ever. Modern statistical software allows researchers to interpret estimates from these models in ways that would not have been feasible in the past. Consequently, we feel that best practice should include a serious investigation of alternative models, without the traditional shackles of ease of computation and interpretation.

In this article, we describe estimation and interpretation of the effects of a natural experiment using two classes of nonlinear statistical models: one for health care expenditures and the other for counts of health care use. For a complete overview of econometric methods for such data, see Deb et al. (14).

We compare estimation and interpretation of the effect of a change in insurance policy on health care expenditures using OLS and a two-part model. The two-part model is based on a statistical decomposition of the density of the outcome into a process that generates zeros and a process that generates positive values. A logit or probit model typically estimates the parameters that determine the threshold between zero and nonzero values of the outcome. In general, alternative specifications of the binary choice model (the first part) yield nearly identical results. However, the choice of model for the distribution of the outcome conditional on it being positive (the second part) is critically important. Different models can yield quite different results. We use a generalized linear model to estimate the parameters that determine positive values. Generalized linear models accommodate skewness in natural ways, give the researcher considerable modeling flexibility, and fit health care expenditures extremely well (9).

Health care use is measured as nonnegative, integer-valued count data. Here, using Poisson, negative binomial, and hurdle models for counts, we describe estimation and interpretation of effects on two counts of health care use: the number of office-based medical practitioner visits and the number of emergency department visits. Whereas the Poisson and negative binomial regressions naturally accommodate zeros, the hurdle model, which is the analog of the two-part model for count data, does so in a more explicit way. See Cameron & Trivedi (10) for a comprehensive discussion of models for count data.

Generally, it is reasonable to expect heterogeneous treatment effects across the distributions of outcomes. In this review, we emphasize one important source of heterogeneity: the one that gives rise to zero versus nonzero expenditures or use. We do so in part because making that distinction in modeling produces better estimates of effects and also because the distinction between extensive margins (zero versus nonzero) and intensive margins (how much if nonzero) is often of substantive policy interest. The two-part and hurdle models explicitly allow for estimation of the extensive and intensive margins separately, along with an overall effect.

We demonstrate these methods in the context of understanding the effects of the young adult health insurance coverage expansion of the Patient Protection and Affordable Care Act (ACA) on health care expenditures and two measures of health care use, i.e., office-based visits and emergency department visits. The young adult health insurance coverage expansion allows dependents to remain on their parents' private health insurance plan until they turn 26 years old. Previously, private insurers often dropped nonstudent dependents at age 19 and student dependents at age 23 (4). This provision took effect on September 23, 2010, which was 6 months after the passage of the ACA on March 23, 2010. In practice, however, it was implemented the following January (in 2011), corresponding to the first open enrollment period following that September date.

This expansion in coverage points to a powerful natural experiment using a difference-in-differences design. Individuals aged 23–25 are considered to be part of the treatment group, whereas those aged 27–29 are considered to be in the control group. We do not use data on 26-year-old individuals because they are in the transition year between being eligible for the ACA provisions and being ineligible. Data from 2008–2010 are used to establish the pre-ACA trends for the difference-in-differences analysis. Data from 2011–2014 are considered to be post-ACA.

Several prior studies have examined the early effects of the ACA's young adult expansion using similar difference-in-differences frameworks and linear statistical models. Amuedo-Dorantes & Yaya (3), Barbaresco et al. (6), Cantor et al. (11), and Sommers et al. (43), among others, document 3–8 percentage-point increases in private health insurance coverage in the treated age group after implementation of the ACA's young adult provision. Barbaresco et al. (6) find significant increases in self-reported health status and regular sources of care and a decrease in the likelihood of foregone needed care. Jhamb et al. (22) estimate a 3% increase in the number of doctor visits. Anderson et al. (4) find that emergency department visits decreased by 40%.

This article makes two important contributions. First, we extend prior analyses to test the effect of the ACA's young adult expansion on three different outcomes—total health care expenditures, office-based visits, and emergency department visits—using nonlinear models that typically fit the distributions of such outcomes better than linear ones. Second, we demonstrate current best econometric practice in modeling these kinds of outcomes, including those that have skewed distributions with a large mass at zero [see also Deb et al. (14)].

## 2. DATA

To answer the question of whether the ACA not only increased health insurance coverage but also affected health care expenditures and use, we need certain data. We need information on a large number of representative young American adults who are younger and older than 26 years and need to observe them in the years before and after the rule change was implemented in 2010. We need accurate measures of health care expenditures and use, as well as detailed measures of health status and other observable characteristics correlated with expenditures and use. The Medical Expenditure Panel Survey (MEPS) (**https://meps.ahrq.gov/mepsweb/**), a national survey on the financing and use of medical care in the United States, has such information. The Agency for Healthcare Research and Quality (AHRQ), a federal government organization in the United States, has collected MEPS data every year since 1996. The data used in these examples are drawn primarily from the Household Component, which contains data on a sample of families and individuals, drawn from a nationally representative subsample of households that participated in the prior year's National Health Interview Survey. AHRQ uses the MEPS to produce annual estimates for a variety of measures of health care expenditures and use, health status, health insurance coverage, and sources of payment for health services in the United States.

The key independent variables for the difference-in-differences analysis are indicators for treatment and control groups and for the pre and post periods. We assume that people aged 23–25 are potentially affected by the ACA policy and are therefore in the treatment group. Those aged 27–29 are in the control group. We eliminate those who are 26 years old because of partial coverage during the year. The three years from 2008 to 2010 are defined as the pre period and the four

