NBER WORKING PAPER SERIES

GETTING THE RIGHT TAIL RIGHT:
MODELING TAILS OF HEALTH EXPENDITURE DISTRIBUTIONS

Martin Karlsson
Yulong Wang
Nicolas R. Ziebarth

Working Paper 31444
http://www.nber.org/papers/w31444

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2023

We thank Anirban Basu, John Mullahy, and Edward Norton for extremely helpful comments and suggestions. Further, we thank all participants of the Annual Health Econometrics Workshop at Emory University for very helpful thoughts. In particular we thank our discussant Lei Liu. We also thank representatives of the German Association of Private Health Insurers for invaluable help with the private insurer claims dataset. We do not have financial interests that would constitute any conflict of interests with this research. Generous funding by the German Federal Ministry of Education and Research (FKZ: 01EH1602A) is gratefully acknowledged. All the remaining errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Getting the Right Tail Right: Modeling Tails of Health Expenditure Distributions
Martin Karlsson, Yulong Wang, and Nicolas R. Ziebarth
NBER Working Paper No. 31444
July 2023
JEL No. C10,C13,I10,I13

ABSTRACT

Health expenditure data almost always include extreme values. Such heavy tails can be a threat to the commonly adopted least squares methods. To accommodate extreme values, we propose the use of an estimation method that recovers the often ignored right tail of health expenditure distributions. We apply the proposed method to a claims dataset from one of the biggest German private health insurers and find that the age gradient in health care spending differs substantially from the standard least squares method. Finally, we extend the popular two-part model and develop a novel three-part model.

Martin Karlsson
CINCH, University of Duisburg-Essen
Weststadttürme Berliner Platz 6-8
45127 Essen
Germany
martin.karlsson@uni-due.de

Yulong Wang
Syracuse University
Maxwell School of Citizenship
and Public Affairs
Economics Department
127 Eggers Hall
Syracuse, NY 13244
ywang402@syr.edu

Nicolas R. Ziebarth
Cornell University
and ZEW Mannheim
L7,1, D-68161 Mannheim,
Germany
and also IZA Bonn, NBER, and
the University of Mannheim
nicolas.ziebarth@zew.de

1 Introduction

For decades, health economists have studied how to appropriately model inherently skewed health expenditure distributions. It is a stylized fact that populations around the world exhibit spending distributions where roughly 50% of total health expenditures falls on 5% of the sickest individuals in society (cf. French and Kelly, 2016; Karlsson, Klein, and Ziebarth, 2016; Finkelstein, 2020). In other words, health expenditure data routinely exhibit heavy right tails (cf. Handel, Kolstad, and Spinnewijn, 2019). Such a distributional property could violate the assumptions required by the ordinary least squares (OLS) estimation. It may lead to a poor finite sample performance.

If a distribution has a heavy tail, random sample draws from this distribution are likely to generate very large values, which are often treated as outliers. Applied economists have adopted several approaches to deal with such outliers. Jones (2011), Manning (2012), and Mihaylova, Briggs, O'Hagan, and Thompson (2011) provide comprehensive overviews of alternative econometric models. A widely used approach is the generalized linear model (GLM) with a log-link function (Mullahy, 1998; Manning and Mullahy, 2001; Manning, Basu, and Mullahy, 2005; Deb, Norton, and Manning, 2017). GLM is one of the most sophisticated health econometric approaches. It attempts to capture particularities of health care spending distributions – including their long right tails and large mass points at zero spending. Further, it is more efficient than the transformed log model (Manning and Mullahy, 2001; Buntin and Zaslavsky, 2004). A much simpler but very popular alternative approach is to trim the top percentiles of the distribution, or to top code them.[1] However, extreme health care spending values are neither classical outliers nor measurement error. Hence, simply deleting or ignoring them means

[1]Other very important approaches include the two-part model, which employs a binary outcome model along with a conditional model for positive spending (Newhouse and Phelps, 1976; Manning, Newhouse, Duan, Keeler, and Leibowitz, 1987; Mullahy, 1998) as well as count data models or latent class models that differentiate between frequent and infrequent users of health care; for example, when modeling the number of outpatient doctor visits (Deb and Trivedi, 1997, 2002; Pohlmeier and Ulrich, 1995).

to potentially ignore valuable information. What's more, the researcher then ignores precisely those individuals who are responsible for the lion's share of per capita health care spending. On the other hand, "how to contain health care costs?" has probably been *the* recurring theme for health economists and policymakers around the world for decades. Therefore, methodological refinements for credible empirical analysis of the right tail of health expenditure distributions harbor great potential to deliver answers to some of the most pressing policy questions.

In this paper, we propose to focus on and specifically model heavy right tails of health expenditure distributions. Our approach intends to accommodate extreme values explicitly. To do so, it focuses on the top 5% of spenders who produce 50% of all health care spending. For that purpose, we use a high-quality claims dataset containing half a million policyholders from one of the biggest German private health insurers. Big insurance pools are required in such a setting where the focus is on the top percentiles of spenders.

We study the heavy tail feature using the following steps: In a first step, using log-rank-log-size plots, we show that the tails of the claims data exhibit clear features of a Pareto distribution. This implies a highly nonlinear relationship between individual i's predictors X_i and their medical spending Y_i. Moreover, we estimate the Pareto exponent, which characterizes the heaviness of the tail of the underlying distribution. We find that the Pareto exponent is around two in our dataset, which implies that the finite second moment condition of OLS is very likely violated, leading to poor performance of OLS and t-tests. Using simulations, we then study the behavior of OLS under the Pareto heavy tail. Moreover, we show that the Pareto and heavy tail features lead to biases in OLS estimates as well as rejection errors when conducting inference.

In the next step, we propose an alternative method that leads to unbiased estimation and asymptotically correct statistical inference. To do so, we exploit the Pareto tail feature and introduce a maximum likelihood estimator (MLE) for the pseudo-true parameter and, more importantly, the marginal effects. This method was initially proposed and studied by Wang and Tsai (2009) and Wang and Li (2013) in the statistics literature. We tailor their approach for the health expenditure context and benchmark it against a simple linear specification. Further, we incorporate our method into the widely used two-part model (e.g., Manning, 1998; Mullahy, 1998) which employs a binary outcome model along with a conditional model for positive spending. We propose a novel three-part model by incorporating our tail MLE as the third part. For empirical users, we also provide a cookbook recipe of the various steps to implement our method.

After that, using the German claims dataset, we estimate the marginal effects and calculate the standard errors for exogenous spending predictors, such as age and gender, both for the standard OLS estimator and for our proposed approach. We provide explicit evidence on the relevance of extreme outliers for the robustness of OLS estimation. Our findings demonstrate the sensitivity of OLS to extreme outliers. Further, coherent with our simulation results, we find that the OLS point estimates of the age-spending nexus lie below the marginal effects estimated by our proposed method along the entire age distribution from age 35 to 75.

We are not the first in the literature to study heavy tail features of health expenditure data. As stated in Mullahy (2009), *"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 some traditionally interesting parameters (means, variances) may not even be finite, a situation never encountered in, e.g., a normal or log-normal world."* Our paper follows this lead and rigorously studies situations in which the variance of the health expenditure is infinite. For alternative models, Manning et al. (2005) propose to use the generalized gamma distribution, and Jones, Lomas, and Rice (2014) propose to use the generalized beta of the second kind (GB2) distribution, which covers Pareto distribution and Burr-Singh-Maddala distribution as special cases. Using Monte Carlo simulations, Jones, Lomas, and Rice (2015) and Jones, Lomas, Moore, Rice et al. (2013) evaluate the empirical performance of a range of different empirical techniques for modelling the distribution of health care expenditures. One performance indicator of their comparison is to accurately represent the distribution of the right tail.

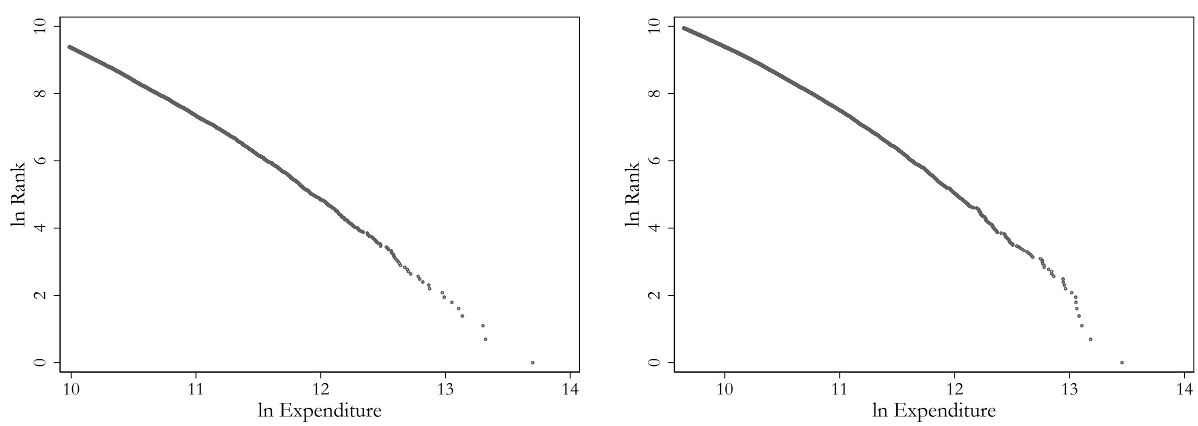
Our proposal is in stark contrast to these existing methods in two ways. First, we only model the *tail* part of the health expenditure data while most of the existing methods model the whole distribution. As our method is based on the Pareto tail approximation, which does not necessarily hold for the whole distribution, it entails more robustness to mis-specification from the non-tail part. In addition, exploiting the Pareto approximation safeguards against mis-specification of the distribution within the right tail. Accordingly, we propose to extend the existing two-part model to a three-part model, in which the additional part applies only to extreme values. Multi-part models for health expenditures have a long tradition in health economics. Duan, Manning, Morris, and Newhouse (1982) proposes a four-part model to deal with different distributional properties of ambulatory and other expenses. Gilleskie and Mroz (2004) propose a flexible estimator of the conditional density of expenditures within a number of set intervals. What is new in our application is to propose a parsimonious estimator that captures key features of the tails of the expenditure distribution. Second, we allow the tail heaviness (more precisely, the Pareto exponent) to depend on covariates such as age and gender, while the existing methods typically consider the tail heaviness to be a fixed constant. Using our claims data, we empirically document that the tail heaviness varies across gender and age, supporting the use of our method.

In Section 2, we preview the heavy tail feature of our data. Using simulation studies, we show that the commonly used least squares method performs poorly under such heavy tails. Next, we introduce our proposed method which explicitly accommodates heavy tails and extreme values. Finally, we extend the two-part model and develop a novel three-part model. Then, we provide an empirical guide for its implementation. In Section 3, we introduce more details about our claims dataset and present descriptive results. In Section 4, we apply the proposed method and present the empirical findings. The mathematical details, additional simulation results, and robustness analysis are in the Appendix.

2 Failure of OLS and How to Model Pareto Tails

2.1 Preview of the Pareto Tail

We start by presenting the Pareto tail feature in our medical expenditure dataset. Let Y_i denote health care spending of individual i for $i = 1, \ldots, n$, where n denotes the total sample size. Also, let $Y_{(1)} \geq Y_{(2)} \geq \ldots \geq Y_{(n)}$ be the descending and ordered expenditure values whose ranks are accordingly $1, 2, ..., n$. Figure 1 plots the natural logarithms of the rank i against $\ln Y_{(i)}$ for the largest 5% of all values. We separately show the plots for females (left) and males (right).



