Modeling Healthcare Expenditures

Anirban Basu, PhD
The University of Chicago & The NBER

HERC CyberSeminar
June 16, 2010
Goals

- Review parametric/semi-parametric statistical models for health care expenditures
- Intent to study how mean costs responds to shifts in covariate levels
- Will assume “selection of observable” throughout
- Main focus is on bias, will look at some efficiency issues too.
Healthcare costs in the U.S.
Healthcare costs in the U.S.

- A substantial fraction of people with no health care costs
- Handful of people with enormous yearly expenditures
- Often top 1% accounts for 25% of the total mean, sometimes the top 10% accounts for half of all costs

Any attempt to study how the mean of such a distribution responds to a covariate would require careful attention to how different parts of the distribution moves in response to the changes in levels of that covariate
Classes of models:

- Single equation models
- Multiple equation models
  - Two-part models
  - Multi-part models
  - Conditional density estimation
- Mixture models
- Propensity scores
- Doubly robust models
Outline

- **Single-equation models**: relate mean of costs to covariates using a single functional form
  - Linear models
  - Transformation models
    - Log-OLS
    - Box-Cox
  - Generalized Linear models
- **Application to datasets of ~1000 to 100,000**
- **Discuss goodness-of-fit tests**
Y = Costs
X = Covariates = (X_T, X_{-T})
X_T = A binary treatment variable
\beta = regression coefficients
\mu = E(Y); \mu(x) = E(Y|X=x)
Marginal effect: d\mu(X)/dX
Incremental effect: \mu(X=1) - \mu(X=0)
An illustrative health-economic study

- Compare the 2-year cost of care under aggressive treatments (radiation or surgery) versus watchful waiting following diagnosis of prostate cancer
- Use SEER-Medicare linked database
- Data: diagnosis between 1995-2002
- Follow-up through 2004.
- No censoring issues
- Ignore issues with death
The parameters of Interest

Main parameters of interest:

- \( \mu(x) = E(y \mid x) \) \( y = \) Expenditures
- Incremental effect of aggressive treatment vs WW on total expenditures \( (\Delta_T) \)
- Avg. Costs for TRT – Avg. Costs for WW

\[
\Delta_T = E_{X_T=1 \mid X_T=1} \left\{ \mu(X_T = 1, X_T) - \mu(X_T = 0, X_T) \right\}
\]

- The interpretation of regression coefficients \( \beta \) is of secondary concern.
Single equation models: Linear model

\[ Y = X\beta + \varepsilon \]
Single equation models: Linear model

- $Y = X\beta + \varepsilon$
- Assumes parallel shift for the entire distribution across levels of any one $X$, and for all $X$'s
- Biased
  - the non-linearity in the response,
  - the instability caused by skewness and kurtosis,
- Inefficient
  - common failure to deal with the heteroscedasticity (variance increases with the mean)
Single equation models: Transformation models – log OLS

\[ \ln(Y) = X\beta + \varepsilon \]
Single equation models: Transformation models – log OLS

\[ \ln(Y) = X\beta + \varepsilon \]

- Assumes proportional shift for the entire distribution across levels of any one X, and for all X’s.

Advantages:
- Usually overcomes skewness issues
- Allows for additive effects in the log-scale, which translates to non-additive effects in the raw scale
- Reduces problems with heteroscedasticity & kurtosis
Transformation models

- Scale of estimation ≠ scale of interest
  - If modeling $, results in log-$

- Model for geometric mean, not arithmetic mean
  - Models $E\{\log(y)\}$ and not $\log(E\{y\})$
Single equation models: Transformation models – log OLS

- \( \ln(Y) = X\beta + \varepsilon \)
- Disadvantages:
  - Creates a model for log-\( Y \), not \( Y \).
  - \( \mathbb{E}(\ln Y) \neq \ln(\mathbb{E}(Y)) \)
  - Require retransformation, \( \mathbb{E}(Y) = \exp(X\beta + \varepsilon) \)
    \[ = \exp(X\beta)^s \]
  - Duan’s smearing estimator provides an estimate for \( s \)
  - \( \hat{s} = \text{Mean}\{\exp(\ln Y - X\hat{\beta})\} \)
Single equation models: Transformation models – log OLS

- \( \ln(Y) = X\beta + \varepsilon \)
- Disadvantages:
  - \( E(\ln Y) = X\beta \)
  - \( \ln E(Y) = X\beta + \ln \{s\} \)
  - If \( s = s(X) \),
    \[ \frac{d\ln E(Y)}{dX} = \beta + \frac{d\ln \{s(X)\}}{dX} \]
  - So, common interpretation of \( \beta \) no longer applies

- If \( \varepsilon \sim \text{Normal}(0, \sigma^2) \), \( \ln(s) = 0.5 \sigma^2 \)
- With log-scale heteroscedasticity,
  \[ \text{Bias} = 0.5d\sigma^2(X)/dX \]
Single equation models: Transformation models – log OLS

Table 1
Univariate statistics on log medical expenditures for users* health insurance experiment

<table>
<thead>
<tr>
<th>Plan</th>
<th>Mean*</th>
<th>Variance**</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>25%</td>
<td>6.128</td>
<td>2.133</td>
<td>0.456</td>
<td>3.539</td>
<td>4.393</td>
</tr>
<tr>
<td>50%</td>
<td>6.039</td>
<td>2.003</td>
<td>0.513</td>
<td>4.067</td>
<td>4.265</td>
</tr>
<tr>
<td>95%</td>
<td>5.998</td>
<td>2.343</td>
<td>0.464</td>
<td>3.226</td>
<td>4.158</td>
</tr>
</tbody>
</table>

* In 1995 US$, adjusted using the medical care component of the Consumer Price Index.
** Plan differences are significant at $F < 0.001$ for both mean and variance, by $F$-tests. Tests for variance based on Park Test.

Manning JHE, 1998

% change in costs between Free & 95%:

From log-OLS regression: $6.349 - 5.998 = 0.351$ (or 35.1%)

Truth (assuming log-normality) = $6.349 - 5.998 + 0.5 \times (2.083 - 2.343) = 0.221$ (or 22.1%)
Plan specific smearing estimate can address bias.

Difficult to come up with appropriate smearing estimator when there are multiple X’s that lead to heteroscedasticity.