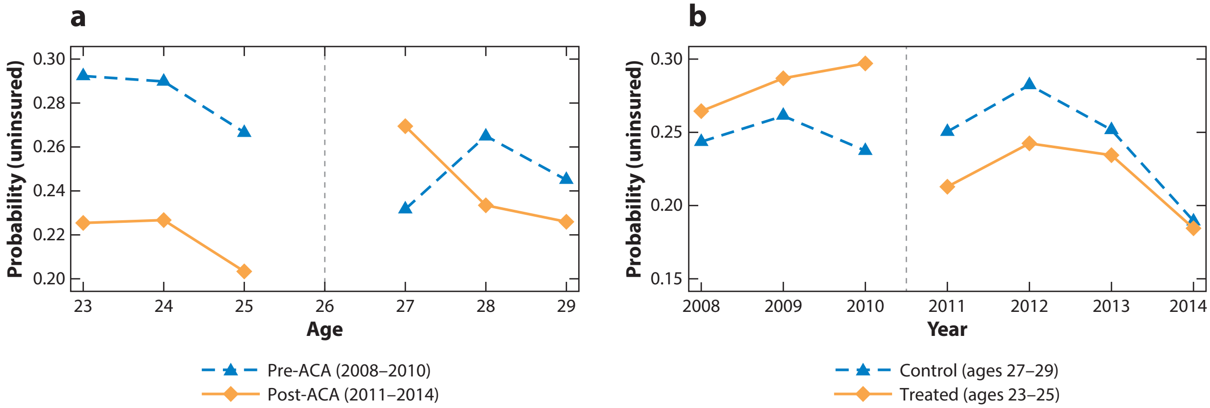


**Figure 1**

The graphs compare the fraction of young adults (above and below age 26) who were uninsured before and after the implementation of the ACA's young adult expansion, first by age and then by year. Authors' own estimates from 2008–2014 MEPS data. Abbreviations: ACA, Affordable Care Act; MEPS, Medical Expenditure Panel Survey.



years from 2011 to 2014 are the post period. We are interested primarily in the effect of the ACA on those up to age 26, that is, in the treatment effect on the treated. Therefore, we can compare those in the treatment group (age 25 and below) to those in the control in the years before and after the policy change. The data are fairly evenly distributed across ages and years.

To help motivate the research questions about the effect of the ACA policy on health care expenditures and use, we first show that the ACA appeared to increase the percentage of young adults who have health insurance. The fraction of people under age 26 who were uninsured dropped after the ACA policy went into effect, whereas the fraction of those above age 26 who were uninsured did not change much (see **Figure 1**). Because of this measurable decrease in the fraction of those just under age 26 who were uninsured, an effect found by others, we would not be surprised if health expenditures and use also increased, owing to the change in insurance on the demand side of the market.

We restrict the MEPS data observations to individuals who were in the scope of the survey design for the entire year and those with valid responses for family size and marital status as basic indicators of data reliability. In addition, we eliminated observations with missing data for two important health status variables (i.e., SF12 physical and mental health scales) because these are important explanatory variables for the analysis of expenditures and use. After dropping the 2,529 observations with missing values for these two variables, the final analytic sample has 17,899 observations.

Because the MEPS sample is drawn using complex survey sampling methods, the use of sampling weights is essential for estimating nationally representative statistics (see 14). Therefore, all our estimates of summary statistics and multivariate models take sampling weights into account.

One of the dependent variables is the total annual health care expenditures, including out-of-pocket payments and third-party payments from all sources. They do not include insurance premiums. Expenditures are measured in nominal US dollars. The distribution of total expenditures is highly skewed with a large mass at zero (see **Figure 2**). More than one-third of observations have zero expenditures, and less than 5% have expenditures in excess of \$9,000. In a very small fraction of observations, 35 to be precise, the expenditure values are greater than \$50,000 (reaching a maximum of \$2,226,997). Although our statistical models are designed to account for skewness,

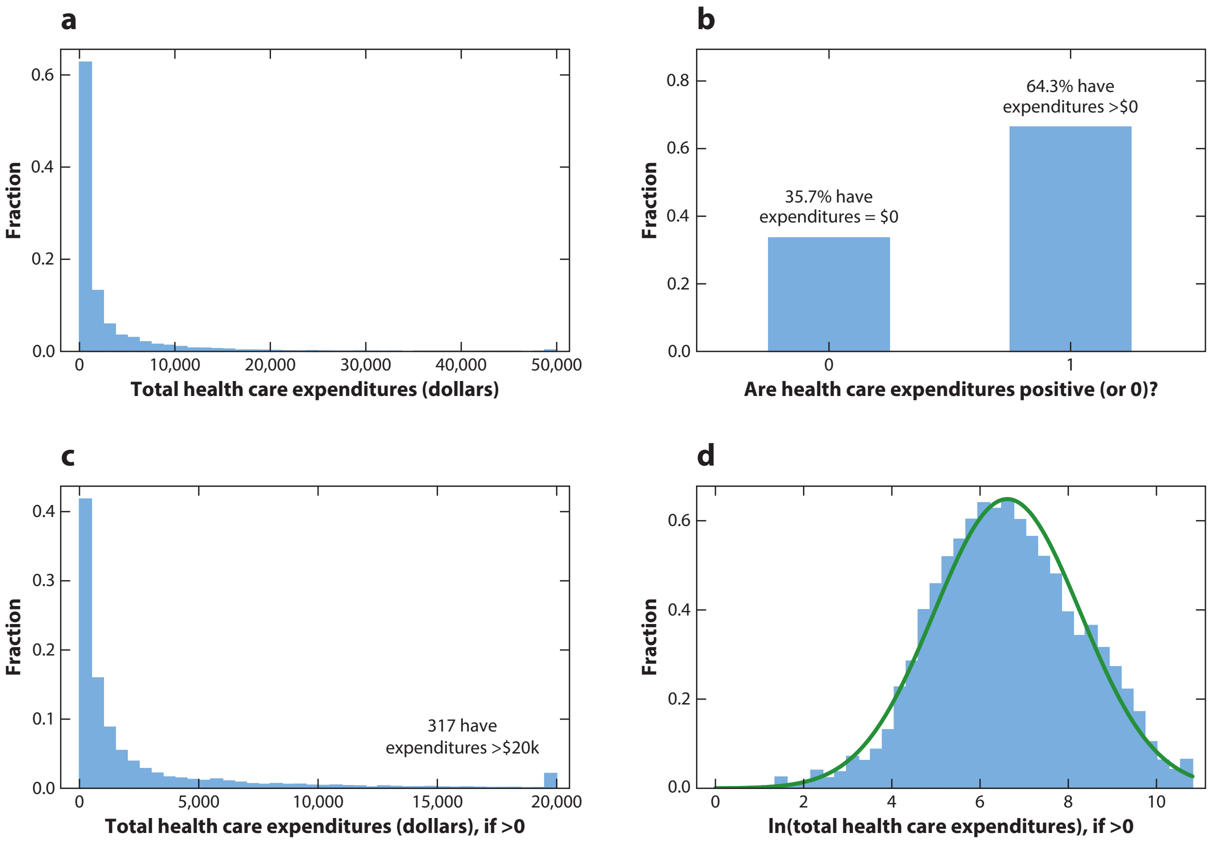


**Figure 2**

The distribution of total health care expenditures is shown four ways: the full histogram including zeros, the split between zeros and nonzeros, the histogram of just positive values, and the histogram of the natural log of positive values. Authors' own estimates from 2008–2014 Medical Expenditure Panel Survey (MEPS) data for individuals ages 23–25 and 27–29.

they are not designed to take extreme values such as these into account. Although it is tempting to drop these observations, we are reluctant to do so because we cannot be sure that they are outliers in any real sense. As a compromise, we topcode the value of the expenditure for each of these 35 observations at $50,000.