(a) Females

(b) Males

Figure 1: Rank-Size Plots of Natural Logarithms of Rank against Expenditures

Notes: The left graph shows plots for females and the right graph shows plots for males. See Section 3 for more details about the German health care claims data.

Both Figure 1a (Females) and b (Males) clearly suggest a linear fit in the rank-size plots. As has been extensively shown (e.g., Gabaix, 2009), this pattern implies that the underlying distribution exhibits a Pareto tail, or equivalently, the power law. More specifically, if Y_i has a Pareto distribution beyond some cutoff value $y_{\min}$, we have that

$$\mathbb{P}\left(Y_i > y | Y_i \geq y_{\min}\right) = \left(\frac{y}{y_{\min}}\right)^{-\alpha}, \tag{2.1}$$

where α is called the *Pareto exponent*, a positive parameter that uniquely characterizes the heaviness of the tail. Given the Pareto tail assumption, the slope of the linear fit in the rank-size plot is equal to $-\alpha$. Furthermore, the parameter $y_{\min}$ determines the cutoff location of the tail, above which the Pareto distribution serves as a good approximation of the underlying distribution. We discuss such Pareto tails from two perspectives.

First, from a theoretical perspective, it has been established in the statistics literature (e.g., Smith, 1987) that many commonly used distributions can be well approximated by Pareto distributions as long as one focuses on a sufficiently far tail region, that is, by considering a sufficiently large $y_{\min}$. Examples include the Student-t, the F and the Cauchy distributions, among many others.[2]

Accordingly, we can treat $y_{\min}$ as a *tuning parameter* that determines the precision of the Pareto tail approximation. Note that this concept is close in spirit to the choice of the bandwidth parameter in nonparametric kernel estimations. In practice, we set $y_{\min}$ as the 95% quantile and present robustness checks with alternative cutoffs in Appendix A1.

Second, from an empirical perspective, the Pareto tail has been widely documented in many other datasets in economics and finance, such as stock returns, city size, firm size, and income, see Gabaix (2009, 2016) for reviews of other datasets that exhibit Pareto tails.

Before introducing our proposed method, we examine the standard OLS estimation method under the Pareto tail. OLS causes two potential problems: The first is a bias due to the strong nonlinearity implied by the Pareto distribution. The second is a potentially infinite variance due to the heavy tail.

First, let X_i denote a vector of exogenous individual spending predictors such as age and

[2] In particular, the Pareto exponent α is equal to the degree of freedom when the underlying distribution is Student-t.

gender. Since the Pareto distribution (2.1) is uniquely characterized by the exponent α, the effect of X_i on Y_i in the *tail* is through $\alpha = \alpha(X_i)$. The power function $y^{\alpha(X_i)}$ naturally generates a nonlinear effect of X_i on the expected value of Y_i. Conversely, a model based on a linear specification, such as a simple OLS, could produce substantially biased results. We present such bias in a simple simulation study in the following subsection.

Second, the slope in Figure 1 is around -2, implying that the underlying distribution of health care expenditure has a very heavy tail. In particular, the tail of the Pareto distribution (2.1) is heavier with a smaller α. Moreover, the Pareto distribution implies that for any $r > 0$ (e.g., Mikosch, 1999):

$$\mathbb{E}[Y_i^r] < \infty \text{ if } r < \alpha \text{ and } \mathbb{E}[Y_i^r] = \infty \text{ if } r > \alpha.$$

Accordingly, when α is less than two, the tail is so heavy that $\mathbb{E}\left[Y_i^2\right]$ becomes infinity! Recall that the asymptotic normality of the OLS estimator and the t-statistic require that the second moment of Y_i is finite, that is, $\mathbb{E}\left[Y_i^2\right] < \infty$.

The heavy tail feature in the data then compromises this *population* moment condition, even though the finite sample variance is always defined. Hence the OLS estimator, the standard t-test, and estimates of any other higher-order moments such as skewness and kurtosis may perform poorly. We also study this consequence of heavy tails using simulation in the next subsection.

2.2 Monte Carlo Simulation Studies

Ordinary Least Squares

The effect of Pareto tails on coefficient bias. We now perform a simple simulation study to illustrate the effect of the Pareto tail. First, we focus on the potential bias of the

OLS estimator due to the nonlinearity. To this end, we generate Y_i from the standard Pareto distribution (2.1) such that

$$\mathbb{P}(Y_i > y | Y_i > y_{\min}, X_i = x) = \left(\frac{y}{y_{\min}}\right)^{\alpha(x)},$$

where X_i is an independent draw from the absolute value of the standard normal distribution. We set $\alpha(x) = \exp(1 + x\beta_0)$ with $\beta_0 = 1$ as the pseudo-true parameter. This setup guarantees that $\alpha(X_i)$ is always positive. Since the Pareto tail is invariant to scale, we set $y_{\min} = 1$ without loss of generality in this simulation.[3] Moreover, the minimum value of $\alpha(x)$ is $\exp(1) = 2.718 > 2$, implying that the variance of Y_i, given X_i, is always finite. Therefore, the potential bias of the OLS method could only originate from misspecification due to nonlinearities in the tail, as we will see in Figures 2 and 3 below.

The Pareto distribution implies that $\mathbb{E}[Y_i | X_i = x] = \alpha(x)/(\alpha(x) - 1)$. Then the marginal effect of X_i on the average of Y_i is

$$\frac{\partial \mathbb{E}[Y_i | X_i = x]}{\partial x} = -\frac{\alpha(x)}{(\alpha(x) - 1)^2}\beta_0,$$

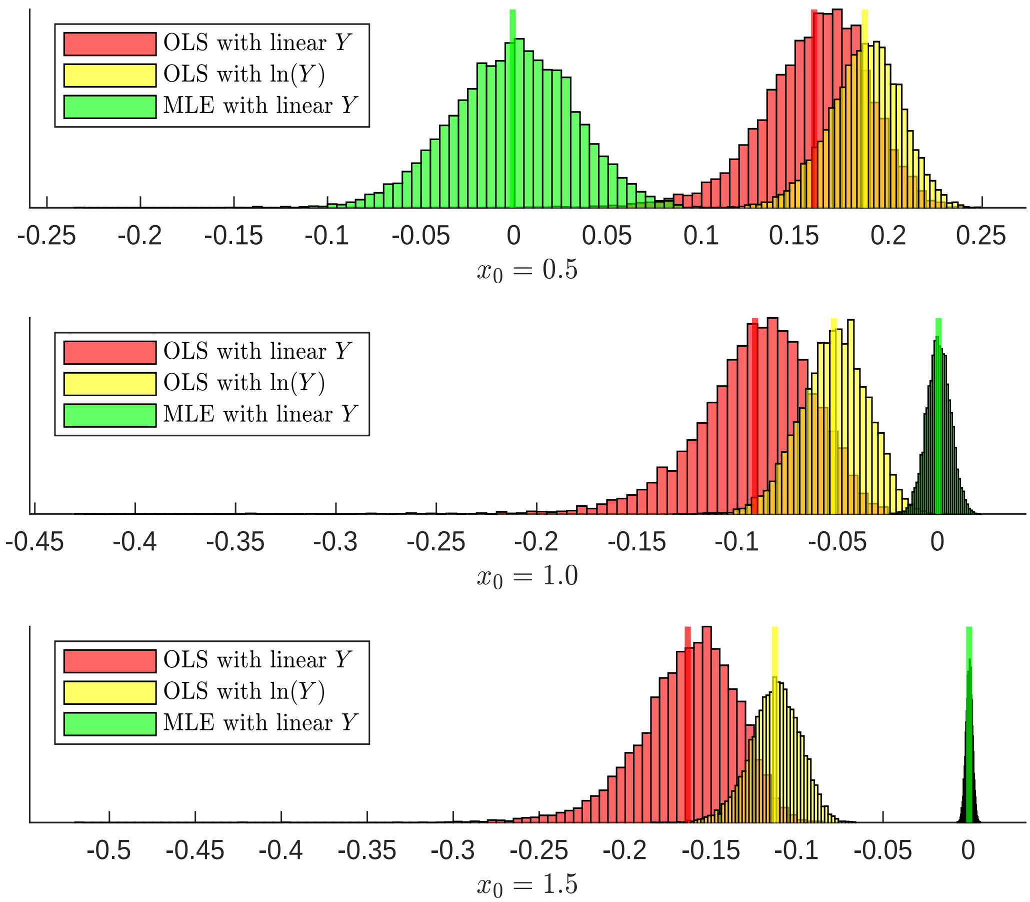
This is the main object of interest. When β_0 is positive, a larger x leads to a larger $\alpha(x)$ and hence a thinner tail. Then, accordingly, the expectation of Y_i conditional on being in the tail is smaller.

To estimate the marginal effect, we implement three methods. First, we use the standard OLS estimator, regressing Y_i on X_i (and a constant). The OLS coefficient estimates the marginal effect. By construction, such an estimated marginal effect is constant regardless of which value of X_i we condition on (red in Figures 2 and 3). Second, we regress $\ln Y_i$ on X_i (and a constant) to obtain the coefficients $(\hat{\beta}_0, \hat{\beta}_1)$. Given the logarithm, the marginal effect of X_i

[3]Conversely, the tail is *not* invariant to an additive transformation like, e.g. the amount of spending above a uniform deductible. However, such transformations are of limited relevance when studying the right tail of health expenditures.

on Y_i evaluated at $X_i = x_0$ is then estimated as $\exp(\hat{\beta}_0 + x_0\hat{\beta}_1)\hat{\beta}_1\overline{\exp(\hat{u})}$, where $\overline{\exp(\hat{u})}$ is the average of the exponential of the residuals (yellow in Figures 2 and 3); hence, the estimated marginal effect varies with X_i even though β does not. Third, we implement our proposed MLE method (green in Figures 2 and 3); for reasons of readability, we postpone the details of our MLE to the next subsection.

Figure 2: Histograms of OLS with Linear or Natural Logarithms of Y and the proposed MLE



Notes: This figure depicts the histograms of the OLS estimator with Y_i (red color) or $\ln Y_i$ (yellow color) and the MLE (green color) for a marginal effect evaluated at $x_0 = 0.5, 1, 1.5$ and $\beta_0 = 2$. The vertical line depicts the averages of the estimators. Results are based on 10,000 simulation draws. See the main text for more details about the data generating process.