Park Test – test to determine heteroscedasticity in log-scale:

- Use squared log-scale residuals as an estimator for variance.
- Study if X’s predict this estimated variance.
- `reg lny x, robust`
- `predict r, res`
- `gen r2= r^2`
- `glm r2 x, link(log) family(gamma) robust`
Let’s try out in Stata
**MODELLING HEALTHCARE EXPENDITURES**

```plaintext
. use week2_datal, clear
. summ totexp, d

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>totexp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>78.32</td>
</tr>
<tr>
<td>5%</td>
<td>907.92</td>
</tr>
<tr>
<td>10%</td>
<td>2702.61</td>
</tr>
<tr>
<td>25%</td>
<td>8912.18</td>
</tr>
<tr>
<td>50%</td>
<td>14536.54</td>
</tr>
<tr>
<td>75%</td>
<td>23375.21</td>
</tr>
<tr>
<td>90%</td>
<td>41139.5</td>
</tr>
<tr>
<td>95%</td>
<td>59268.49</td>
</tr>
<tr>
<td>99%</td>
<td>112434.3</td>
</tr>
</tbody>
</table>

Smallest: 2.78
Largest: 279487.5
Obs: 7129
Sum of Wgt.: 7129

Mean: 20265.17
Std. Dev.: 22937.29
Variance: 5.26e+08
Skewness: 4.537538
Kurtosis: 39.71721
```
. kdensity totexp, saving(grp1, replace)
. help aggr_trt: sum totexp, d

-> aggr_trt = 0

totexp

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>50.8</td>
</tr>
<tr>
<td>5%</td>
<td>648.2</td>
</tr>
<tr>
<td>10%</td>
<td>1423.23</td>
</tr>
<tr>
<td>25%</td>
<td>4738.09</td>
</tr>
<tr>
<td>50%</td>
<td>12317.91</td>
</tr>
<tr>
<td>75%</td>
<td>23087.59</td>
</tr>
<tr>
<td>90%</td>
<td>43853.15</td>
</tr>
<tr>
<td>95%</td>
<td>65831.95</td>
</tr>
<tr>
<td>99%</td>
<td>114372.3</td>
</tr>
</tbody>
</table>

Obs: 2219
Sum of Wgt.: 2219
Mean: 19201.76
Std. Dev.: 24102.25
Variance: 5.81e+08
Skewness: 3.529187
Kurtosis: 24.20243
- aggr_trt = 1

**totexp**

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>99.18</td>
</tr>
<tr>
<td>5%</td>
<td>1172.94</td>
</tr>
<tr>
<td>10%</td>
<td>5141.575</td>
</tr>
<tr>
<td>25%</td>
<td>10071</td>
</tr>
<tr>
<td>50%</td>
<td>15343.07</td>
</tr>
<tr>
<td>75%</td>
<td>23522.77</td>
</tr>
<tr>
<td>90%</td>
<td>39927.79</td>
</tr>
<tr>
<td>95%</td>
<td>57092.22</td>
</tr>
<tr>
<td>99%</td>
<td>110871.9</td>
</tr>
</tbody>
</table>

Obs: 4910
Sum of Wgt.: 4910

Mean: 20745.77
Largest: 263867.3
Std. Dev.: 22376.88
Variance: 5.01e+08
Skewness: 5.114373
Kurtosis: 48.88715

.di (20745.77 - 19201.76)/ 19201.76

.08040982
. /* OLS */
.
. reg totexp aggr_trt, robust

Linear regression

|            | Coef. | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-------------|-------|-----------|-------|-------|----------------------|
| totexp      |       |           |       |       |                      |
| aggr_trt    | 1544.008 | 603.106   | 2.56  | 0.010 | 361.7408 2726.274    |
| _cons       | 19201.76  | 511.6135  | 37.53 | 0.000 | 18198.85 20204.67    |
. `reg` `totexp` `aggr_trt` `agec` `white` `black` `single` `married` `well` `mod` `stage1` `stage2` `lnc`, `robust`

|       |    Coef.   |      Std. Err. |      t    |     P>|t|     |      [95% Conf. Interval] |
|-------|------------|----------------|-----------|---------|---------------------------|
| totexp |            |                |           |         |                           |
| aggr_trt | 3490.596   | 636.6122       | 5.48      | 0.000   | 2242.646                 |
| agec   | 205.2538   | 52.4853        | 3.91      | 0.000   | 102.367                  |
| white  | 1269.85    | 1027.947       | 1.24      | 0.217   | -745.2315                |
| black  | 3706.038   | 1426.357       | 2.60      | 0.009   | 909.9544                 |
| single | 946.6497   | 1241.496       | 0.76      | 0.446   | -1487.051                |
| married| -1048.995  | 696.5203       | -1.51     | 0.132   | -2414.382                |
| well   | -4717.357  | 1261.522       | -3.74     | 0.000   | -7190.314                |
| mod    | -2995.753  | 745.5633       | -4.02     | 0.000   | -4457.279                |
| stage1 | -304.7167  | 602.6593       | -0.51     | 0.613   | -1486.108                |
| stage2 | -1803.271  | 804.8815       | -2.24     | 0.025   | -3381.078                |
| lnc    | 7871.044   | 502.2525       | 15.67     | 0.000   | 6886.48                  |
| _cons  | 15268.15   | 1349.965       | 11.31     | 0.000   | 12621.82                 |

Slide 24
. /* Log-OLS */
. capture drop lny
. gen lny = ln(totexp)
. reg lny aggr_trt, robust

Linear regression       Number of obs =    7129
                       F(  1,  7127) =   93.48
                       Prob > F      =  0.0000
                       R-squared     =  0.0150
                       Root MSE      =  1.2894

-------------------------------------------------------------------------------
|               Robust          lny |      Coef.   Std. Err.      t    P>|t|    [95% Conf. Interval] |
-------------+---------------------------------------------------------------
           aggr_trt |   .3431855    .035495     9.67   0.000    .2736047    .4127662 |
          _cons     |  9.147605    .0310711   294.41   0.000    9.086697    9.208514 |
-------------------------------------------------------------------------------