The other two dependent variables are counts of office-based medical practitioner visits and emergency department visits. Both office-based visits and emergency department visits have a large fraction of zeros and a declining density (see **Figure 3**). Thirty-four observations have values of office-based visits exceeding 60 (the maximum is 215). We topcode the value of office-based visits for each of these observations at 60. One observation has 22 as the value of emergency department visits. As the next highest value is 12, we topcode the value for this observation at 12. Only a modest number of observations have more than 8 visits, whereas very few observations have more than 3 emergency department visits.

Other covariates control for demographics, education, poverty, and health status. Taking sampling weights into account, about 51% of the sample is female (see **Table 1**). In the control group, 13% are black, whereas in the treated group 15% are black. Just under 20% of the population is Hispanic. Not surprisingly, there is a significant imbalance in marital status across control and

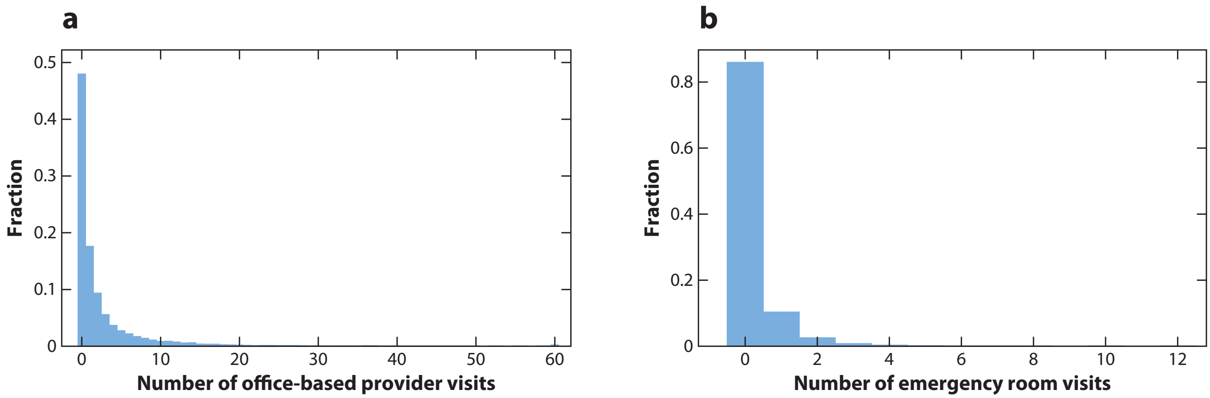


**Figure 3**

The distributions of the two dependent variables measuring health care use: office-based visits and emergency department visits. Authors' own estimates from 2008–2014 Medical Expenditure Panel Survey (MEPS) data for individuals ages 23–25 and 27–29.

treated groups. About 43% of individuals in the (older) control group are married, and about 7% are widowed, separated, or divorced. In the (younger) treated group, these rates are 22% and 3%, respectively. The poverty category is on a scale of 1 (poorest) to 5 (richest). Physical and mental health statuses are derived from SF12 measures to be scaled from 0 to 100 with means of about 50 in the overall population (44). A higher number indicates better health. In our sample, the weighted means of the physical and mental health scales are about 54 and 51, respectively. They are quite balanced across treated and control groups. Both distributions are skewed left, with a median three to four points above the mean. There are dummy control variables for arthritis, asthma, cancer, high cholesterol, diabetes, high blood pressure, and the presence of any of five health conditions that are rare for this age group (angina, congestive heart disease, emphysema, myocardial infarction, and stroke).

## 3. EXPENDITURE MODELS

Modeling health care expenditures usually has several challenges related to the distribution of the dependent variable. Health expenditure data, for those with any health care use, are generally extremely skewed. In the United States, a small fraction of the population accounts for a substantial fraction of total expenditures. Berk & Monheit (8) report that 5% of the population accounts for the majority of health expenditures and that the severely right-skewed concentration of health care expenditures has remained stable over decades. In this analysis, the dependent variable has highly skewed positive values and may be heteroskedastic.

Although one could potentially use OLS to model skewed positive values, there are better alternatives [see Deb et al. (14)]. For highly skewed data, generalized linear models (GLM) offer a range of alternative functional forms to match the relationship between the expected value of the dependent variable and the linear index of covariates. GLMs are more general than ordinary linear regression models (28). The GLM generalizes the ordinary linear regression model by allowing the expectation of the outcome variable to be a function (known as the link function) of the linear index of covariates, not simply a linear function of the index. Expenditure data, for example, often fit best with a log link, meaning that the natural logarithm of the expected value of the dependent variable is modeled as the linear index. We check the log link against several other functional form alternatives.