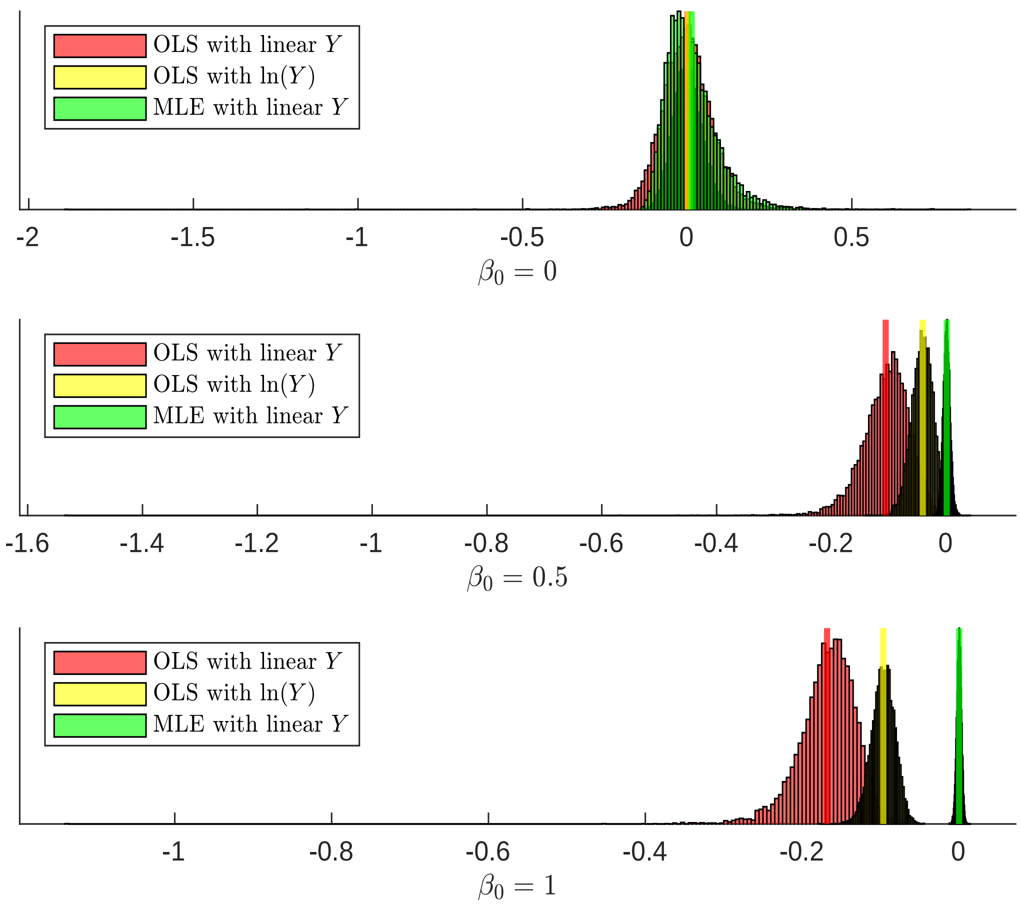
Figure 2 depicts the histograms of the OLS estimators—the true marginal effect substracted with 0 indicating no bias. The figure shows the results of the first method of Y_i (red color) as well as the second method $\ln Y_i$ (yellow color) and our proposed MLE (green color) . The histrograms are based on 500 observations in each simulation and 10000 simulation draws. The top/middle/bottom panel corresponds to the marginal effect evaluated at $x_0 = 0.5/1/1.5$ and $\beta_0 = 2$.

We find the following: It is evident that the OLS estimator is substantially biased, regardless of whether we use Y_i or $\ln Y_i$. By constrast, our proposed MLE estimator is unbiased. The bias here is due to the fact that the marginal effect is highly nonlinear in X_i, while the OLS method specifies a linear model. We emphasize that such a bias exists only in the tail but not necssarily below $y_{\min}$ where a linear model is more reasonable and OLS could still perform well. Therefore, we consider our proposed method as a useful complement to study tail features of heavily skewed distributions such as medical spending.

Next, we repeat the previous analysis with data generated from the same process as in Figure 2. We maintain that $x_0 = 2$, but now vary $\beta_0 = 0, 0.5, 1$. The histograms of the OLS estimators with Y_i and $\ln Y_i$ and our proposed MLE are in Figure 3. In the top panel, where $\beta_0 = 0$, we know that—by construction—X_i does not have any effect on Y_i. Therefore $\mathbb{E}[Y_i|X_i]$ is essentially linear in X_i. In this scenario, the OLS method does not suffer from any misspecification due to nonlinearity and hence the histograms are basically identical as expected. As β_0 increases from zero to one when moving to the bottom panel in Figure 3, the nonlinearity becomes more significant. Hence the bias of the OLS method becomes more severe.

In summary, we know from Figure 1 that Y_i exhibits a Pareto tail. This implies that $\mathbb{E}[Y_i|Y_i > y_{\min}] = y_{\min}\alpha/(\alpha - 1)$. Considering that $\alpha = \alpha(X_i)$ is a function of X_i, the Pareto tail imposes a nonlinear effect of X_i on the tail expectation of Y_i. Such nonlinear effect cannot be well approximated by the linear regression model, except in some special cases. This observation

Figure 3: Histograms of OLS with Linear or Natural Logarithms of Y and the proposed MLE



Notes: This figure depicts the histograms of the OLS estimator with Y_i (red color) or $\ln Y_i$ (yellow color) and the MLE (green color) for the marginal effect with $\beta_0 = 0, 0.5, 1$. The vertical line depicts the averages of the estimators. Results are based on 10000 simulation draws. See the main text for more details about the data generating process.

is the first motivation for our proposed MLE that explicitly takes advantage of the Pareto tail regardless of its heaviness.

The Effect of Heavy Tails on Variance. After having examined the bias, we now evaluate the effect of a heavy tail on the variance of the OLS estimation. To rule out that the effect stems from nonlinearities, we generate data from the standard linear regression model that

$$Y_i = \beta_0 + \beta_1 X_i + u_i,$$

with $(\beta_0, \beta_1) = (1, 0)$ and X_i being i.i.d. standard normal. To characterize the potential heavy tail, we independently generate u_i from a two-sided generalized Pareto distribution that satisfies $\mathbb{P}(u_i \geq u) = \mathbb{P}(u_i \leq -u) = 0.5(1 + \xi u)^{-1/\xi}$ for $u > 0$.

The parameter ξ is called the tail index and equals the reciprocal of the Pareto exponent α. Using the standard OLS, we estimate the coefficients β_0 and β_1 and construct the standard t-statistic and the 95% confidence interval based on heteroskedasticity robust standard errors. We implement our simulations with a wide range of sample sizes $n \in \{500, 1000, 5000, 10^4, 10^5, 10^6\}$. Let $\hat{\beta}_1(s)$ denote the OLS estimator for β_1 in the sth simulated draw. Further, let $\hat{\sigma}(s)$ denote the estimated robust standard error of the OLS estimator $\hat{\beta}_1(s)$. Accordingly, let $t(s) = \hat{\beta}_1(s)/\hat{\sigma}(s)$ denote the t-statistic.

Panel A of Table 1 depicts the mean absolute deviation (MAD): $S^{-1}\sum_{s=1}^{S}|\hat{\beta}_1(s) - \beta_1|$ across $S = 10,000$ simulation draws. Panel B depicts the root mean squared error (RMSE): $(S^{-1}\sum_{s=1}^{S}|\hat{\beta}_1(s) - \beta_1|^2)^{1/2}$ of the OLS estimator. Panel C depicts the average rejection rate for the null that $\beta_1 = 0$, that is $S^{-1}\sum_{s=1}^{S} 1[|t(s)| > 1.96]$. Finally, Panel D of Table 1 depicts the average length of the 95% confidence intervals, that is, $S^{-1}\sum_{s=1}^{S} 2 \times 1.96\hat{\sigma}(s)$.

Table 1 shows the following. First, note that the error term u_i has a finite variance when $\xi < 0.5$ $(\alpha > 2)$. Thus, in the first five rows of Panels A and B, the MAD and the RMSE are reasonably small. In comparison, they become substantially larger in the bottom five rows where $\xi > 0.5$.

Second, when the variance of u_i is finite, we expect the t-statistic to be approximately normally distributed, as implied by the central limit theorem. Therefore, we would expect that the rejection probability is around 5%, as seen in the first five rows of Panel C. However, when

Table 1: OLS Simulation Results with Generalized Pareto Distribution

n	500	1000	5000	10^4	10^5	10^6	500	1000	5000	10^4	10^5	10^6
$\xi(1/\alpha)$	Panel A: MAD						Panel B: RMSE					
0.09	0.06	0.04	0.02	0.01	0.00	0.00	0.07	0.05	0.02	0.02	0.01	0.00
0.19	0.07	0.05	0.02	0.02	0.01	0.00	0.09	0.06	0.03	0.02	0.01	0.00
0.29	0.09	0.06	0.03	0.02	0.01	0.00	0.12	0.08	0.04	0.03	0.01	0.00
0.39	0.13	0.09	0.04	0.03	0.01	0.00	0.18	0.12	0.05	0.04	0.01	0.00
0.49	0.18	0.14	0.07	0.05	0.02	0.01	0.27	0.25	0.12	0.09	0.02	0.01
0.59	0.32	0.24	0.13	0.10	0.05	0.02	1.31	0.68	0.25	0.16	0.08	0.03
0.69	0.63	0.50	0.28	0.23	0.12	0.06	5.09	3.36	0.89	0.79	0.47	0.49
0.79	1.14	0.94	0.74	0.64	0.38	0.24	4.47	3.98	5.43	4.63	1.84	1.40
0.89	2.87	5.19	5.02	3.43	1.88	1.19	46.9	260	291	147	30.0	13.5
0.99	5.61	6.69	5.49	5.68	5.00	7.75	44.5	87.7	44.1	38.7	37.1	156
$\xi(1/\alpha)$	Panel C: Rejection Prob.						Panel D: Length of 95% CI					
0.09	0.05	0.05	0.05	0.05	0.05	0.05	0.28	0.20	0.09	0.06	0.02	0.01
0.19	0.05	0.05	0.05	0.05	0.05	0.05	0.34	0.25	0.11	0.08	0.02	0.01
0.29	0.05	0.05	0.05	0.05	0.05	0.05	0.44	0.31	0.14	0.10	0.03	0.01
0.39	0.05	0.05	0.05	0.05	0.05	0.05	0.59	0.43	0.20	0.14	0.05	0.01
0.49	0.04	0.04	0.05	0.05	0.05	0.05	0.85	0.65	0.32	0.24	0.08	0.03
0.59	0.04	0.03	0.04	0.04	0.04	0.04	1.43	1.08	0.58	0.44	0.18	0.07
0.69	0.03	0.03	0.04	0.04	0.03	0.04	2.75	2.17	1.23	1.01	0.51	0.27
0.79	0.03	0.03	0.03	0.03	0.03	0.03	4.81	3.97	3.14	2.73	1.62	1.03
0.89	0.03	0.03	0.03	0.02	0.03	0.03	11.8	20.8	20.1	13.8	7.70	4.92
0.99	0.02	0.02	0.02	0.02	0.02	0.02	22.9	27.2	22.3	23.1	20.4	31.1

Notes: The table depicts the average mean absolute deviation (MAD), average root mean squared error (RMSE), average rejection probability of the standard t-test, and the average length of the standard 95% confidence intervals. The results are based on 10000 simulation draws. See the main text for details about the data generating process.

$\xi > 0.5$, the rejection probability becomes substantially smaller than 5%.

Third, following the previous point, the underrejection results from large standard errors, as reflected in the long confidence intervals in Panel D. Remember that the confidence interval is expected to shrink at the root-n rate when $\xi < 0.5$. However, when $\xi > 0.5$, the standard error is not well-defined and hence the confidence interval becomes too wide to be informative. More specifically, since the variance of u_i is infinite, we may alternatively consider its inter-quantile range as a benchmark. In our data generating process, the inter-quantile range of u_i is one when $\xi = 1$ and the standard deviation of X_i is always one. So the average length of the confidence interval should be of the order of magnitude $n^{-1/2}$ if the central limit theorem provides a good approximation. However, the average length of the standard CI is substantially larger than $n^{-1/2}$. In the last row of Table 1, where ξ (and α) is approximately one, the average length of the confidence interval is even above 20. Therefore, we believe that the standard OLS-based inference is not performing satisfactorily under the heavy tail distribution.

Finally, our simulations in Table 1 assume a correctly specified model. Given the poor performance of the linear model in the presence of heavy tails, an applied researcher might be tempted to follow the common practice of taking the logarithm of Y_i as the dependent variable as in Figure 2. The performance of such a transformed specification crucially depends on the true data generating process, which is typically unknown. In Appendix A2 we conduct simulations based on a logarithmic specification applied to a linear data generating process with heavy tails. We find that approximating the linear model with a heavy-tailed error by the log-linear model could lead to substantial misspecification errors, which are not even diminishing with the sample size.