. predict xb, xb
. gen mu = exp(xb)
. summ mu totexp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>7129</td>
<td>12040.35</td>
<td>1780.561</td>
<td>9391.919</td>
<td>13237.27</td>
</tr>
<tr>
<td>totexp</td>
<td>7129</td>
<td>20265.17</td>
<td>22937.29</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>
** Overall smearing
  . drop mu
  . gen smr = exp(lny - xb)
  . summ smr

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>smr</td>
<td>7129</td>
<td>1.7158</td>
<td>2.016464</td>
<td>.00021</td>
<td>33.72997</td>
</tr>
</tbody>
</table>

. gen smear = r(mean)
. gen mu = exp(xb)*smear
. summ mu totexp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>7129</td>
<td>20658.62</td>
<td>3055.054</td>
<td>16114.49</td>
<td>22712.27</td>
</tr>
<tr>
<td>totexp</td>
<td>7129</td>
<td>20265.17</td>
<td>22937.29</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>
. bysort aggr_trt: summ mu totexp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>2219</td>
<td>16114.49</td>
<td>0</td>
<td>16114.49</td>
<td>16114.49</td>
</tr>
<tr>
<td>totexp</td>
<td>2219</td>
<td>19201.76</td>
<td>24102.25</td>
<td>3.63</td>
<td>316789.1</td>
</tr>
</tbody>
</table>

-> aggr_trt = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>4910</td>
<td>22712.27</td>
<td>0</td>
<td>22712.27</td>
<td>22712.27</td>
</tr>
<tr>
<td>totexp</td>
<td>4910</td>
<td>20745.77</td>
<td>22376.88</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>

/* Notice difference in log scale Variance */

. bysort aggr_trt: summ lny

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lny</td>
<td>2219</td>
<td>9.147605</td>
<td>1.463769</td>
<td>1.289233</td>
<td>12.66599</td>
</tr>
</tbody>
</table>

-> aggr_trt = 0

| lny       | 4910 | 9.490791     | 1.202411  | 1.022451 | 12.88387 |
. gr twoway (kdensity lny if aggr_trt==0) (kdensity lny if aggr_trt==1), saving
. ***  Trt specific smearing
. drop mu
. summ smr if aggr_trt ==0

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>smr</td>
<td>2219</td>
<td>2.044498</td>
<td>2.566275</td>
<td>.0003865</td>
<td>33.72997</td>
</tr>
</tbody>
</table>

. gen smear0=r(mean)
. summ smr if aggr_trt ==1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>smr</td>
<td>4910</td>
<td>1.567224</td>
<td>1.690446</td>
<td>.00021</td>
<td>29.75754</td>
</tr>
</tbody>
</table>

. gen smear1=r(mean)
. gen mu =exp(xb)*smear0 if aggr_trt ==0
. replace mu =exp(xb)*smear1 if aggr_trt ==1

. summ mu totexp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>7129</td>
<td>20265.17</td>
<td>714.9419</td>
<td>19201.76</td>
<td>20745.77</td>
</tr>
<tr>
<td>totexp</td>
<td>7129</td>
<td>20265.17</td>
<td>22937.29</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>
. bysort aggr_trt: summ mu totexp

aggr_trt = 0

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>2219</td>
<td>19201.76</td>
<td>0</td>
<td>19201.76</td>
<td>19201.76</td>
</tr>
<tr>
<td>totexp</td>
<td>2219</td>
<td>19201.76</td>
<td>24102.25</td>
<td>3.63</td>
<td>316789.1</td>
</tr>
</tbody>
</table>

aggr_trt = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>4910</td>
<td>20745.77</td>
<td>0</td>
<td>20745.77</td>
<td>20745.77</td>
</tr>
<tr>
<td>totexp</td>
<td>4910</td>
<td>20745.77</td>
<td>22376.88</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>
. **** ln-OLS regression
. drop mu xb
. reg lny aggr_trt agec white black single married well mod stage1 stage2 lnc, robust

|                  | Coef.  | Std. Err. |  t    | P>|t|  | [95% Conf. Interval] |
|------------------|--------|-----------|-------|------|----------------------|
| lny              |        |           |       |      |                      |
| aggr_trt         | .460918 | .0372838  | 12.36 | 0.000 | .3878303   .5340052 |
| agec             | .0166473 | .0027592   | 6.03  | 0.000 | .0112385  .0220561  |
| white            | .2875066 | .0703636  | 4.09  | 0.000 | .1495731  .4254401  |
| black            | .239347  | .0876897   | 2.73  | 0.006 | .0674492  .4112448  |
| single           | .0702465 | .066109    | 1.06  | 0.288 | -.0593468  .1998397 |
| married          | -.0016466 | .0381762  | -0.04 | 0.966 | -.0764833  .07319   |
| well             | -.3312179 | .0710317  | -4.66 | 0.000 | -.4704613  -.1919746 |
| mod              | -.1428405 | .0378877  | -3.77 | 0.000 | -.2171118  -.0685693 |
| stage1           | -.0147308 | .0337002  | -0.44 | 0.662 | -.0807933  .0513317 |
| stage2           | -.0830997 | .0480434  | -1.73 | 0.084 | -.177279  .0110797  |
| lnc              | .5280045 | .0233166  | 22.65 | 0.000 | .4822971  .5737119  |
| _cons            | 8.627953 | .0881616  | 97.87 | 0.000 | 8.455131  8.800776  |
. ** Park_test - detect heteroscedasticity in log-scale
. predict xb, xb
. gen res2 = (lny- xb)^2
. glm res2 aggr_trt  agec  white black single married well mod stage1 stage2 lnc, family(gamma) link(log) robust

|     | Coef.   | Std. Err. | z     | P>|z|   [95% Conf. Interval] |
|-----|---------|-----------|-------|-------|-------------------------|
| aggr_trt | -0.5202051 | 0.0652508 | -7.97 | 0.000 | -0.6480943 -0.3923159   |
| agec  | -0.0253315 | 0.0054215 | -4.67 | 0.000 | -0.0359575 -0.0147055   |
| white | -0.4874537 | 0.1051322 | -4.64 | 0.000 | -0.6935091 -0.2813983   |
| black | -0.1256915 | 0.1388546 | -0.91 | 0.365 | -0.3978414 0.1464585    |
| single| -0.0352572 | 0.1358728 | -0.26 | 0.795 | -0.3015629 0.2310485    |
| married| -0.1367799 | 0.073717  | -1.86 | 0.064 | -0.2812624 0.0077027    |
| well  | 0.100519   | 0.1393567 | 0.72  | 0.471 | -0.172615 0.373653      |
| mod   | -0.0869734 | 0.0785356 | -1.11 | 0.268 | -0.2409005 0.6699536    |
| stage1| -0.057802  | 0.0720468 | -0.80 | 0.422 | -0.1990111 0.0834072    |
| stage2| -0.0547015 | 0.1094026 | -0.50 | 0.617 | -0.2691267 0.1597236    |
| lnc   | -0.5347174 | 0.046929  | -11.39| 0.000 | -0.6266966 -0.4427382   |
| _cons | 1.597288   | 0.1376163 | 11.61 | 0.000 | 1.3275655 1.867011      |