**Table 1 Summary statistics for the Medical Expenditure Panel Survey data. Means of the dependent and independent variables used in the analyses, stratified by control group (ages 27–29) and treatment group (ages 23–25). Values are rates unless otherwise specified**

| Variables | Ages 27–29 | Ages 23–25 |
|---|---|---|
| Total health care expenditures (US$) | 2,193.7 | 1,951.6 |
| Number of office-based provider visits | 3.339 | 2.984 |
| Number of emergency department visits | 0.186 | 0.206 |
| Age (in years) | 28.01 | 24.00 |
| Female | 0.510 | 0.510 |
| Black | 0.132 | 0.147 |
| Hispanic | 0.197 | 0.199 |
| Married | 0.434 | 0.216 |
| Widowed, separated, divorced | 0.066 | 0.026 |
| High school education | 0.497 | 0.638 |
| Some college education | 0.248 | 0.177 |
| College graduate | 0.082 | 0.013 |
| Poverty category: Poor | 0.143 | 0.186 |
| Poverty category: Near poor | 0.046 | 0.055 |
| Poverty category: Low income | 0.149 | 0.169 |
| Poverty category: Middle income | 0.333 | 0.335 |
| Poverty category: High income | 0.328 | 0.255 |
| Physical health scale (0–100) | 53.93 | 54.30 |
| Mental health scale (0–100) | 50.95 | 51.34 |
| Arthritis | 0.052 | 0.033 |
| Asthma | 0.103 | 0.098 |
| Cancer | 0.020 | 0.014 |
| High cholesterol | 0.071 | 0.042 |
| Diabetes | 0.012 | 0.006 |
| High blood pressure | 0.099 | 0.071 |
| Rare diseases | 0.012 | 0.011 |
| *Sample size N* | 8,983 | 8,916 |

In addition, GLMs also explicitly model the heteroskedasticity. GLMs allow the variance of the outcome to be a function of its predicted value by the choice of an appropriate distribution family. Health economists have been increasingly interested in applying GLMs to health care expenditures and costs. The work of Mullahy (30) and that of Blough et al. (9) were among the first applications in health economics.

The other main modeling challenge is the large fraction of observations with zero expenditures. The health econometrics literature has settled on the two-part model as the best way to model a dependent variable with a large mass at zero and many positive values (7). Therefore, we first model the probability that a person has any health care expenditures with a logit model using the full sample. Then we estimate a GLM on the subset of people who have any expenditures. The two-part model allows for separate investigation of the effect of covariates on the extensive margin (logit model, if any expenditures) and on the intensive margin (GLM, amount of expenditures if any).

The two-part model has a long history in empirical analysis (7). Newhouse & Phelps (31) published a paper that is the first known example of the two-part model in health economics. Their empirical model estimated price and income elasticities of medical care. The two-part model became widely used in health economics and health services research after a team at the RAND Corporation used it, and the related four-part model, to model health care expenditures in the context of the Health Insurance Experiment (15). More recently, the two-part model with a logit or probit first part and a GLM second part has been used in a variety of empirical work in health services research (12, 16, 26). See Mihaylova et al. (29) for more on the widespread use of the two-part model for health care expenditure data.

There are four main modeling choices in two-part models. The first is between logit and probit for the first part of the two-part model. This choice is generally innocuous in that there is never a substantial difference between logit and probit. The second and third choices are for the link function and distribution family for the GLM. We use standard specification tests to make these choices because, with different data sets and dependent variables, different link functions and distribution families are most appropriate. The fourth choice is about the specification of the linear index: in particular, whether to include interaction terms and whether to allow flexible nonlinear specification of continuous variables. The general practice is that any variable (including interactions terms and higher-order terms) that is in either the first-part or the second-part model will be in both. No variables are included in one part but excluded from the other.

## 3.1. The Treatment Effect

In nonlinear models, the interpretation of the interaction effect of two variables—such as between treatment and the implementation of the ACA as in this difference-in-differences study design—is complicated. Ai & Norton (1) showed that the full interaction effect, calculated by taking the double derivative with respect to both interacted variables, is not equal to the marginal effect of the change in just the interaction term. The full interaction effect can even be the opposite sign of the coefficient on the interaction term.

However, Puhani (36) argued, using the potential outcomes framework, that the treatment effect on the treated in the difference-in-difference regression equals the expected value of the dependent variable for the treatment group in the post period with treatment compared with the hypothetical expected value of the dependent variable for the treatment group in the post period if they had not received treatment. In nonlinear models, the treatment effect on the treated equals the difference in two predicted values. It always has the same sign as the coefficient on the interaction term. Because we estimate many nonlinear models using a difference-in-differences study design, we report the treatment effect on the treated in all tables of results.

## 3.2. Specification Tests for GLM

For the first choice, we arbitrarily choose the logit model. If, instead, we had chosen the probit, the results in terms of marginal effects would be virtually identical. Although the probit has estimated coefficients that appear quite different from the logit, the predicted values and marginal effects are essentially the same (32).

The other modeling choices are more consequential (14). The GLM requires choosing both a link function and a distribution family. The link function relates the expected value of the dependent variable to the linear index of covariates and coefficients. For example, one choice is the linear link so that the expected value of the dependent variable is the linear index. However, with financial data such as health care expenditures, the appropriate link is often the natural logarithm. That is, the log of the expected value of expenditures is the linear index. Other link functions are possible, including the square root.

We use a Box-Cox test to see what power function will transform the dependent variable health care expenditures to be closest to symmetric. In brief, the Box-Cox approach tests which scalar power, $\delta$, of the dependent variable, $y^{\delta}$, results in the most symmetric distribution. A power of $\delta = 1$ corresponds to a linear model, $\delta = 0.5$ corresponds to the square root transformation, and $\delta \to 0$ corresponds to the natural log transformation model (14). We do the Box-Cox test two ways, with and without controlling for covariates, in both cases, limiting the test to observations with positive values. It turns out that the results are not appreciably different. In both tests, the estimated coefficient is close to zero, corresponding to the natural log transformation. These results are not surprising, given the near symmetric distribution of the log of health care expenditures (shown in the lower-right part of **Figure 2**).

The next specification test is to determine the distribution family, that is, the relationship between the mean and the variance. GLMs allow for heteroskedasticity in which the variance is not a constant for all observations but is a function of the mean. A higher variance makes intuitive sense for health care expenditures that have a higher expected value. For example, the variance could be proportional to the square or the cube of the mean.

We use a modified Park test (33), which empirically tests the relationship between the mean and the variance. The test is conducted after running a GLM (in which we use the log link and the gamma distribution). We compute the expected value (mean) for each observation, conditional on the covariates. We compute the squared error (variance) for each observation. The regression of the logarithm of the squared error on the expected value provides the test. One should use the Gaussian distribution in the GLM when the coefficient on the expected value is close to 0.0 because the variance is unrelated to the mean. One should use a Poisson-type distribution, the Gamma distribution, or the inverse-Gaussian distribution when the coefficient is close to 1.0, close to 2.0, or close to 3.0, respectively. For our sample, we observed an estimated coefficient of 1.83. In summary, the specification tests supported the use of the log link and the gamma distribution.