Generalized Linear Model

Next, we repeat the previous exercise using the generalized linear model (GLM). The GLM is also widely used in health economics to model health expenditure distributions, see, for instance, Manning and Mullahy (2001) and Buntin and Zaslavsky (2004). More specifically, we generate the data from

$$Y_i = \exp(\beta_0 + \beta_1 X_i) + u_i,$$

with $(\beta_0, \beta_1) = (1, 0)$ and (X_i, u_i) following the same distribution as before. This model implies that the conditional mean $\mathbb{E}[Y_i|X_i = x] = \exp(\beta_0 + \beta_1 x)$, which is nonlinear in X_i. To discipline the estimators, we impose the infeasible bound that the estimators are within $[-50, 50]$.

Table 2: GLM Simulation Results with Generalized Pareto Distribution

n	500	1000	5000	10^4	10^5	10^6	500	1000	5000	10^4	10^5	10^6
$\xi(1/\alpha)$	Panel A: MAD						Panel B: RMSE					
0.09	0.02	0.02	0.01	0.00	0.00	0.00	0.03	0.02	0.01	0.01	0.00	0.00
0.19	0.03	0.02	0.01	0.01	0.00	0.00	0.03	0.02	0.01	0.01	0.00	0.00
0.29	0.04	0.02	0.01	0.01	0.00	0.00	0.07	0.03	0.01	0.01	0.00	0.00
0.39	0.05	0.03	0.02	0.01	0.00	0.00	0.27	0.05	0.02	0.01	0.00	0.00
0.49	0.07	0.05	0.03	0.02	0.01	0.00	0.30	0.19	0.04	0.03	0.01	0.00
0.59	0.12	0.10	0.05	0.04	0.01	0.01	0.53	0.46	0.11	0.07	0.03	0.01
0.69	0.23	0.16	0.10	0.08	0.04	0.02	0.92	0.46	0.30	0.29	0.09	0.06
0.79	0.39	0.33	0.21	0.18	0.11	0.08	1.27	1.14	0.60	0.40	0.24	0.22
0.89	0.62	0.56	0.42	0.39	0.28	0.23	1.89	1.54	1.03	0.99	0.55	0.45
0.99	0.96	0.89	0.72	0.70	0.61	0.52	2.44	2.12	1.51	1.45	1.12	0.89
$\xi(1/\alpha)$	Panel C: Rejection Prob.						Panel D: Length of 95% CI					
0.09	0.05	0.05	0.05	0.05	0.05	0.05	0.10	0.07	0.03	0.02	0.01	0.00
0.19	0.05	0.05	0.05	0.05	0.05	0.05	0.13	0.09	0.04	0.03	0.01	0.00
0.29	0.05	0.05	0.05	0.05	0.05	0.05	0.16	0.12	0.05	0.04	0.01	0.00
0.39	0.05	0.04	0.05	0.05	0.05	0.05	0.22	0.16	0.07	0.05	0.02	0.01
0.49	0.04	0.04	0.04	0.04	0.05	0.05	0.33	1.03	0.12	0.09	0.03	0.01
0.59	0.05	0.05	0.04	0.04	0.04	0.04	66.6	0.42	0.23	0.16	0.07	0.03
0.69	0.06	0.06	0.05	0.05	0.04	0.04	$> 10^3$	$> 10^3$	409	$> 10^3$	1.94	$> 10^3$
0.79	0.08	0.08	0.07	0.07	0.05	0.04	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$
0.89	0.10	0.11	0.11	0.11	0.09	0.08	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$
0.99	0.13	0.13	0.14	0.14	0.13	0.14	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$	$> 10^3$

Notes: The table depicts the average mean absolute deviation (MAD), average root mean squared error (RMSE), average rejection probability of the standard t-test, and the average length of the standard 95% confidence intervals. The results are based on 10000 simulation draws. See the main text for details about the data generating process.

Table 2 depicts the same performance measures as in Table 1 for the GLM method. The findings are also similar. In particular, the GLM estimator has a small bias and RMSE when the error term does not have a heavy tail and $\xi < 0.5$ in Panels A and B. The confidence interval is short and shrinking with the sample size.

In contrast, in Panel C, the t-test substantially overrejects when the error term has a heavy tail. Moreover, the confidence interval is exploding when ξ exceeds 0.6. Such wide confidence intervals originate from the poor standard error estimates in the GLM. To see this, note that we obtain the GLM estimator $\hat{\beta}$ by solving the nonlinear least squares problem

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (Y_i - \exp(\beta_0 + \beta_1 X_i))^2.$$

Some derivation shows that the asymptotic variance of $\hat{\beta}$ becomes

$$\mathbb{E}[u_i^2] \begin{pmatrix} \mathbb{E}[\exp(2(\beta_0 + \beta_1 X_i))] & \mathbb{E}[\exp(2(\beta_0 + \beta_1 X_i))X_i] \\ \mathbb{E}[\exp(2(\beta_0 + \beta_1 X_i))] & \mathbb{E}[\exp(2(\beta_0 + \beta_1 X_i))X_i^2] \end{pmatrix}^{-1}.$$

When the error has a heavy tail, $\mathbb{E}[u_i^2]$ becomes extremely large. Furthermore, in this case, the estimator can be numerically unstable so that the above matrix is not invertible. Both features lead to large standard errors and hence wide and uninformative confidence intervals. The results become even worse when we relax the restriction that $\hat{\beta}_j \in [-50, 50]$ for $j = 0, 1$.

In summary, ignoring heavy tails in heavily skewed data, such as health expenditure data, can lead to substantial estimation biases as well as rejection errors in the statistical inference of unknown parameters. As shown, such errors could become even more severe in nonlinear GLM than linear OLS models, motivating our proposed method that explicitly focuses on the heavy tail.

2.3 Modeling Pareto Tails and the Proposed Maximum Likelihood Estimator

Given the failure of OLS under a Pareto tail, we move forward to construct a valid alternative that explicitly accommodates extreme values. For illustrative purposes, we preview the new approach in this subsection by assuming an exact Pareto tail. In fact, the econometric derivation only requires an *approximate* Pareto tail, which holds for many commonly used distributions such as Student-t, F, Gamma, *et cetera*. For reasons of readability, we relegate the technical details and the primitive assumptions to Appendix A3 .

The new method is based on the tail index regression proposed by Wang and Tsai (2009). First, assuming Y_i has an exact Pareto tail above $y_{\min}$, we obtain

$$\mathbb{P}\left(Y_i > y | Y_i > y_{\min}, X_i = x\right) = \left(\frac{y}{y_{\min}}\right)^{-\alpha(x)}, \tag{2.2}$$

where $\alpha\left(x\right)$ is the Pareto exponent that depends on the characteristics $X_i = x$. We adopt the model $\alpha = \exp\left(X_i'\beta_0\right)$, where β_0 again denotes the pseudo-true coefficient. The exponential function guarantees that the Pareto exponent is always positive, and the linear index form is adopted mainly for computational simplicity. Using only the observations above $y_{\min}$, we then obtain the following negative log-likelihood function:

$$\mathcal{L}\left(\beta\right) = n^{-1}\sum_{i=1}^{n}\left\{\exp(X_i'\beta)\log\left(Y_i/y_{\min}\right) - X_i'\beta\right\}\mathbf{1}\left[Y_i > y_{\min}\right], \tag{2.3}$$

where $\mathbf{1}\left[\cdot\right]$ denotes the indicator function. Then, the maximum likelihood estimator (MLE) of β_0 is

$$\hat{\beta} = \arg\max_{\beta}\mathcal{L}\left(\beta\right).$$

We can estimate its asymptotic variance as

$$\hat{\Sigma}_{\beta} = \left(n_0^{-1}\sum_{i=1}^{n}X_iX_i'\mathbf{1}\left[Y_i > y_{\min}\right]\right)^{-1}, \tag{2.4}$$

where $n_0 = \sum_{i=1}^{n}\mathbf{1}\left[Y_i > y_{\min}\right]$ denotes the total number of tail observations.

We provide two remarks about the proposed MLE. First, the Pareto tail assumption (2.2) implies that the conditional expectation of health expenditures beyond $y_{\min}$ is

$$\mathbb{E}\left[Y_i | Y_i > y_{\min}, X_i = x\right] = y_{\min}\frac{\alpha\left(x'\beta_0\right)}{\alpha\left(x'\beta_0\right) - 1}. \tag{2.5}$$

The tail cutoff $y_{\min}$ is again a tuning parameter chosen by the econometrician. In our subsequent analysis, we use the 95% quantile as $y_{\min}$. Appendix A3 provides more details about the choice of this parameter.

Given the Pareto tail (2.2), the marginal effect of X_i on tail expenditures is

$$
\begin{aligned}
M(x;\beta_0) &\equiv \frac{\partial \mathbb{E}\left[Y_i|Y_i > y_{\min}, X_i = x\right]}{\partial x} \\
&= -y_{\min}\frac{\exp\left(x'\beta_0\right)}{\left(\exp\left(x'\beta_0\right)-1\right)^2}\beta_0 \\
&= -\frac{\mathbb{E}\left[Y_i|Y_i > y_{\min}, X_i = x\right]}{\left(\exp\left(x'\beta_0\right)-1\right)}\beta_0, \qquad (2.6)
\end{aligned}
$$

which we estimate by replacing β_0 with our MLE $\hat{\beta}$ in (2.3).

As seen, the marginal effect is a function of X_i. Hence the proposed estimator allows for a nonlinear impact of individual characteristics, such as age, on expected spending in the tail. This may be a desirable feature as average health care spending increases with age at a faster rate for seniors. In contrast, an OLS specification assumes that the marginal effect is constant, unless higher-order polynomials are included in the regression equation. We emphasize that the non-linearity is not generic but specifically due to the Pareto tail. One could include higher-order and interaction terms in $\alpha(x)$ for more flexibility. It should also be noted that whenever the model includes several independent variables, the assumption is that their interaction effects are non-zero (apart from the special case where the marginal effect is zero).

Using the Delta method, we can construct the standard errors for the marginal effects. In particular, for the marginal effect of the jth component of X_i, we have that

$$\begin{aligned}\nabla_j M(x,\beta_0) &\equiv \left.\frac{\partial M(x;\beta)}{\partial \beta_j}\right|_{\beta=\beta_0} \\ &= -y_{\min}\left[e_j \frac{\exp(x'\beta_0)}{(\exp(x'\beta_0)-1)^2}\right. \\ &\quad \left. -2x\frac{\exp(2x'\beta_0)}{(\exp(x'\beta_0)-1)^3}\beta_{0j} + x\frac{\exp(x'\beta_0)}{(\exp(x'\beta_0)-1)^2}\beta_{0j}\right],\end{aligned}$$

where e_j denotes the jth standard unit vector. Then the estimate of the standard error is

$$\hat{\Sigma}_{M_j} = \nabla_j M\left(x,\hat{\beta}\right)' \hat{\Sigma}_\beta \nabla_j M\left(x,\hat{\beta}\right). \tag{2.7}$$

In summary, we propose the following steps:

1. Given $y_{\min}$, say the 95% quantile of Y_i, select all Y_i's that are larger than $y_{\min}$.
2. Construct the MLE by numerically solving (2.3) and estimate the standard error using (2.4).
3. Estimate the marginal effect (2.6) and the standard error (2.7).
4. Perform robustness check by using different $y_{\min}$.
5. Generate the counterfactual of the conditional tail expectation using (2.5).

2.4 Extension to a Three-Part Model

So far, we have focused on the tail solely using observations $Y_i > y_{\min}$. In this subsection, we generalize the previous analysis to model the whole distribution and extend the existing two-part model (cf., Mullahy, 1998) to a three-part model.[4]

[4]We thank Anirban Basu and Edward Norton for proposing this extension.