. test
     chi2( 11) =  283.67
     Prob > chi2 =  0.0000

Slide 32
Single equation models: Transformation models – Box-Cox

\[ \frac{(Y^\lambda - 1)}{\lambda} = X\beta + \varepsilon \]

- If \( \lambda \rightarrow 0 \), Box-Cox \( \rightarrow \) log-OLS
- If \( \lambda \) is known apriori, can use Duan’s smearing estimator
- If \( \lambda \) is estimated, use different estimator

\[
E(Y|X) = (X\beta\lambda + 1)^{1/\lambda} \left\{ \frac{1 + \frac{\sigma^2(1 - \lambda)}{2(1 + X\beta\lambda)^2}}{2(1 + X\beta\lambda)^2} \right\}
\]

- All concerns about retransformation persist
g\{E(Y | X)\} = X\beta, g\{\cdot\} \text{ is a link function}

No retransformation problem as link function applies to E(Y) and NOT Y.

Estimation of GLM
- Parametric – FIML
- Semi-parametric – Quasi-likelihood

FIML:
- Gamma, Exponential, Inverse Gaussian, Poisson, Neg. Binomial
- \texttt{glm y x, link(log) family(gamma) robust}
Single equation models: Generalized Linear Models

Quasi-likelihood approach

1. Relate $\mu(X)$ to the linear predictor $X\beta$ using a link function. $g(\mu(X)) = X\beta$

2. $V(Y) \sim h(\mu)$; e.g. $V(Y) = \phi \mu^2 \Rightarrow$ Gamma Variance

3. Does not make any distributional assumption beyond 1st and 2nd moments

4. Estimate coefficients by solving quasi-score equations

$$\sum_{i=1}^{N} G_{\beta_j}^i = \sum_{i=1}^{N} (y_i - \mu_i)V_i^{-1}(\partial \mu_i / \partial \beta_j) = 0$$

• `glm y x, link(log) fam(gamma) robust irls`
Example: Gamma with log link

- \( \ln(\mu) = X\beta \) or \( \mu = \exp(X\beta) \)
- \( \nu = \theta_1 \mu^2 \)
- If mean and variance functions are correctly specified, \( \beta_1 \) provides a consistent estimator of \( \partial \ln(\mathbb{E}\{y|X\}) / \partial x_1 \)
- If the empirical distribution of \( Y \) is truly Gamma, then FIML will generate the MVUE estimator for \( \beta \)
- If the empirical distribution of \( Y \) is deviates from Gamma, but mean and variance functions are appropriately specified, quasi-likelihood will be a more robust estimator.
To Stata..
/* Generalized Linear Models */
 . glm totexp aggr_trt, link(log) family(gamma) robust

Iteration 3:   log pseudolikelihood = -77820.337

Generalized linear models
Optimization : ML
               Residual df = 7127
Scale parameter = 1.291687

Deviance = 7586.035219  1/df) Deviance = 1.064408
Pearson = 9205.852281  1/df) Pearson = 1.291687

Variance function: V(u) = u^2  [Gamma]
Link function    : g(u) = ln(u)  [Log]

AIC             =  21.83261
Log pseudolikelihood = -77820.3371  BIC             = -55644.18

-----------------------------------------------------------------------
<table>
<thead>
<tr>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>totexp</td>
</tr>
</tbody>
</table>
-----------------------------------------------------------------------
| aggr_trt  |  0.0773403 | 0.0307692  | 2.51 | 0.012 | 0.0170338    | 0.1376469 |
| _cons     |  9.862757  | 0.0266422  | 370.19 | 0.000 | 9.810539    | 9.914975  |
-----------------------------------------------------------------------
. predict mu, mu
. summ mu totexp

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>7129</td>
<td>20265.17</td>
<td>714.9419</td>
<td>19201.76</td>
<td>20745.77</td>
</tr>
<tr>
<td>totexp</td>
<td>7129</td>
<td>20265.17</td>
<td>22937.29</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>

. bysort aggr_trt: summ mu totexp

-> aggr_trt = 0

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>2219</td>
<td>19201.76</td>
<td>0</td>
<td>19201.76</td>
<td>19201.76</td>
</tr>
<tr>
<td>totexp</td>
<td>2219</td>
<td>19201.76</td>
<td>24102.25</td>
<td>3.63</td>
<td>316789.1</td>
</tr>
</tbody>
</table>

-> aggr_trt = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu</td>
<td>4910</td>
<td>20745.77</td>
<td>0</td>
<td>20745.77</td>
<td>20745.77</td>
</tr>
<tr>
<td>totexp</td>
<td>4910</td>
<td>20745.77</td>
<td>22376.88</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>
** GAMMA FIML

* glm totexp aggr_trt agec white black single married well mod stage1 stage2 lnc, link(log) family(gamma) robust