Finally, we test the specification of the explanatory variables. Although we know that we want to control for demographics, education, and health, the preferred functional form of these variables in the regression specifications is not known. Our main specification includes numerous controls for these important variables. However, interactions between these variables and higher-order terms of continuous covariates may greatly improve model fit. There are two ways to test this possibility. One way is to conduct Pregibon's link test (35) and Ramsey's regression equation specification error test (RESET) (37). These tests are typically done in OLS models, but modified versions can be done in GLM models. The logic of these tests is to regress the dependent variable on the predicted value and powers of the predicted value. This test could reveal whether there are important omitted variables that are correlated with higher-order terms (14).

Pregibon's link test assesses whether the coefficient on the squared term is significantly different from zero. Ramsey's RESET is a joint test of the squared, cubed, and fourth-order terms. Neither test is significant at the traditional 5% level.

However, Pregibon's and Ramsey's tests are suggestive but may miss important nonlinearities. For example, nearly all the covariates are binary. Including higher-order terms for binary variables is not possible. The physical and mental health variables, in contrast, are continuous variables and could, in principle, have nonlinear effects on health care expenditures. Therefore, we ran an additional model with squared terms and the interaction between mental and physical health. These tests also did not reveal significant effects of these squared and interaction terms. Consequently, we did not add any more variables to the model specification.

**Table 2** Results from OLS and two-part models for total health care expenditures[a]

| Variables | OLS[b] | Two-part model[b] Logit | GLM | Overall |
|---|---|---|---|---|
| | **Coefficients[c]** | | | |
| Treated | 365.3** (93.0) | 0.184** (0.091) | 0.044 (0.029) | |
| ACA | 294.2** (78.6) | 0.037 (0.101) | 0.121* (0.066) | |
| Treated × ACA | −323.4*** (76.0) | −0.061 (0.108) | −0.115* (0.066) | |
| Physical health scale | −166.0*** (5.8) | −0.032*** (0.006) | −0.048*** (0.003) | |
| Mental health scale | −45.9*** (8.9) | −0.028*** (0.004) | −0.018*** (0.004) | |
| | **Effects among treated in ACA period[d]** | | | |
| Treatment effect | −323.4*** (76.0) | −0.010 (0.018) | −321.4* (191.2) | −245.6* (139.9) |
| Physical health scale | −166.0*** (5.8) | −0.005*** (0.001) | −126.9*** (11.4) | −98.7*** (8.5) |
| Mental health scale | −45.9*** (8.9) | −0.005*** (0.001) | −47.1*** (11.7) | −42.9*** (8.4) |
| *N* | 17899 | 17899 | 11885 | 17899 |

Abbreviations: ACA, Patient Protection and Affordable Care Act; GLM, generalized linear models; OLS, ordinary least squares.

[a]All models control for indicators for age and year, female, black, Hispanic, marital status, education levels, family poverty levels, and a number of health conditions.

[b]***, statistical significance at the 1% level; **, statistical significance at the 5% level; *, statistical significance at the 10% level.

[c]Shows the coefficients and cluster-robust standard errors.

[d]Shows marginal effects, including the combined marginal effects from both parts of the two-part model and standard errors.

We used Stata's twopm command to estimate the two-part model (7). The Stata code and the data set are available upon request so that readers can reproduce our results.

## 3.3. Results

We used Stata to estimate OLS, logit, and GLM models to obtain parameter estimates and marginal effects for two-part expenditure models using methods described in Deb et al. (14). The estimated coefficients for a few key variables and the associated cluster-robust standard errors are shown in **Table 2**. The simple OLS model implies that the treatment effect is a reduction in spending of about $320 per person ages 23–25. Moving to the two-part model, the logit indicates that, in the pre period, those in the treated group (ages 23–25) are more likely to have at least some spending. Among those who spend something, the GLM model indicates that there is an increase in spending after implementation of the ACA for the control group, although this result is statistically significant at only the 10% level.

For the overall treatment effect combining both parts of the two-part model, we calculated the treatment effect on the treated (see **Table 2**, Overall column, first row). The number −$246 means that the effect of the ACA on expenditures for the target population was a reduction in health care spending by about $250, statistically significant at the 10% level. This result is considerably smaller, and less significant, than the OLS estimate. The two-part model indicates that most of the effect was due to a reduction in spending by those who spend something, not due to a reduction in the probability of spending.

**Table 2** also shows estimates and marginal effects for two important continuous covariates, the physical and mental health scores. The results show that young adults who are in better health spend significantly less than those in poorer health. The two-part model results show that they are both less likely to spend and to spend less when they do spend.

The other explanatory variables generally have effects in the expected directions (results not reported in the tables). Women spend more than men. Blacks and Hispanics spend less money than do whites and non-Hispanics. Married people spend more than nonmarried people. Higher education is associated with more spending, with less than high school education as the omitted category. However, it is difficult to have a causal interpretation of any of these variables because the models do not control for health insurance, which is likely an important omitted confounder.

## 4. COUNT MODELS

The number of office-based medical practitioner visits and the number of emergency department visits are measured as nonnegative integers or count variables. Both have distributions that place probability mass only at nonnegative integer values and are severely skewed, are intrinsically heteroskedastic, and have variances that increase with the mean. For both, the observations are concentrated on a few small discrete values, typically zero and a few small positive integers, although the right tail for office-based visits goes out a long way.

If one is interested only in the prediction of the conditional mean or in the response of the conditional mean to a covariate, it may be tempting to ignore the discreteness and skewness and simply estimate the responses of interest using linear or generalized linear regression methods. However, models that ignore discreteness can be quite inefficient, leading to substantial losses in statistical power (24). Equally important is the consideration that, in the case of discrete data, substantive interest may lie in the estimation of event probabilities. In these situations, formal estimation of a count data process is essential. Regression models for count data are comprehensively described in Cameron & Trivedi (10), Hardin & Hilbe (21), and Winkelmann (46). Deb et al. (14) describe a variety of count data models with a specific focus on measures of health care use.