In particular, the widely used two-part model is designed to capture that many observations of Y_i are zero. To model this, consider

$$\mathbb{E}[Y_i|X_i = x] = \mathbb{P}(Y_i > 0|X_i = x) \times \mathbb{E}[Y_i|Y_i > 0, X_i = x], \tag{2.8}$$

provided that $Y_i > 0$ almost surely.

In the first part, we fit the binary outcome $1[Y_i = 0]$ with a standard logit or probit model. Then we estimate the partial effect on $\mathbb{P}(Y_i > 0|X_i = x)$ of X_i. In the second part, we run regressions of Y_i (or $\ln Y_i$) on X_i. We obtain the overall marginal effect $\partial\mathbb{E}[Y_i|X_i = x]/\partial x$ by combining the estimates from both parts.

Given the Pareto tail, we can extend (2.8) and propose the following three-part model:

$$\begin{aligned}\mathbb{E}[Y_i|X_i = x] \quad = \quad & \mathbb{E}[Y_i|0 < Y_i \leq y_{\min}, X_i = x] \times \mathbb{P}[0 < Y_i \leq y_{\min}|X_i = x] \\ & + \mathbb{E}[Y_i|Y_i > y_{\min}, X_i = x] \times \mathbb{P}[Y_i > y_{\min}|X_i = x]. \end{aligned} \tag{2.9}$$

Three-Part Model. In the first part, we estimate the conditional probabilities $\mathbb{P}[0 < Y_i \leq y_{\min}|X_i = x]$ and $\mathbb{P}[Y_i > y_{\min}|X_i = x]$ by running a multinomial logistic regression. More specifically, denote $Y^* = 0, 1, 2$ if $Y_i = 0, Y_i \in (0, y_{\min}), Y_i > y_{\min}$, respectively. Then, the multinomial logistic regression fits

$$\mathbb{P}(Y_i^* = j|X_i = x) = \frac{\exp(x'\theta_j)}{1 + \sum_{j=0}^{2} \exp(x'\theta_j)}, \tag{2.10}$$

for $j = 1, 2$ and θ_0 is understood as zero for normalization. Denote the estimated coefficient as $\hat{\theta}_j$. In the second part, we run a linear regression of Y_i on X_i with observations $Y_i \in (0, y_{\min})$. Denote the regression coefficient as $\hat{\gamma}$. Given the upper bound $y_{\min}$, we do not have to consider $\ln Y_i$. In the third part, we implement the MLE method as described in the previous subsection.

Combining all three parts, we then estimate conditional expectation by

$$\begin{aligned}\hat{\mathbb{E}}[Y_i|X_i = x] &= x'\hat{\gamma} \times \frac{\exp(x'\hat{\theta}_1)}{1 + \sum_{j=0}^{2} \exp(x'\hat{\theta}_1)} \\ &\quad + y_{\min} \frac{\exp(x'\hat{\beta})}{\exp(x'\hat{\beta}) - 1} \times \frac{\exp(x'\hat{\theta}_2)}{1 + \sum_{j=0}^{2} \exp(x'\hat{\theta}_2)}.\end{aligned}$$

Finally, we obtain the partial effect $\partial \mathbb{E}[Y|X = x]/\partial x$ by taking the derivative, and obtain the standard error by bootstrapping.

Limitations. As a final note, we discuss the underlying distributional assumptions and potential limitations of the three-part model.

In the existing two-part model (2.8), we assume that $\mathbb{P}(Y_i > 0|X_i = x)$ is characterized by a parametric binary probability model like logit or probit. And that $\mathbb{E}[Y_i|Y_i > 0, X_i = x]$ is a linear or log-linear function of x. See, for example, Mullahy (1998) and Manning (1998).

In a similar fashion, our three-part model assumes that (i) $\mathbb{P}(Y_i \in (0, y_{\min})|X_i = x)$ and $\mathbb{P}(Y_i > y_{\min})|X_i = x)$ are governed by a parametric multinomial logit model, (ii) that $\mathbb{E}[Y_i|Y_i > y_{\min}, X_i = x]$ is governed by the Pareto tail as in (2.5), and (iii) that $\mathbb{E}[Y_i|Y_i \in (0, y_{\min}), X_i = x]$ is linear in x. Obviously, these assumptions are stronger than those for the two-part model. They possibly lead to more bias in estimating the marginal effects.

Again following Mullahy (1998), one could relax these assumptions by considering alternative models in all three parts. In particular, one could consider the Heckman selection model or the modified two part-model for estimating the conditional probabilities $\mathbb{P}(Y_i > 0|X_i = x)$. Then, for observations $Y_i > 0$, one could further decompose the data based on $Y_i > y_{\min}$ or not. However, as known, the Heckman model requires a valid exlcusion restriction.

In this sense, we consider our proposed three-part model as one of the potential extensions of the popular two-part model, one that is specifically designed to accommodate the heavy tail feature of health care spending data. An extensive study of other extensions is beyond the scope of this paper. It is an important topic for future research.

3 Data

This section describes the claims data used in this paper. The main working sample focuses on the privately insured in the German health care system. Note that the policyholders do not have supplemental private insurance, but comprehensive long-term health insurance over their lifecycles until death. For more details on the German two-tier health care system and German private health insurance, please see Atal, Fang, Karlsson, and Ziebarth (2023).

The claims data are administrative records on the universe of policies and claims between 2005 and 2011 from one of the largest private health insurers in Germany. In total, our dataset includes more than 2.6 million enrollee-year observations from 620 thousand unique policyholders along with detailed information on plan parameters such as premiums, claims, and diagnoses. Atal, Fang, Karlsson, and Ziebarth (2019) provide more details about the dataset. The data also contain the age and gender of all policyholders as well as their occupational group. We convert all monetary values to 2016 U.S. dollars (USD).

Sample Selection. We focus on primary policyholders. In other words, we disregard insured children and those who are younger than 25 years (555,690 enrollee-year observations).[5] Moreover, due to a 2009 portability reform (Atal et al., 2019), we disregard inflows after 2008 (253,325 enrollee-year observations). The final sample consists of 1,867,465 enrollee-year observations from 362,783 individuals.

[5] Children obtain their own individual risk-rated policies. However, if parents purchase the policy within two months of birth, no risk-rating applies. Under the age of 21, insurers do not have to budget and charge for old-age provisions.

Descriptive Statistics. Table 3 presents the descriptive statistics. The mean age of the sample is 45.5 years. The oldest policyholder is 99 years old. Thirty-four percent of the sample are high-income employees, 49% are self-employed and 13% are civil servants. The majority of policyholders (72 percent) are male, because women are underpresented among the self-employed and high-income earners in Germany. On average, policyholders have been clients of the insurer for 13 years and have been enrolled in their current health plan for 7 years.[6] The majority of individuals sign join private insurance around the age of 30, when most Germans have fully entered the labor market but are still healthy and are charged reasonable premiums in this risk-rated market (risk rating is only imposed at contract inception and all subsequent premium increases are community rated).

Table 3 shows that the average *annual premium* is $4,749 and slightly lower than the average premium for a single plan in the U.S. group market at the time (Kaiser Family Foundation, 2019). Note that the *annual premium* is the total premium—including employer contributions for privately insured high-income earners.[7] The average *deductible* is $675 per year.

In terms of benefits covered, we simplify the rich data and focus on a plan generosity indicator provided by the insurer. It classifies plans into three coverage tiers: *TOP*, *PLUS*, and *ECO* plans. *ECO* plans are the lowest coverage tier; they lack coverage for services such as single rooms in hospitals and treatments by a leading senior M.D. For *ECO* and *PLUS* plans, a 20% coinsurance rate applies if enrollees see a specialist without referral from their primary care physician. About 38% of all policyholders have a *TOP* plan, 34% a *PLUS* plan, and 29% an *ECO* plan. Because these plan characteristics have mechanical effects on claim sizes and correlate with policyholders' age, we control for them in our estimation of health care costs.

[6] Our insurer doubled the number of clients between the 1980s and 1990s and has thus a relatively young enrollee population, compared to all privately insured in Germany. Gotthold and Gräber (2015) report that a quarter of all privately insured are either retirees or pensioners.

[7] Employers cover roughly one half of the total premium and the self-employed pay the full premium.

Table 3: Summary Statistics: German Claims Panel Data

	Mean	SD	Min	Max	N
Health Plan Parameters					
Total Claims (USD)	3,289	8,577	0	2,345,126	1,867,465
Annual premium (USD)	4,749	2,157	0	33,037	1,867,318
Deductible (USD)	675	659	0	3,224	1,867,465
Annual risk penalty (USD)	157	453	0	21,752	1,867,465
TOP Plan	0.377	0.485	0.0	1.0	1,867,465
PLUS Plan	0.338	0.473	0.0	1.0	1,867,465
ECO Plan	0.285	0.451	0.0	1.0	1,867,465
Socio-Demographics					
Age (in years)	45.5	11.4	25.0	99.0	1,867,465
Female	0.276	0.447	0.0	1.0	1,867,465
Policyholder since (years)	6.5	5.0	1.0	40.0	1,867,465
Client since (years)	12.8	11.0	1.0	86.0	1,867,465
Employee	0.336	0.473	0.0	1.0	1,867,465
Self-Employed	0.486	0.500	0.0	1.0	1,867,465
Civil Servant	0.132	0.338	0.0	1.0	1,867,465
Health Risk Penalty	0.358	0.480	0.0	1.0	1,867,465
Pre-Existing Condition Exempt	0.016	0.126	0.0	1.0	1,867,465

Source: German Claims Panel Data. *Policyholder since* is the number of years since the client has enrolled in the current plan; *Client since* is the number of years since the client joined the company. *Employee* and *Self-Employed* are dummies for the policyholders' current occupation. *Health Risk Penalty* is a dummy that is one if the initial underwriting led to a health-related risk penalty on top of the factors age, gender, and type of plan; *Pre-Existing Conditions Exempt* is a dummy that is one if the initial underwriting led to exclusions of pre-existing conditions. The mutually exclusive dummies *TOP Plan*, *PLUS Plan* and *ECO Plan* capture the generosity of the plan. *Annual premium* is the annual premium, and *Annual Risk Penalty* is the amount of the health risk penalty charged. *Deductible* is the deductible and *Total Claims* the sum all claims in a calendar year. See Section 3 for further details.

4 Results

4.1 The Marginal Effect of Age

Figure 4 depicts estimates of marginal age effects on health care spending. In particular, we implement our MLE as described in equation (2.6) and set the tail cutoff $y_{\min}$ as the 95% quantile of Y_i as the benchmark value. In addition to age, the specification controls for female sex, for plan generosity type, and includes six year dummies. We use the final year 2011 as the reference category.

We summarize the results in Figure 4 as follows: First, the OLS estimate of the marginal age effect is biased: for all ages, it is outside the 95% confidence interval of the MLE estimate. Second, the bias is economically large and meaningful. The smallest marginal effect of the MLE, which is 112 Euro for 29-year-olds, is more than twice the OLS estimate of 51 Euro.