Iteration 3: log pseudolikelihood = -77563.068

|                  | Coef. | Robust Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------------------|-------|------------------|-------|------|------------------------|
| totexp           |       |                  |       |      |                        |
| aggr_trt         | 0.1794719 | 0.0320742       | 5.60  | 0.000 | 0.1166076 - 0.2423362  |
| agec             | 0.010508  | 0.0025216       | 4.17  | 0.000 | 0.0055658 - 0.0154501  |
| white            | 0.0854313 | 0.0532399       | 1.60  | 0.109 | -0.0189169 - 0.1897796 |
| black            | 0.1795773 | 0.067756        | 2.65  | 0.008 | 0.0467779 - 0.3123767  |
| single           | 0.0333528 | 0.0559583       | 0.60  | 0.551 | -0.0763235 - 0.143029  |
| married          | -0.0542054 | 0.0347019      | -1.56 | 0.118 | -0.1222198 - 0.013809  |
| well             | -0.2484179 | 0.0642366      | -3.87 | 0.000 | -0.3743194 - 0.1225165 |
| mod              | -0.1447476 | 0.0333845      | -4.34 | 0.000 | -0.21018 - 0.0793151   |
| stage1           | -0.0276048 | 0.0291643      | -0.95 | 0.344 | -0.0847657 - 0.0295562 |
| stage2           | -0.091044  | 0.0429937      | -2.12 | 0.034 | -0.1753101 - 0.0067779 |
| lnc              | 0.3698705  | 0.0214508      | 17.24 | 0.000 | 0.3278276 - 0.4119133  |
| _cons            | 9.628452   | 0.0680018      | 141.59| 0.000 | 9.495171 - 9.761733    |
. ** GAMMA QMLE

. glm totexp aggr_trt agec white black single married well mod stage1 stage2 lnc, link(log) family(gamma) robust irls

Iteration 6:   deviance =  7071.498

|            | Coef.   | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|------------|---------|-----------|------|------|----------------------|
| totexp     |         |           |      |      |                      |
| aggr_trt   | .1794717| .0327351  | 5.48 | 0.000| .1153121 .2436313    |
| agec       | .010508 | .0025526  | 4.12 | 0.000| .005505 .015511     |
| white      | .0854311| .0535442  | 1.60 | 0.111| -.0195136 .1903758  |
| black      | .1795774| .0681181  | 2.64 | 0.008| .0460683 .3130864   |
| single     | .033353 | .0559614  | .60  | 0.551| -.0763292 .1430353  |
| married    | -.0542054| .0349642 | -1.55| 0.121| -.1227339 .0143232 |
| well       | -.2484182| .0648256 | -3.83| 0.000| -.375474 -.1213623  |
| mod        | -.1447475| .0337717 | -4.29| 0.000| -.2109388 -.0785563 |
| stage1     | -.0276047| .0292326 | -0.94| 0.345| -.0848996 .0296902  |
| stage2     | -.0910438| .0436959 | -2.08| 0.037| -.1766862 -.0054015 |
| lnc        | .3698705 | .0216682 | 17.07| 0.000| .3274016 .4123394   |
| _cons      | 9.628452 | .0676074 | 142.42| 0.000| 9.495944 9.76096   |
Method of recycled predictions

- In order to calculate effect in the additive scale
- Estimate parameters in 
  \[ \exp(\beta_0 + \beta_1 \times \text{TRT} + \beta_2 \times X) \]
- Predict after turning on & off the TRT
  \[ \mu_1 = \exp(\beta_0 + \beta_1 \times 1 + \beta_2 \times X) \]
  \[ \mu_0 = \exp(\beta_0 + \beta_1 \times 0 + \beta_2 \times X) \]
- Summ \( \mu_1 \), \( \mu_0 \)
** THE METHOD OF RECYCLED PREDICTIONS - TO ESTIMATE EFFECT IN ADDITIVE-SCALE

- `gen trtold=aggr_trt`
- `replace aggr_trt =1`
  (2219 real changes made)
- `predict mu1, mu`

- `replace aggr_trt=0`
  (7129 real changes made)
- `predict mu0, mu`

- `replace aggr_trt = trtold`
  (4910 real changes made)

- `summ mu1 mu0`

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>mu1</td>
<td>7129</td>
<td>21447.39</td>
<td>6183.882</td>
<td>12332.27</td>
<td>59634.99</td>
</tr>
<tr>
<td>mu0</td>
<td>7129</td>
<td>17923.83</td>
<td>5167.942</td>
<td>10306.23</td>
<td>49837.65</td>
</tr>
</tbody>
</table>
Single equation models: Generalized Linear Models

- **Limitations**
  - Inefficient when empirical distribution of Y has fat tails (kurtotic)
  - Log Link is the most common link function used, but assumption that log(µ) = Xβ may be violated
  - Appropriate link or variance functions not known a-priori
  - Various tests exist, not always sufficient to overcome these problems
Proposed by Manning & Mullahy (2001)

- $\text{Var}(Y|X) = \theta_1 \mu^{\theta_2}$
- For Gamma, we assume $\theta_1 = 1$, $\theta_2 = 2$
- Test: $\ln(\text{Var}(Y|X)) = \ln(\theta_1) + \theta_2 \ln(\mu)$
- Use Residual $^2$ as an estimate for variance.
- \texttt{glm r2 lnmu, link(log) family(gamma) robust}
/* Modified Park test */
predict mu, mu
gen lnmu = ln(mu)
gen r2 = (totexp - mu)^2
glm r2 lnmu, link(log) family(gamma) robust

|        | Coef.     | Std. Err. |    z  |  P>|z|   |   [95% Conf. Interval] |
|--------|-----------|-----------|-------|--------|------------------------|
| lnmu   | 1.468981  | 0.2492923 | 5.89  | 0.000  | 0.9803774  1.957585 |
| _cons  | 5.419501  | 2.497738  | 2.17  | 0.030  | 0.5240253  10.31498 |

test lnmu=1
    chi2(  1) =  3.54
    Prob > chi2 =  0.0599

test lnmu=2
    chi2(  1) =  4.54
    Prob > chi2 =  0.0332

test lnmu=3
    chi2(  1) = 37.72
    Prob > chi2 =  0.0000
.drop mu lnmu r2
Generalized Gamma FIML

Three parameter distribution

- Mean is a function of all three parameters
- Only of them is related to $X\beta$, implicit log-link
- Special cases are Gamma, Exponential, Weibull, log-normal
- Except for log-normal case, no retransformation problem

- `stset y`
- `gengam2 x`
. /* Generalized Gamma Regression */

. stset totexp
. ****** Download -gengam2- command from http://home.uchicago.edu/~abasu/index_files/Pagec317.htm AND place in C:\ado\Personal\.