We begin our discussion of regression models for count data with the Poisson regression model. It is the canonical regression model for count data and should be the starting point of any analysis. The Poisson distribution is a member of the linear exponential family and thus has a powerful robustness property, indicating that its parameters are consistently estimated as long as the conditional mean is specified correctly, even if the true data-generating process is not Poisson (19, 20). This robustness comes at an efficiency cost, however (10). Next, we discuss the negative binomial regression model, which is the canonical model for overdispersed count data. We contrast results obtained from negative binomial regressions with those obtained from Poisson regressions. The negative binomial regression model relaxes the restrictive mean-variance property of the Poisson regression and thus can be substantially more efficient.

Finally, we estimate hurdle models for the count variable outcomes. The hurdle model is the count-data analog of the two-part model. It is often motivated as arising from a principal–agent mechanism (34). Such justification is not required, however. The hurdle model is well justified on statistical grounds.

Pohlmeier & Ulrich (34) and Gerdtham & Trivedi (18) provided early papers that used Poisson, negative binomial, and hurdle models for counts of health care use. More recently, various count data regression models have been applied to a variety of policy and descriptive studies of determinants of health care use (5, 17, 25, 39, 40, 45).

## 4.1. The Treatment Effect

For nonlinear count models, the interpretation of the interaction term between treatment and the implementation of the ACA requires the same care as with nonlinear expenditures models (1, 36). As explained above, the treatment effect on the treated equals the expected value of the dependent variable for the treatment group in the post period with treatment compared with the hypothetical expected value of the dependent variable for the treatment group in the post period if they had not received treatment. We report this treatment effect on the treated in all tables of results.

## 4.2. Specification Tests and Model Selection

As with models for health care expenditures, it is important to test the specification of the explanatory variables because the exact functional form is not known, even in the context of a natural experiment in which the primary focus is on a treatment variable. Interactions and polynomials of covariates may improve the model fit substantially. We use Pregibon's link test and Ramsey's RESET test in the context of Poisson regression models to determine that there are no substantial gains from nonlinear functions of covariates given our data. Note that this finding should not be seen as a general statement of preference for linear index specifications. Almost all the covariates in our specification are binary. In addition, we have a full set of age and year indicators in our specification, which limit the scope of the power of other covariates considerably.

We use the Akaike Information Criterion (AIC) (2) and the Bayesian Information Criterion (BIC), also known as the Schwarz Bayesian Criterion (SBC) (41), to evaluate the performance of Poisson, negative binomial, and hurdle count-data models. These criteria have been shown, in a variety of circumstances, to have desirable properties, including robustness to model misspecification and to when the data have additional statistical issues, such as clustering and weighting (23, 27, 42).

## 4.3. Results

We used Stata to estimate Poisson and negative binomial models and to obtain parameter estimates and marginal effects for hurdle models using methods described in Deb et al. (14). Estimated coefficients, associated cluster-robust standard errors, and measures of model fit of models for office-based visits and emergency department visits are shown in **Tables 3** and **4**, respectively.

The results for the models of office-based visits shown in **Table 3** suggest that the negative binomial-based models are substantially better than the Poisson-based models. Evidence also favors the hurdle negative binomial model relative to the standard one. Thus, there is evidence of heterogeneity in extensive versus intensive margins of the decision-making process.

The single-equation Poisson and negative binomial count models imply that the treatment effect is an increase in office-based visits by roughly one-quarter to one-third of a visit per person ages 23–25 (see **Table 3**). Moving to the hurdle model, the logit indicates that there is no difference in the treated and the control groups in their likelihood of having any office-based visits. Nor is there a difference over time in the control group. However, for those who have any office-based

**Table 3** **Results for count models for number of office-based visits[a]**

| Variables | Poisson[b] | NB[b] | Hurdle model[b] | | |
|---|---|---|---|---|---|
| | | | Logit | Truncated negative binomial | Overall |
| | **Coefficients[c]** | | | | |
| Treated | 0.015 (0.053) | −0.048 (0.051) | 0.073 (0.045) | −0.135** (0.058) | |
| ACA | −0.002 (0.031) | 0.016 (0.024) | −0.021 (0.024) | −0.003 (0.016) | |
| Treated × ACA | 0.082** (0.040) | 0.126*** (0.032) | 0.034 (0.035) | 0.199*** (0.021) | |
| Physical health scale | −0.037*** (0.004) | −0.043*** (0.004) | −0.032*** (0.005) | −0.047*** (0.006) | |
| Mental health scale | −0.022*** (0.003) | −0.032*** (0.003) | −0.027*** (0.003) | −0.032*** (0.004) | |
| | **Effects among treated in ACA period[d]** | | | | |
| Treatment effect | 0.239** (0.113) | 0.370*** (0.087) | 0.007 (0.014) | 0.363*** (0.122) | 0.427*** (0.059) |
| Physical health scale | −0.111*** (0.012) | −0.134*** (0.016) | −0.007*** (0.001) | −0.095*** (0.009) | −0.132*** (0.016) |
| Mental health scale | −0.067*** (0.009) | −0.099*** (0.011) | −0.006*** (0.000) | −0.064*** (0.006) | −0.094*** (0.011) |
| *N* | 17899 | 17899 | 17899 | 9316 | 17899 |
| K | 44 | 45 | 44 | 45 | 89 |
| BIC | 134027.7 | 73717.5 | | | 72964.7 |

Abbreviations: ACA, Patient Protection and Affordable Care Act; BIC, Bayesian information criterion; K, number of parameters in the model.
[a]All models also control for indicators for age and year, female, black, Hispanic, marital status, education levels, family poverty levels, and a number of health conditions.
[b]***, statistical significance at the 1% level; **, statistical significance at the 5% level.
[c]Shows the coefficients and cluster-robust standard errors.
[d]Shows marginal effects, including the combined marginal effects from both parts of the two-part model and standard errors.

visits, the treated group has fewer visits in the pre period, and there is a large positive coefficient on the interaction term, statistically significant at the 5% level.

For the overall treatment effect combining both parts of the hurdle model, we calculated the treatment effect on the treated (see **Table 3**, Overall column, first row). The number 0.427 means that the effect of the ACA on office-based visits for the target population was an increase in visits by almost half a visit, statistically significant at the 1% level. This result is larger in magnitude and has a smaller *p*-value than the results from the single-equation models. The hurdle model indicates that most of the effect of the ACA on office-based visits was due to an increase in visits for those who had any visits.