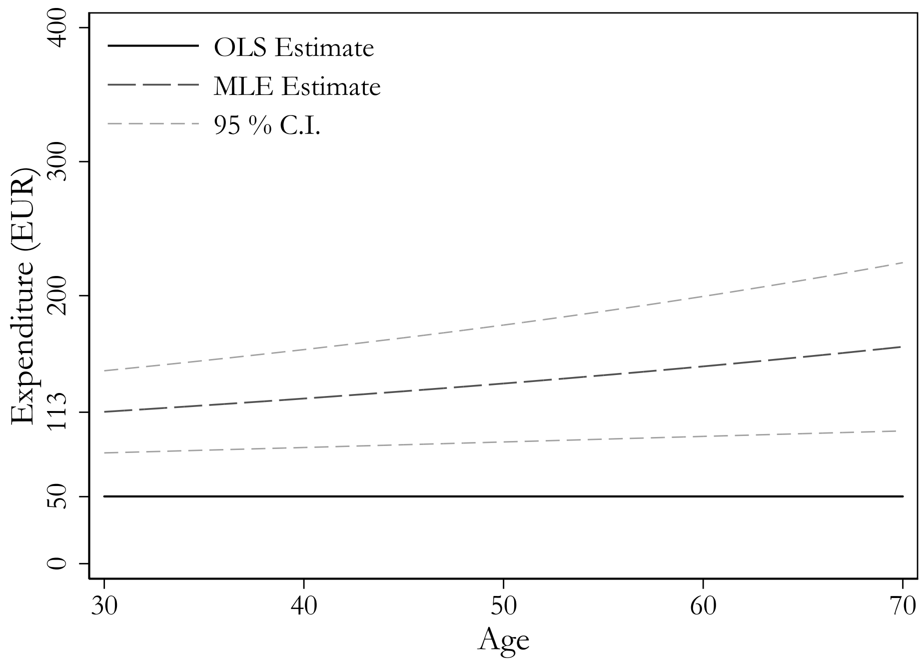
In Figure 5, we split the sample by gender and compare results for males and females. The OLS estimates for females presented in Figure 5a paint a slightly different picture compared to 4 above: over the entire age range, the OLS estimate is within the confidence interval of the MLE estimate. On the other hand, the bias is larger in this subsample: for the youngest females, the OLS estimate is 77% downward biased, compare to 54% in the pooled sample.

In Figure 5b for males, the relative bias is slightly lower at 51%; however, also within this subsample, the OLS estimate lies outside the MLE confidence interval for all ages.

4.2 Discussions and Guidance for Empirical Analysis

The previous simulations and empirical results suggest a substantial difference between our proposed MLE and the classic OLS methods. Both specifications are based on some functional form assumptions regarding the relationship between health care spending and age. More specifically, the OLS assumptions are that expenditures are linear in age, and that the variance

Figure 4: Comparison of OLS and MLE Estimates of Marginal Age Effects.



Notes: See Section 3 for more details about the German health insurance data. The graph compares the marginal age effects on health care spending, along with 95% confidence intervals. The marginal MLE effect is larger across the whole age domain and the difference to OLS increases in age.

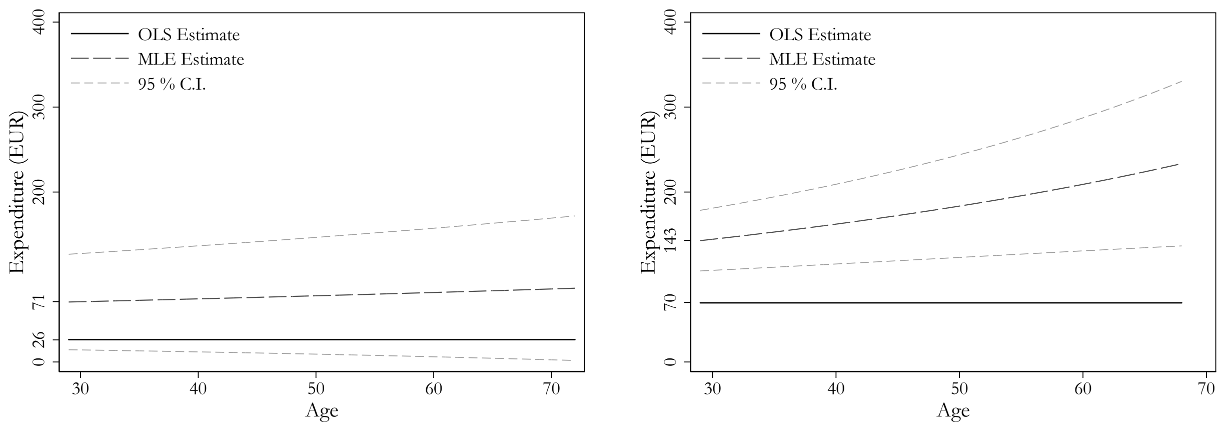
is finite, whereas the MLE estimate allows for a non-linear relationship but requires the Pareto tail, see equation (2.5).

In general, both these and any other functional form assumptions may be incorrect. In cases where the distributional assumptions required for OLS are satisfied, we know that OLS estimates represents the best linear approximation to $\mathbb{E}[Y_i|X_i]$ (Angrist and Pischke, 2008). In cases where the true distribution has an exact Pareto tail above $y_{\min}$, the MLE estimator will be the most efficient one among all consistent estimators. We close this section by some heuristic discussions about the underlying distributional assumptions and provide some empirical guidance.

To compare the proposed MLE and the OLS methods, we essentially need to address two

Figure 5: Marginal Age Effects by Sex, OLS versus MLE.

(a) Females (b) Males



Notes: Own calculations based on German health insurance data. The graph compares the marginal age effects on health care spending, along with 95% confidence intervals.

relevant issues: (1) whether expected expenditures can be well approximately by a linear function of age and other controls, and (2) whether the data exhibit a sufficiently heavy tail so that the variance of the expenditure is possibly infinite.

Note that the sample variance is always finite given any data set, while the unknown population variance could be infinite. In this scenario, the best linear approximation property fails and the sample variance and any other higher moment such as sample skewness and kurtosis are not informative about their population analogs.

The first question, regarding the functional form, depends on the true data generating process, which is usually unknown. In order to shed some light on this issue, we return to the log-size-log-rank plot in Figure 1. If the distribution of expenditures has a Pareto-type tail, we expect to see a linear fit in the plot. In this scenario, age and other control variables could only affect the expenditure through the Pareto exponent $\alpha = \alpha(X)$ and hence

$$\mathbb{E}[Y_i | Y_i > y_{\min}, X_i = x] = y_{\min} \frac{\alpha(x)}{\alpha(x) - 1}.$$

Although the functional form of $\alpha(x)$ is unknown, the conditional mean is in general nonlinear in x and thus results in a bias in the OLS estimation. Thus, the proposed MLE should be applied when the data shows a clear Pareto pattern in the tail. Moreover, we follow Wang and Tsai (2009) to consider the single-index form $\alpha(X) = \exp(X'\beta)$. This additional assumption facilitates the estimation but could again be restrictive.

As a robustness analysis, we now relax the functional form assumptions imposed so far. To do so, we summarize the age information by five age quintile dummies. We then use the entire sample to construct point estimates and confidence intervals for the predicted expenditure values at each age quintile.[8] We plot predicted values as the partial effects of age dummies are poorly defined in the MLE specification.

Figure 6 presents the results. For the youngest group, aged 31.6 years on average, the OLS estimate is close to the MLE estimate of EUR 37,900. For the second youngest group, the OLS point estimate is just outside the 95% CI of the MLE estimate, and there is still considerable overlap between the CI's of the two estimators. For the three oldest groups, however, the two CI's are completely disconnected, and the OLS estimates are substantially below their MLE counterparts. This finding is again coherent with our simulation results.

As a remark, quantile regression is another commonly adopted method to study nonlinear effects. However, we argue that it might not be appropriate in our context. To see why, let $Q_{Y|X=x}(\tau)$ denote the quantile function of Y_i conditional on $X_i = x$ for $\tau \in (0, 1)$.

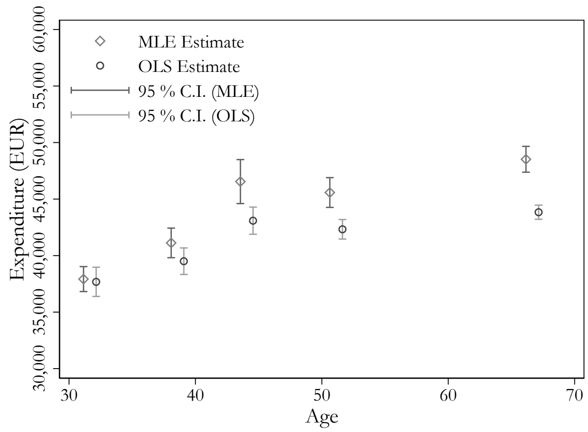
The classical quantile regression model imposes that

$$Q_{Y|X=x}(\tau) = x'\beta(\tau)$$

for some pseudo-true coefficient $\beta(\tau)$. Wang and Li (2013, Proposition 2.1) show that the Pareto exponent $\alpha(x)$ has to be constant for the above quantile regression model to be appropriate.

[8]Thereby it is asserted that the other independent variables representing gender, year and plan type are constant across age quintiles.

Figure 6: Predicted Health Spending by Age Quintiles.



Note: Own calculations based on German health insurance data.

However, Figure 1 clearly shows different slopes for males and females, suggesting that the Pareto exponent changes at least across gender. This feature is not coherent with quantile regression and hence our proposed MLE is more suitable, at least in our data.

Coming back to the second question about tail heaviness, we can estimate the unconditional Pareto exponent of Y_i by existing estimators. Common choices include Hill (1975) and Gabaix and Ibragimov (2011). The slope in Figure 1 is also a consistent estimator of α. As mentioned earlier, the second moment $\mathbb{E}[Y_i^2]$ is infinite if the true α is less than two. Therefore, Figure 1 raises the concern of an infinite variance and the performance of OLS since the slope is merely above two. Sasaki and Wang (2022) provide a formal test about the finite moment condition.

As a final remark, our proposed MLE is specially designed for learning about the *tail* properties of medical expenditures, but not the whole sample. For the non-tail observations satisfying $Y_i < y_{\min}$, the support and the variance are naturally bounded and hence the least squares methods could perform well.

We summarize these discussions in the following guidance for an empirical implementation.

Guidance for Empirical Implementation

Step 1 Given any potentially skewed dataset, e.g. on claims or incomes, run the log-size-log-rank plot as in Figure 1. A linear fit in the tail suggests that the underlying distribution has a Pareto-type tail.

Step 2 If the slope of the linear fit and other estimators of the Pareto exponent such as Hill (1975) and Gabaix and Ibragimov (2011) are below (or approximately) two, the underlying distribution might have a heavy tail and the population variance of expenditure might be infinite.

Step 3 Select the tail part of the data by $Y_i > y_{\min}$ and the associated X_i. Run our proposed MLE as in Section 2.3 to estimate the conditional expectation $\mathbb{E}[Y_i|Y_i > y_{\min}, X_i]$.

Step 4 For the non-tail part, estimate the three-part model as described in Section 2.4.

5 Conclusion

Health expenditure data typically involve extreme outliers. They represent heavy tail features in the underlying distribution of the data. Simple truncation of such extreme values could lead to substantial bias when estimating any type of effect on health expenditures. In general, extreme data values can be a threat to the commonly adopted least squares methods.

In this paper, using simulation studies, we first show that when the underlying distribution has heavy tails, then the commonly used OLS and GLS methods may suffer from large biases; further, the corresponding confidence intervals may be too wide to be informative. Second, to accommodate extreme values, we propose the use of a new econometric method that allows us to recover information about the right tail of a health expenditure distributions, which is entirely ignored in many standard approaches such as top-coding. Third, we apply the proposed method to high-quality claims data from one of the biggest German private health insurers with half a million policyholders.

We estimate marginal effects of exogenous age predictors that substantially differ from those of the potentially biased least squares methods. In general, OLS tends to underestimate the age gradient in health spending. However, both estimators require careful consideration of functional form assumptions. In a next step, we extend the standard two-part model and propose a novel three-part model to model health expenditure distributions. Finally, we provide guidance and a cook-book recipe for applied economists on how to test for heavy tail features, and how to implement our proposed method.