. gengam2 aggr_trt agec white black single married well mod stage1 stage2 lnc, robust

<table>
<thead>
<tr>
<th></th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.  Std. Err.  z  P&gt;</td>
</tr>
<tr>
<td>_t</td>
<td></td>
</tr>
<tr>
<td>aggr_trt</td>
<td>0.193495  0.0333561  5.80  0.000  0.1281183  0.2588717</td>
</tr>
<tr>
<td>agec</td>
<td>0.0104401  0.002439  4.28  0.000  0.0056598  0.0152205</td>
</tr>
<tr>
<td>white</td>
<td>0.0913009  0.0521438  1.75  0.080  -0.0108992  0.1935009</td>
</tr>
<tr>
<td>black</td>
<td>0.1798185  0.0662471  2.71  0.007  0.0499766  0.3096605</td>
</tr>
<tr>
<td>single</td>
<td>0.0355918  0.0546193  0.65  0.515  -0.07146  0.1426436</td>
</tr>
<tr>
<td>married</td>
<td>-0.0508413  0.0332982 -1.53  0.127  -0.1161046  0.0144221</td>
</tr>
<tr>
<td>well</td>
<td>-0.2535959  0.0625437 -4.05  0.000  -0.3761792  -0.1310125</td>
</tr>
<tr>
<td>mod</td>
<td>-0.1453943  0.0321754 -4.52  0.000  -0.2084568  -0.0823318</td>
</tr>
<tr>
<td>stage1</td>
<td>-0.0261988  0.0281348 -0.93  0.352  -0.0813419  -0.0289444</td>
</tr>
<tr>
<td>stage2</td>
<td>-0.0900821  0.0410532 -2.19  0.028  -0.1705449  -0.0096192</td>
</tr>
<tr>
<td>lnc</td>
<td>0.3746245  0.0203089  18.45  0.000  0.3348198  0.4144292</td>
</tr>
<tr>
<td>_cons</td>
<td>9.578895  0.0675292  141.85  0.000  9.44654  9.711249</td>
</tr>
</tbody>
</table>

Slide 48
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ln_sig</td>
<td>-.053114</td>
<td>.0147078</td>
<td>-3.61</td>
<td>0.000</td>
<td>-.0819408</td>
<td>-.0242872</td>
</tr>
<tr>
<td>kappa</td>
<td>.8766324</td>
<td>.0314132</td>
<td>27.91</td>
<td>0.000</td>
<td>.8150637</td>
<td>.9382012</td>
</tr>
<tr>
<td>sigma</td>
<td>.9482719</td>
<td>.013947</td>
<td></td>
<td></td>
<td>.9213265</td>
<td>.9760053</td>
</tr>
</tbody>
</table>

Tests for identifying distributions

<table>
<thead>
<tr>
<th>Distributions</th>
<th>chi2</th>
<th>df</th>
<th>Prob&gt;chi2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Gamma (kappa = sigma)</td>
<td>3.27</td>
<td>1</td>
<td>0.0704</td>
</tr>
<tr>
<td>Log Normal (kappa = 0)</td>
<td>778.77</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Weibull (kappa = 1)</td>
<td>15.42</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>Exponential (kappa = sigma = 1)</td>
<td>50.89</td>
<td>2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

New Variables for Schoenfield Residuals created: _h*, _c*

```
.dropr smr smear smear0 smear1 res2 *
```
Goodness-of-fit tests

- **Pearson Correlation**
  - Correlation between raw-scale predictions and residuals

- **Pregibon’s Link Test / Reset Test**
  - Run same model with $X\beta$ and $X\beta^2$ as covariates

- **Hosmer-Lemeshow Test**
  - Plot mean residuals across deciles of $X\beta$.
  - Perform joint test that all means are zero

- **Copa’s Test**
. /* GOODNESS OF FIT TESTS */
. ** For OLS
. reg totexp aggr_trt agec white black single married well
  mod stagel stage2 lnc, robust
. predict xb, xb
. gen mu_ols= xb
. gen res_ols = totexp - mu_ols
. ** Pearson Corr
. pwcorr res_ols mu_ols, sig
   | res_ols   mu_ols -------------+------------------
   | res_ols |   1.0000
   | mu_ols  |  -0.0000   1.0000               |   1.0000
. ** Link Test
. gen xb2=xb^2
. reg totexp xb xb2, robust

| totexp | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|----------------------|
| xb     | 0.2645214 | .3449794 | 0.77  | 0.443 | -.4117406 -.9407833 |
| xb2    | 0.0000171 | 8.43e-06 | 2.03  | 0.043 | 5.46e-07 .0000336   |
| _cons  | 7353.987  | 3357.482 | 2.19  | 0.029 | 772.3251 13935.65   |
** H-L test

```
xtile xbtile =xb, nq(10)
qui tab xbtile, gen(xbt)
reg res_ols xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10, nocons robust
```

|     | Coef.   | Std. Err. |   t    |   P>|t|   | [95% Conf. Interval] |
|-----|---------|-----------|--------|--------|----------------------|
| xbt1| 857.5334| 614.4044  | 1.40   | 0.163  | -346.8819 2061.949   |
| xbt2| 681.1362| 674.0399  | 1.01   | 0.312  | -640.1824 2002.455   |
| xbt3| -525.1787| 642.1524 | -0.82  | 0.413  | -1783.988 733.631    |
| xbt4| -142.7663| 699.025  | -0.20  | 0.838  | -1513.063 1227.531   |
| xbt5| 938.962 | 933.5517  | 1.01   | 0.315  | -891.0768 2769.001   |
| xbt6| -638.9313| 716.1985 | -0.89  | 0.372  | -2042.893 765.0306   |
| xbt7| -1635.909| 696.5204 | -2.35  | 0.019  | -3001.296 -270.5221  |
| xbt8| -2250.963| 710.2455 | -3.17  | 0.002  | -3643.255 -858.6706  |
| xbt9| 707.1719 | 1138.647 | 0.62   | 0.535  | -1524.915 2939.259   |
| xbt10| 2007.334| 1231.866 | 1.63   | 0.103  | -407.489 4422.158    |

```
.test xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10
```

```
F( 10, 7119) = 2.41
Prob > F = 0.0074
```

```
.drop xb xb2 xbt* xbtile
```