We repeat this exercise for the count measure of emergency department visits (see **Table 4**). Relative to office-based visits, this distribution has a much smaller domain and is not substantially overdispersed. The hurdle negative binomial is better than the hurdle Poisson, which is better than the standard Poisson. The standard negative binomial regression beats them all, although the differences in BIC are not nearly as pronounced as they were for the models of office-based visits.

**Table 4 Results for count models for emergency department visits[a]**

| Variables | Poisson[b] | NB[b] | Hurdle model[b] | | |
|---|---|---|---|---|---|
| | | | Logit | Truncated negative binomial | Overall |
| | Coefficients[c] | | | | |
| Treated | 0.261*** (0.067) | 0.285*** (0.055) | 0.340*** (0.074) | 0.022 (0.143) | |
| ACA | 0.080* (0.046) | 0.077 (0.056) | 0.053 (0.099) | 0.187 (0.136) | |
| Treated × ACA | −0.171** (0.067) | −0.206*** (0.067) | −0.218** (0.101) | −0.109 (0.202) | |
| Physical health scale | −0.043*** (0.005) | −0.047*** (0.005) | −0.051*** (0.006) | −0.028*** (0.007) | |
| Mental health scale | −0.021*** (0.001) | −0.023*** (0.001) | −0.023*** (0.003) | −0.016** (0.006) | |
| | Effects among treated in ACA period[d] | | | | |
| Treatment effect | −0.037** (0.016) | −0.045*** (0.016) | −0.026** (0.011) | −0.026 (0.044) | −0.043*** (0.014) |
| Physical health scale | −0.009*** (0.001) | −0.009*** (0.001) | −0.006*** (0.000) | −0.006*** (0.002) | −0.009*** (0.001) |
| Mental health scale | −0.004*** (0.000) | −0.004*** (0.000) | −0.003*** (0.000) | −0.004** (0.001) | −0.004*** (0.000) |
| *N* | 17899 | 17899 | 17899 | 2522 | 17899 |
| K | 44 | 45 | 44 | 45 | 89 |
| BIC | 18443.9 | 17548.3 | | | 17839.5 |

Abbreviations: ACA, Patient Protection and Affordable Care Act; BIC, Bayesian information criterion; K, number of parameters in the model.

[a]All models also control for indicators for age and year, female, black, Hispanic, marital status, education levels, family poverty levels and a number of health conditions.

[b]***, statistical significance at the 1% level; **, statistical significance at the 5% level; *, statistical significance at the 10% level.

[c]Shows the coefficients and cluster-robust standard errors.

[d]Shows marginal effects, including the combined marginal effects from both parts of the two-part model and standard errors.

For emergency department visits, the single-equation models show a substantial small but statistically significant treatment effect measured as a decline in emergency department use. The logit model shows substantial baseline differences in the treated and control groups in the probability of individuals going to the emergency department. The younger treatment group is more likely to go to the emergency department prior to the enactment of the ACA. The negative coefficient on the interaction term in the logit model implies that there may be a negative treatment effect. However, the second part of the hurdle count model finds no effect of any of the three policy variables.

The overall treatment effect on the treated is negative and statistically significant at the 1% level. There is a reduction of about 0.04 in the amount of emergency department use for young adults after ACA implementation. This is a reduction by about 1 visit for every 23 young adults. The hurdle model shows that this reduction is due almost entirely to reducing the probability of having any visits, not reducing the number of visits for the few people who ever go to the emergency department.

**Tables 3** and **4** also show estimates and marginal effects for the physical and mental health scores. The results show that young adults in better health are less likely to visit an office-based practitioner, they are less likely to go to the emergency department, and, when they do use such care, they use less of it. The other explanatory variables generally produce effects in the expected directions (results not reported in the tables).

## 5. CONCLUSIONS

This article describes methods for estimating models of health care expenditures and use that take researchers beyond linear regression methods. GLMs, two-part models, Poisson regressions, negative binomial regressions, and hurdle models are shown to be superior to linear regression methods in a large body of work [see also Deb et al. (14)]. We view them as best practice methods for such outcomes.

Using these methods, we find that the ACA young adult expansion lowered health care expenditures, increased office-based visits, and decreased emergency department visits. Modeling the large mass of zeros, through two-part and hurdle count models, greatly improves the fit of the models and allows for better understanding of the results.

We encourage researchers to use a battery of specification checks and model selection criteria to narrow down the model specification along the following dimensions: specification of covariates, functional relationship between the outcome and covariates, and statistical distributions for the outcome (or error term, as appropriate). We view this approach as a critical component of a good empirical analysis (13, 23, 38).

### SUMMARY POINTS

1. This article describes best practice methods for modeling health care expenditures and counts of use.
2. Modeling the large mass of zeros, through two-part and hurdle count models, often greatly improves the fit of the models and allows for better understanding of the results.
3. Model checks are important for choosing the best model.
4. The ACA young adult expansion lowered health care expenditures, increased office-based visits, and decreased emergency department visits.
5. The single-equation OLS and Poisson models cannot reveal what the two-part and hurdle models show: The changes in expenditures and office-based visits were on the intensive margin, but the change in emergency department visits was due only to the change in the probability of an emergency department visit.

### FUTURE ISSUES

1. Over the coming years, additional years of data will allow researchers to test for the long-run effect of the ACA.
2. As the range of statistical models available for health care expenditures and use continues to grow, it is important to understand their strengths and limitations. Statistical tests can help choose which model is best for the particular data set and research question.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We gratefully acknowledge years of insights and mentoring from Willard G. Manning.