References

Angrist, J. D. and J.-S. Pischke (2008): "Mostly harmless econometrics," in *Mostly Harmless Econometrics*, Princeton university press.

Atal, J., H. Fang, M. Karlsson, and N. R. Ziebarth (2019): "Exit, voice or loyalty? An investigation into mandated portability of front-loaded private health plans," *Journal of Risk and Insurance*, 86, 697–727.

——— (2023): "Long-term health insurance: Theory meets evidence," Tech. Rep. 26870, https://www.nber.org/papers/w26870, retrieved June 22, 2023.

Buntin, M. B. and A. M. Zaslavsky (2004): "Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures," *Journal of Health Economics*, 23, 525–542.

de Haan, L. and A. Ferreira (2006): *Extreme value theory: An introduction*, Springer Series in Operations Research and Financial Engineering, NY: Springer.

Deb, P., E. C. Norton, and W. G. Manning (2017): *Health econometrics using Stata*, Stata Press College Station, TX.

Deb, P. and P. Trivedi (1997): "Demand for medical care by the elderly: A finite mixture approach," *Journal of Applied Econometrics*, 12, 313–336.

——— (2002): "The structure of demand for health care: Latent class versus two-part models," *Journal of Health Economics*, 21, 601–625.

Duan, N., W. Manning, C. Morris, and J. Newhouse (1982): "A Comparison of Alternative Models for the Demand for Health Care," *Rand Corp. Publication R-2754-HHS*.

Finkelstein, A. (2020): "A strategy for improving US health care delivery—conducting more randomized, controlled trials," *New England Journal of Medicine*, 382, 1485–1488.

French, E. and E. Kelly (2016): "Medical spending around the developed world," *Fiscal Studies*, 37, 327–344.

Gabaix, X. (2009): "Power laws in economics and finance," *Annual Review of Economics*, 1, 255–293.

——— (2016): "Power laws in economics: An introduction," *Journal of Economic Perspectives*, 30, 185–206.

Gabaix, X. and R. Ibragimov (2011): "Rank-1/2: a simple way to improve the OLS estimation of tail exponents," *Journal of Business Economics and Statistics*, 29, 24–39.

Gilleskie, D. B. and T. A. Mroz (2004): "A flexible approach for estimating the effects of covariates on health expenditures," *Journal of health economics*, 23, 391–418.

Gotthold, K. and B. Gräber (2015): "So kommen Sie zurück in die gesetzliche Kasse," https://www.welt.de/finanzen/verbraucher/article138148142/So-kommen-Sie-zurueck-in-die-gesetzliche-Kasse.html, last retrieved on October 18, 2018.

Handel, B. R., J. T. Kolstad, and J. Spinnewijn (2019): "Information frictions and adverse selection: Policy interventions in health insurance markets," *Review of Economics and Statistics*, 101, 326–340.

Hill, B. M. (1975): "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 3, 1163–1174.

Jones, A. M. (2011): "Models for health care," in *The Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, Oxford University Press, chap. 44, 473–480.

Jones, A. M., J. Lomas, P. Moore, N. Rice, et al. (2013): "A quasi-Monte Carlo comparison of developments in parametric and semi-parametric regression methods for heavy tailed and non-normal data: with an application to healthcare costs," *Health Econometrics and Data Group Working Paper*, 13, 30.

Jones, A. M., J. Lomas, and N. Rice (2014): "Applying beta-type size distributions to healthcare cost regressions," *Journal of Applied Econometrics*, 29, 649–670.

——— (2015): "Healthcare cost regressions: going beyond the mean to estimate the full distribution," *Health economics*, 24, 1192–1212.

Kaiser Family Foundation (2019): *2019 Employer health benefits survey*, https://www.kff.org/health-costs/report/2019-employer-health-benefits-survey/, retrieved on December 14, 2019.

Karlsson, M., T. J. Klein, and N. R. Ziebarth (2016): "Skewed, persistent and high before death: medical spending in Germany," *Fiscal Studies*, 37, 527–559.

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

——— (2012): "Dealing with skewed data on costs and expenditures," in *The Elgar Companion to Health Economics*, ed. by A. M. Jones, Edward Elgar Publishing, chap. 44, 473–480, 2 ed.

Manning, W. G., A. Basu, and J. Mullahy (2005): "Generalized modeling approaches to risk adjustment of skewed outcomes data," *Journal of Health Economics*, 24, 465–488.

Manning, W. G. and J. Mullahy (2001): "Estimating log models: to transform or not to transform?" *Journal of Health Economics*, 20, 461–494.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, and A. Leibowitz (1987): "Health insurance and the demand for medical care: Evidence from a randomized experiment," *American Economic Review*, 77, 251–27.

Mihaylova, B., A. Briggs, A. O'Hagan, and S. G. Thompson (2011): "Review of statistical methods for analysing healthcare resources and costs," *Health Economics*, 20, 897–916.

Mikosch, T. (1999): *Regular variation, subexponentiality and their applications in probability theory*, vol. 99, Eindhoven University of Technology Eindhoven, The Netherlands.

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

——— (2009): "Econometric modeling of health care costs and expenditures: a survey of analytical issues and related policy considerations," *Medical care*, 47, S104–S108.

Müller, U. K. and Y. Wang (2017): "Fixed-k asymptotic inference about tail properties," *Journal of the American Statistical Association*, 112, 1334–1343.

Newhouse, J. P. and C. E. Phelps (1976): "New Estimates of Price and Income Elasticities of Medical Care Services," in *The Role of Health Insurance in the Health Services Sector*, National Bureau of Economic Research, Inc, NBER Chapters, chap. 7, 261–320.

Pohlmeier, W. and V. Ulrich (1995): "An econometric model of the two-part decision-making process in the demand for health care," *Journal of Human Resources*, 30, 339–36.

Sasaki, Y. and Y. Wang (2022): "Diagnostic testing of finite moment conditions for the consistency and root-n asymptotic normality of the gmm and m estimators," *Journal of Business & Economic Statistics (forthcoming)*.

Smith, R. L. (1987): "Estimating tails of probability distributions," *Annals of Statistics*, 15, 1174–1207.

Wang, H. and C.-H. Tsai (2009): "Tail index regression," *Journal of the American Statistical Association*, 104, 1233–1240.

Wang, H. J. and D. Li (2013): "Estimation of extreme conditional quantiles through power transformation," *Journal of the American Statistical Association*, 108, 1062–1074.

Appendix

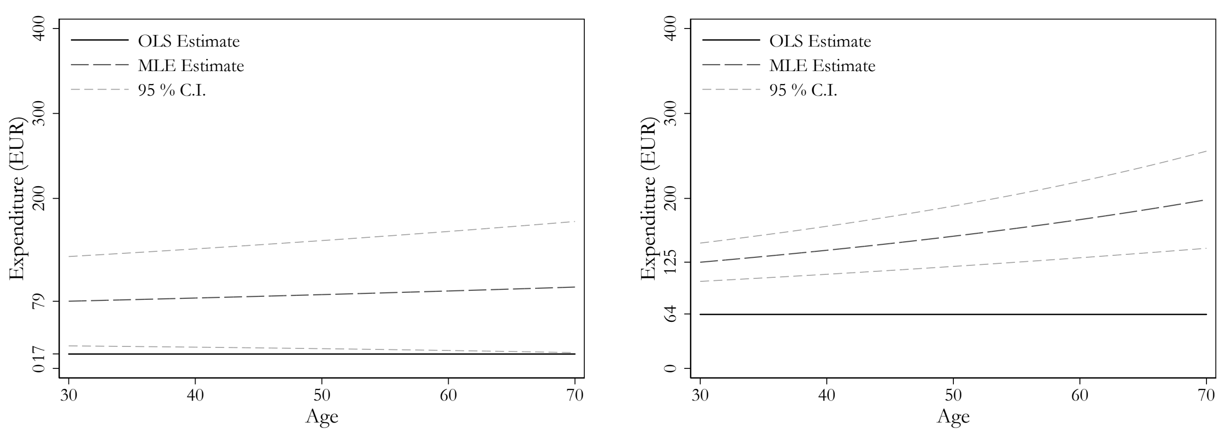
The Appendix consists of three sections. Appendix A1 contains additional empirical robustness checks. Appendix A2 contains additional simulation results. Appendix A3 contains econometric details about the proposed estimator.

A1 Robustness Checks

We first check the robustness to the choice of $y_{\min}$. In Section 4 of the main text we use the top 5% percentile, and in this section we use top 3% and 7% percentiles. Figure A1 depicts the estimates of the marginal effects based on OLS and our proposed MLE. Figures A2 and A3 repeat the analysis with female and male subsamples. The findings are similar to those reported in Section 4. In particular, the OLS estimates are substantially below the MLE regardless of the subsample and the choice of $y_{\min}$.

Figure A1: Marginal Age Effects for Different Cutoffs, OLS versus MLE.

(a) Top 3% (b) Top 7%

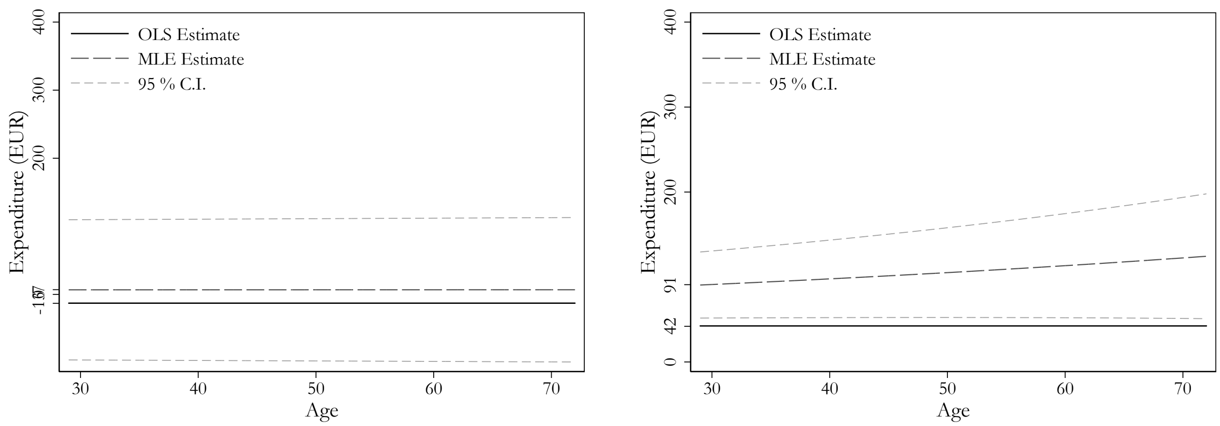


Notes: Own calculations based on German health insurance data. The graph compares the marginal age effects on health care spending, along with 95% confidence intervals.

Second, we repeat the exercise in Section 4.2 by splitting the sample by gender. Recall

Figure A2: Marginal Age Effects for Different Cutoffs – Females. OLS versus MLE.