Slide 52
** For lnOLS

```
. reg lny aggr_trt agec white black single married well mod stage1 stage2 lnc, robust
. predict xb, xb
. gen smr = exp(lny- xb)
. qui summ smr
. gen smear = r(mean)
. gen mu_lols= exp(xb)*smear
. gen res_lols = totexp - mu_lols

** Pearson Corr

. pwcorr res_lols mu_lols, sig
```

```
<table>
<thead>
<tr>
<th></th>
<th>res_lols</th>
<th>mu_lols</th>
</tr>
</thead>
<tbody>
<tr>
<td>res_lols</td>
<td>1.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>mu_lols</td>
<td>-0.1776</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

** Link Test

. gen xb2=xb^2

. reg lny xb xb2, robust
```

```
Robust

|      | Coef.     | Std. Err. | t     | P>|t|   | [95% Conf. Interval]
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>lny</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>7.586948</td>
<td>1.443073</td>
<td>5.26</td>
<td>0.000</td>
<td>4.758097</td>
</tr>
<tr>
<td>xb2</td>
<td>-.3492898</td>
<td>.0757058</td>
<td>-4.61</td>
<td>0.000</td>
<td>-.4976956</td>
</tr>
</tbody>
</table>
```
** H-L test**

```
xtile xbtile =xb, nq(10)
qui tab xbtile, gen(xbt)
reg res_lols xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10, nocons robust
```

|        | Robust       | Coef.   | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|--------------|---------|-----------|-------|-------|---------------------|
| xbt1   | 2479.074     | 635.594 |           | 3.90  | 0.000 | 1233.121 3725.027   |
| xbt2   | 2585.675     | 755.5887|           | 3.42  | 0.001 | 1104.497 4066.853   |
| xbt3   | 608.5663     | 694.4084|           | 0.88  | 0.381 | -752.6807 1969.813  |
| xbt4   | 573.4585     | 733.1452|           | 0.78  | 0.434 | -863.724 2010.641   |
| xbt5   | 1438.745     | 775.6945|           | 1.85  | 0.064 | -81.84715 2959.336  |
| xbt6   | -215.099     | 826.1292|           | -0.26 | 0.795 | -1404.36 1404.36    |
| xbt7   | -2386.189    | 757.4182|           | -3.15 | 0.002 | -3870.954 -901.4244 |
| xbt8   | -3506.212    | 839.8621|           | -4.17 | 0.000 | -5152.592 -1859.833 |
| xbt9   | -3966.24     | 1005.915|           | -3.94 | 0.000 | -5938.133 -1994.347 |
| xbt10  | -10547.35    | 1207.501|           | -8.73 | 0.000 | -12914.42 -8180.294 |

```
test xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10
F( 10,  7119) =   15.10
Prob > F =    0.0000
```

```
drop xb xb2 xbt* xbtile
```

Slide 54
. ** For log-GLM
. glm totexp aggr_trt agec white black single married well mod stage1 stage2 lnc, link(log) family(gamma) robust
. predict xb, xb
. gen mu_glm = exp(xb)
. gen res_glm = totexp - mu_glm
. ** Pearson Corr
. pwcorr res_glm mu_glm, sig

<table>
<thead>
<tr>
<th></th>
<th>res_glm</th>
<th>mu_glm</th>
</tr>
</thead>
<tbody>
<tr>
<td>res_glm</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>mu_glm</td>
<td>-0.0062</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0.6022</td>
<td></td>
</tr>
</tbody>
</table>

. ** Link Test
. gen xb2=xb^2
. glm totexp xb xb2, link(log) family(gamma) robust

|        | Coef.     | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|--------|-----------|-----------|-------|-------|----------------------|
| totexp |           |           |       |       |                      |
| xb     | 2.601821  | 3.09824   | 0.84  | 0.401 | -3.470618            | 8.67426  |
| xb2    | -0.0806185| .155559   | -0.52 | 0.604 | -.3855025            | .2242655 |
| _cons  | -7.950714 | 15.41964  | -0.52 | 0.606 | -38.17266            | 22.27123 |
** H-L test

xtile xtile =xb, nq(10)
qui tab xtile, gen(xbt)
reg res_glm xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10, nocons robust

|      | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|------|-------|-----------|-------|-----|---------------------|
| xbt1 | -781.7378 | 554.103 | -1.41 | 0.158 | -1867.944  304.4688 |
| xbt2 |  473.2616 | 736.601 | 0.64  | 0.521 | -970.6954  1917.219 |
| xbt3 | -323.851  | 626.755 | -0.52 | 0.605 | -1552.477  904.751 |
| xbt4 |  49.60152 | 702.7822| 0.07  | 0.944 | -1328.061  1427.264 |
| xbt5 | 1288.524  | 797.3756| 1.62  | 0.106 | -274.5689  2851.618 |
| xbt6 |  207.3119 | 861.6822| 0.24  | 0.810 | -1481.841  1896.465 |
| xbt7 | -558.6503 | 690.9381| -0.81 | 0.419 | -1913.094  795.7937 |
| xbt8 |-1955.442  | 718.1396| -2.72 | 0.006 | -3363.209  -547.674 |
| xbt9 | 1364.908  | 1148.222| 1.19  | 0.235 | -885.9489  3615.764 |
| xbt10|  156.9733 | 1227.114| 0.13  | 0.898 | -2248.535  2562.482 |

.test xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10
F(10, 7119) = 1.48
Prob > F = 0.1382

drop xb xb2 xbt* xtile
Single equation models: Generalized Linear Models - EEE

Flexible Mean Model

\[
\frac{(\mu^\lambda - 1)}{\lambda} = X\beta \quad \text{if } \lambda \neq 0
\]

\[
\log(\mu) = X\beta \quad \text{if } \lambda = 0
\]

Flexible Variance Model

\[
V(Y) = \phi \mu^\theta
\]

\[
\begin{align*}
\theta = 1 & \Rightarrow \text{Poisson} \\
\theta = 2 & \Rightarrow \text{Gamma} \\
\theta = 3 & \Rightarrow \text{Inv. Gauss}
\end{align*}
\]
Goal: estimate $\gamma = (\beta^T, \lambda, \phi, \theta)^T$

- EEE estimator estimates:
  Mean model parameters $(\beta, \lambda)$ and
  Var. model parameters $(\phi, \theta)$
  simultaneously using estimating equations.