## LITERATURE CITED

1. Ai C, Norton EC. 2003. Interaction terms in logit and probit models. *Econ. Lett.* 80:123–29
2. Akaike H. 1970. Statistical predictor identification. *Ann. Inst. Stat. Math.* 22:203–17
3. Amuedo-Dorantes C, Yaya ME. 2016. The impact of the ACA's extension of coverage to dependents on young adults' access to care and prescription drugs. *South. Econ. J.* 83:25–44
4. Anderson M, Dobkin C, Gross T. 2012. The effect of health insurance coverage on the use of medical services. *Am. Econ. J. Econ. Policy* 4:1–27
5. Antón JI, Muñoz de Bustillo R. 2010. Health care utilisation and immigration in Spain. *Eur. J. Health Econ.* 11:487–98
6. Barbaresco S, Courtemanche CJ, Qi Y. 2015. Impacts of the Affordable Care Act dependent coverage provision on health-related outcomes of young adults. *J. Health Econ.* 40:54–68
7. Belotti F, Deb P, Manning WG, Norton EC. 2015. twopm: two-part models. *Stata J.* 15:3–20
8. Berk ML, Monheit AC. 2001. The concentration of health care expenditures, revisited. *Health Aff.* 20:9–18
9. Blough DK, Madden CW, Hornbrook MC. 1999. Modeling risk using generalized linear models. *J. Health Econ.* 18:153–71
10. Cameron AC, Trivedi PK. 2013. *Regression Analysis of Count Data*. Cambridge, UK: Cambridge Univ. Press
11. Cantor JC, Monheit AC, DeLia D, Lloyd K. 2012. Early impact of the Affordable Care Act on health insurance coverage of young adults. *Health Serv. Res.* 47:1773–90
12. Cawley J, Meyerhoefer C. 2012. The medical care costs of obesity: an instrumental variables approach. *J. Health Econ.* 31:219–30
13. Claeskens G, Hjort NL. 2008. *Model Selection and Model Averaging*. Cambridge, UK: Cambridge Univ. Press
14. Deb P, Norton EC, Manning WG. 2017. *Health Econometrics Using Stata*. College Station, TX: Stata Press
15. Duan N, Manning WG Jr., Morris CN, Newhouse JP. 1984. Choosing between the sample-selection model and the multi-part model. *J. Bus. Econ. Stat.* 2:283–89
16. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. 2009. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Aff.* 28:w822–31
17. Gerdtham UG. 1997. Equity in health care utilization: further tests based on hurdle models and Swedish micro data. *Health Econ.* 6:303–19
18. Gerdtham UG, Trivedi PK. 2001. Equity in Swedish health care reconsidered: new results based on the finite mixture model. *Health Econ.* 10:565–72
19. Gourieroux C, Monfort A, Trognon A. 1984. Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica* 52:701–20
20. Gourieroux C, Monfort A, Trognon A. 1984. Pseudo maximum likelihood methods: theory. *Econometrica* 52:681–700
21. Hardin JW, Hilbe JM. 2007. *Generalized Linear Models and Extensions*. College Station, TX: Stata Press
22. Jhamb J, Dave D, Colman G. 2015. The Patient Protection and Affordable Care Act and the utilization of health care services among young adults. *Int. J. Health Econ. Dev.* 1:8–25
23. Kadane JB, Lazar NA. 2004. Methods and criteria for model selection. *J. Am. Stat. Assoc.* 99:279–90
24. King G. 1988. Statistical models for political science event counts: bias in conventional procedures and evidence for the exponential Poisson regression model. *Am. J. Pol. Sci.* 32:838–63

25. Lahiri K, Xing G. 2004. An econometric analysis of veterans' health care utilization using twopart models. *Empir. Econ.* 29:431–49
26. Lê Cook B, McGuire TG, Lock K, Zaslavsky AM. 2010. Comparing methods of racial and ethnic disparities measurement across different settings of mental health care. *Health Serv. Res.* 45:825–47
27. Leroux BG. 1992. Consistent estimation of a mixing distribution. *Ann. Stat.* 20:1350–60
28. McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Boca Raton, FL: CRC Press. 2nd ed.
29. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. 2011. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 20:897–916
30. Mullahy J. 1998. Much ado about two: reconsidering retransformation and the two-part model in health economics. *J. Health Econ.* 17(3):247–81
31. Newhouse JP, Phelps CE. 1976. New estimates of price and income elasticities of medical care services. In *The Role of Health Insurance in the Health Services Sector*, ed. RN Rosett, pp. 261–320. Cambridge, MA: Natl. Bur. Econ. Res.
32. Norton EC, Dowd BE. 2017. Log odds and the interpretation of logit models. *Health Serv. Res.* **https://doi.org/10.1111/1475-6773.12712**. In press
33. Park RE. 1966. Estimation with heteroscedastic error terms. *Econometrica* 34:888
34. Pohlmeier W, Ulrich V. 1995. An econometric model of the two-part decisionmaking process in the demand for health care. *J. Hum. Resour.* 30:339–61
35. Pregibon D. 1980. Goodness of link tests for generalized linear models. *J. R. Stat. Soc. C* 29:15–23
36. Puhani PA. 2012. The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. *Econ. Lett.* 115:85–87
37. Ramsey JB. 1969. Tests for specification errors in classical linear least-squares regression analysis. *J. R. Stat. Soc. Ser. B* 31:350–71
38. Rao CR, Wu Y, Konishi S, Mukerjee R. 2001. On model selection. *Lect. Notes-Monogr. Ser.* 38:1–64
39. Sarma S, Simpson W. 2006. A microeconometric analysis of Canadian health care utilization. *Health Econ*. 15:219–39
40. Schmid C. 2015. Consumer health information and the demand for physician visits. *Health Econ*. 24:1619–31
41. Schwarz G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–64
42. Sin C-Y, White H. 1996. Information criteria for selecting possibly misspecified parametric models. *J. Econom.* 71:207–25
43. Sommers BD, Buchmueller T, Decker SL, Carey C, Kronick R. 2013. The Affordable Care Act has led to significant gains in health insurance and access to care for young adults. *Health Aff*. 32:165–74
44. Ware JE Jr., Kosinski M, Keller SD. 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34:220–33
45. Winkelmann R. 2004. Health care reform and the number of doctor visits—an econometric analysis. *J. Appl. Econom.* 19:455–72
46. Winkelmann R. 2008. *Econometric Analysis of Count Data*. Berlin/Heidelberg: Springer

Annual Review of Public Health

Volume 39, 2018

# Contents

## Symposium



## Epidemiology and Biostatistics

## Social Environment and Behavior



## Environmental and Occupational Health

## Public Health Practice and Policy

## Health Services



## Indexes



## Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at http://www.annualreviews.org/errata/publhealth