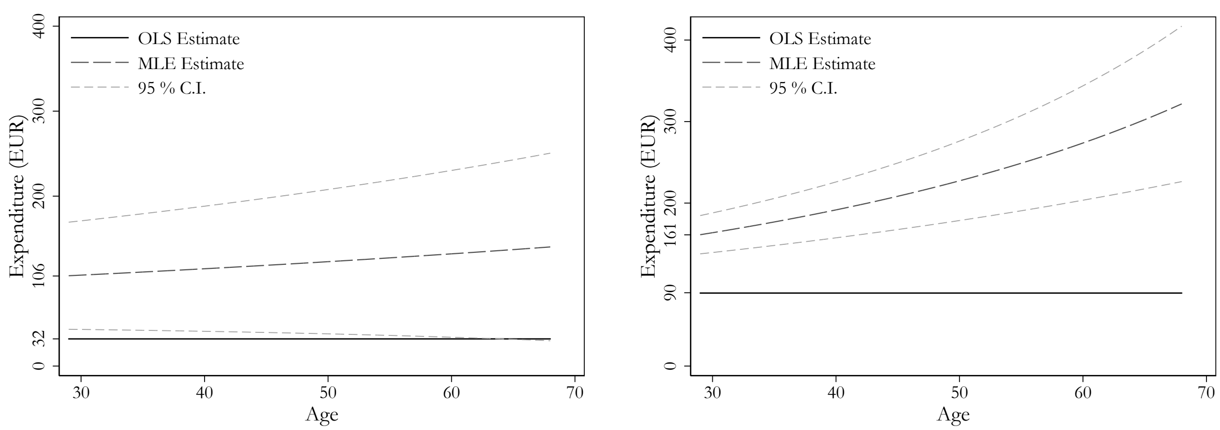
(a) Top 3% (b) Top 7%



Notes: Own calculations based on German health insurance data. The graph compares the marginal age effects on health care spending, along with 95% confidence intervals.

Figure A3: Marginal Age Effects for Different Cutoffs – Males. OLS versus MLE.

(a) Top 3% (b) Top 7%

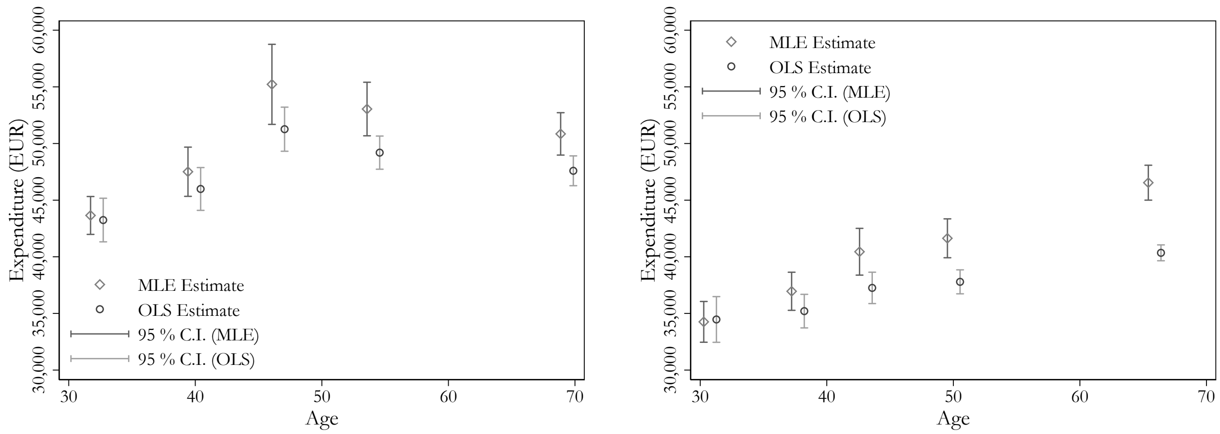


Notes: Own calculations based on German health insurance data. The graph compares the marginal age effects on health care spending, along with 95% confidence intervals.

that age is replaced with five dummy variables presenting the quintiles. Figure A4 depicts the estimated health spending at different age quintiles based on the proposed MLE and the classic OLS methods. The OLS estimates and confidence intervals are again below their MLE counterparts as we find in Section 4.2.

Figure A4: Predicted Health Spending by Age Quintiles

(a) Females (b) Males



Notes: Own calculations based on German health insurance data.

A2 Additional Simulations

In this section, we conduct simulation studies of the OLS estimator with log-transformed data. The data are still generated from

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \qquad \text{(A2.1)}$$

with $(\beta_0, \beta_1) = (1, 1)$. To make sure that Y_i is positive, we generate X_i from the absolute value of the standard normal distribution, and u_i from the standard Pareto distribution with exponent $1/\xi$, that is, $\mathbb{P}(u_i \geq u) = (1 + \xi u)^{-1/\xi}$ for $u > 0$. The other model specifications are the same as in Tables 1 and 2.

Note that the OLS estimator of β_1—when regressing $\ln(Y_i)$ on X_i (and a constant)—is *not* measuring the marginal effect of X_i on Y_i, given the nonlinear setup. To make a reasonable comparison, we treat $\bar{Y}\hat{\beta}_1$ as the estimator of the marginal effect where $\bar{Y}$ denotes the sample average of Y_i and compares it with the true parameter β_1. The standard error of the estimator of the marginal effect is adjusted accordingly by multiplying $\bar{Y}$ to that of $\hat{\beta}_1$. Since the marginal effect depends on the value of X_i, this estimator essentially estimates the average marginal effect over X_i. We also implement the estimator for $X_i = 1$ with $\beta_1 = 1$ as in Figures 2 and 3. The results are very similar and hence omitted.

Table A1 depicts the performance of such as marginal effect estimator in terms of MAD, RMSE, average rejection probability of the standard t-test, and the average length of the standard 95% confidence intervals.

Table A1: Log-Y Simulation Results with Generalized Pareto Distribution

n	500	1000	5000	10^4	10^5	10^6	500	1000	5000	10^4	10^5	10^6
$\xi(1/\alpha)$	Panel A: MAD						Panel B: RMSE					
0.09	0.07	0.06	0.05	0.05	0.05	0.05	0.09	0.07	0.05	0.05	0.05	0.05
0.19	0.09	0.08	0.08	0.08	0.08	0.08	0.11	0.10	0.08	0.08	0.08	0.08
0.29	0.13	0.12	0.12	0.12	0.12	0.12	0.16	0.14	0.13	0.12	0.12	0.12
0.39	0.19	0.18	0.18	0.18	0.18	0.18	0.22	0.20	0.19	0.19	0.18	0.18
0.49	0.28	0.27	0.27	0.27	0.27	0.27	0.33	0.30	0.28	0.28	0.27	0.27
0.59	0.41	0.41	0.41	0.41	0.41	0.41	0.50	0.46	0.42	0.42	0.41	0.41
0.69	0.66	0.64	0.64	0.66	0.64	0.64	1.57	1.07	0.81	1.88	0.68	0.65
0.79	1.61	1.03	1.23	1.11	1.09	1.10	47.9	2.65	11.2	4.48	1.27	1.29
0.89	1.88	2.05	2.13	2.12	2.07	2.26	8.90	19.7	9.34	11.8	3.48	6.63
0.99	4.43	4.37	6.47	9.83	6.04	6.02	65.5	61.5	129	342	104	37.6
$\xi(1/\alpha)$	Panel C: Rejection Prob.						Panel D: Length of 95% CI					
0.09	0.10	0.15	0.57	0.86	1.00	1.00	0.27	0.19	0.09	0.06	0.02	0.01
0.19	0.15	0.27	0.87	0.99	1.00	1.00	0.32	0.23	0.10	0.07	0.02	0.01
0.29	0.24	0.44	0.98	1.00	1.00	1.00	0.37	0.27	0.12	0.08	0.03	0.01
0.39	0.35	0.62	1.00	1.00	1.00	1.00	0.45	0.32	0.14	0.10	0.03	0.01
0.49	0.49	0.77	1.00	1.00	1.00	1.00	0.54	0.38	0.17	0.12	0.04	0.01
0.59	0.61	0.89	1.00	1.00	1.00	1.00	0.68	0.48	0.22	0.15	0.05	0.02
0.69	0.73	0.95	1.00	1.00	1.00	1.00	0.91	0.63	0.28	0.20	0.06	0.02
0.79	0.81	0.98	1.00	1.00	1.00	1.00	1.94	0.88	0.43	0.29	0.09	0.03
0.89	0.86	0.99	1.00	1.00	1.00	1.00	2.02	1.51	0.68	0.48	0.15	0.05
0.99	0.90	1.00	1.00	1.00	1.00	1.00	4.63	3.03	1.84	1.94	0.38	0.12

Notes: The table depicts the average mean absolute deviation (MAD), the average root mean squared error (RMSE), the average rejection probability of the standard t-test, and the average length of the standard 95% confidence intervals. The results are based on 10000 simulation draws. See the main text for details about the data generating process.

The results can be summarized as follows. First, the large bias and RMSE indicate that the misspecification error of approximating the linear model with a heavy-tailed error term by the log-linear model could be substantial. Such misspecification error is small when the tail is thin, i.e, ξ is close to zero. This is what has been documented in the existing literature. However, when the tail of the error term becomes heavy, the misspecification error is amplified substantially.

Second, accordingly, the rejection probability of the t-test becomes substantially larger than the nominal 5% level (Panel C). When ξ is above 0.5, the variance of the error term is not well-defined and hence the t-test is hardly informative.

Third, although the length of the confidence interval is shrinking with the sample size, the bias and the overrejection do not (Panel D). This suggests that the estimator substantially deviates from the true value with the bias strictly dominating the randomness.

A3 More Details about the MLE

Our maximum likelihood estimator is based on the tail index regression proposed by Wang and Tsai (2009). The key condition is that the conditional distribution of Y_i on $X_i = x$ has an approximate Pareto tail. In particular, we assume that uniformly over x,

$$1 - \mathbb{P}\left(Y_i > y | X_i = x\right) = c\left(x\right) y^{-\alpha(x)} \left(1 + o(1)\right) \text{ as } y \to \infty, \tag{A3.1}$$

for some constant $c\left(x\right) > 0$ and $\alpha\left(x\right) > 0$. This condition is mild and satisfied by many commonly used distributions such as Student-t, Gamma, F distributions. In particular, if Y is Student-t distributed conditional on $X = x$ with $v\left(x\right)$ degrees of freedom, then $\alpha\left(x\right)$ is simply $v\left(x\right)$. Chapter 1 in de Haan and Ferreira (2006) provides a complete review of the literature.

Note that the condition (A3.1) requires that the right tail of the conditional distribution is well approximated by a Pareto distribution with component $\alpha(x)$. Such an approximation becomes more accurate as we move further towards the tail (i.e., $y \to \infty$). In this sense, we consider our method a semiparametric method that does not hinge on any specific distribution. This is important for empirical applications as, *a priori*, researchers do not know the true health expenditure distribution.

Under Condition (A3.1), we can further approximate the conditional probability density of Y given X and $Y > y_{\min}$ for some large tail threshold $y_{\min}$ by $\alpha(x)(y/y_{\min})^{-\alpha(x)} y^{-1}$. This leads to the negative likelihood function in Equation (2.3). Under (A3.1) and some additional technical conditions, Wang and Tsai (2009) establish that by solving (2.3), the MLE $\hat{\beta}$ is consistent and asymptotically normal. The standard error can be estimated by (2.4).

As a tuning parameter, the econometrician chooses the tail threshold $y_{\min}$ which affects the estimation result, especially when the sample size is only moderate. However, the optimal selection of $y_{\min}$ is challenging and has stimulated a large literature in statistics and econometrics. On the one hand, a large $y_{\min}$ ensures that the tail Pareto approximation performs well, and hence the bias is small. On the other hand, a small $y_{\min}$ ensures enough tail observations for asymptotic normality, and hence the variance is small.

This bias-variance trade-off indicates a delicate balance in the choice of $y_{\min}$ (and equivalently the number of tail observations n_0). It turns out that a theoretically optimal choice of $y_{\min}$ does not exist if no other condition is imposed on the true underlying distribution (Müller and Wang, 2017). Therefore, we recommend to vary $y_{\min}$ in sensitivity analysis. Wang and Tsai (2009) also provide a data-driven method of choosing $y_{\min}$, whose theoretical properties need further investigation.