- Extended Estimating Equations (EEE) ....
Single equation models: Generalized Linear Models - EEE

\[ G_{\beta_j}^i = (y_i - \mu_i)V_i^{-1} \left( \frac{\partial \mu_i}{\partial \beta_j} \right) \]

\[ G_{\lambda}^i = (y_i - \mu_i)V_i^{-1} \left( \frac{\partial \mu_i}{\partial \lambda} \right) \]

\[ G_{\phi}^i = \left[ (y_i - \mu_i)^2 - V_i \right] V_i^{-2} \left( \frac{\partial V_i}{\partial \phi} \right) \]

\[ G_{\theta}^i = \left[ (y_i - \mu_i)^2 - V_i \right] V_i^{-2} \left( \frac{\partial V_i}{\partial \theta} \right) \]
Single equation models: Generalized Linear Models - EEE

Latest Stata code available from
http://home.uchicago.edu/~abasu

```
. summarize totcost, meanonly
. global scale=r(mean)
. generate y=totcost/$scale
. pglm y trt x
. pglm y trt x [pw=marsupwt]
. pglmprevent y_hat, mu scale($sc)
. pglmprevent ietrt, ie(trt) scale($sc)
. pglmprevent me_x, me(x) scale($sc)
```
. /* EEE - EXTENDED ESTIMATING EQUATIONS */
. ****** Download -pglm- command from http://home.uchicago.edu/~abasu/index_files/Pagec317.htm AND place in C:\ado\Personal\n
. qui sum totexp, meanonly
. gen y = totexp/r(mean)
. global scale = r(mean)
. summ y

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>7129</td>
<td>1.000</td>
<td>1.131858</td>
<td>0.001372</td>
<td>19.4377</td>
</tr>
</tbody>
</table>

. pglm y aggr_trt agec white black single married well
mod stage1 stage2 lnc

Extended GEE with Power Variance Function
Optimization: Fisher's Scoring
Variance: (theta1*mu^theta2)
Link: (mu^lambda - 1)/lambda
Std Errors: Robust

Slide 61
| y       | Coef.   | Std. Err.   | z     | P>|z|   | [95% Conf. Interval] |
|---------|---------|-------------|-------|-------|---------------------|
| aggr_trt | 0.176795 | 0.0319912   | 5.53  | 0.000 | 0.1140935 - 0.2394965       |
| aec     | 0.010301 | 0.0025097   | 4.10  | 0.000 | 0.005382 - 0.01522        |
| white   | 0.0772856 | 0.0524606   | 1.47  | 0.141 | -0.0255354 - 0.1801065    |
| black   | 0.1771791 | 0.0671802   | 2.64  | 0.008 | 0.0455084 - 0.3088498     |
| single  | 0.0369893 | 0.0563494   | 0.66  | 0.512 | -0.0734535 - 0.1474322    |
| married | -0.053268 | 0.0341707   | -1.56 | 0.119 | -0.1202413 - 0.0137054    |
| well    | -0.2421367 | 0.063453   | -3.82 | 0.000 | -0.3665023 - 0.117711     |
| mod     | -0.144254 | 0.0339966   | -4.24 | 0.000 | -0.2108861 - 0.077622     |
| stage1  | -0.0237874 | 0.0289403   | -0.82 | 0.411 | -0.0805093 - 0.0329345    |
| stage2  | -0.0913624 | 0.0418396   | -2.18 | 0.029 | -0.1733666 - 0.0093583    |
| lnc     | 0.3719294 | 0.0221651   | 16.78 | 0.000 | 0.3284867 - 0.4153721     |
| _cons   | -0.2769968 | 0.0682525   | -4.06 | 0.000 | -0.4107693 - 0.1432244    |
| lambda  |         |             |       |       |                     |
| _cons   | 0.1928841 | 0.303307    | 0.64  | 0.525 | -0.4015866 - 0.7873549    |
| theta1  |         |             |       |       |                     |
| _cons   | 1.166449 | 0.0810133   | 14.40 | 0.000 | 1.007666 - 1.325233       |
| theta2  |         |             |       |       |                     |
| _cons   | 1.457585 | 0.2222653   | 6.56  | 0.000 | 1.021953 - 1.893217       |

. predict xb, xb
. pglmpredict mu_pglm, mu scale($scale)
. `summ` `mu_pglm` `totexp`

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>mu_pglm</code></td>
<td>7129</td>
<td>20265.42</td>
<td>5696.584</td>
<td>10770.24</td>
<td>51476.94</td>
</tr>
<tr>
<td><code>totexp</code></td>
<td>7129</td>
<td>20265.17</td>
<td>22937.29</td>
<td>2.78</td>
<td>393908.4</td>
</tr>
</tbody>
</table>

. `pglmpredict` `trteffect`, `ie(aggr_trt)` `scale($scale)`

. `summ` `trteffect`

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>trteffect</code></td>
<td>7129</td>
<td>3470.224</td>
<td>790.525</td>
<td>2190.088</td>
<td>7521.893</td>
</tr>
</tbody>
</table>

. `gen` `res_pglm = totexp - mu_pglm`

**GOODNESS OF FIT**

. **Pearson Corr**

. `pwcorr` `res_pglm` `mu_pglm`, `sig`

<table>
<thead>
<tr>
<th></th>
<th>res_pglm</th>
<th>mu_pglm</th>
</tr>
</thead>
<tbody>
<tr>
<td>res_pglm</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>mu_pglm</td>
<td>-0.0013</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>0.9158</td>
<td></td>
</tr>
</tbody>
</table>
** H-L test

xtile xtile =xb, nq(10)
qui tab xtile, gen(xbt)
reg res_pglm xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10, nocons robust

|               | Coef.   | Std. Err. | t    | P>|t|   | [95% Conf. Interval] |
|---------------|---------|-----------|------|-------|----------------------|
| res_pglm      |         |           |      |       |                      |
| xbt1          | -609.5808 | 553.8617  | -1.10 | 0.271 | -1695.314    476.1529 |
| xbt2          | 646.0331  | 742.0791  | 0.87  | 0.384 | -808.6624    2100.729 |
| xbt3          | -471.0719 | 618.8141  | -0.76 | 0.447 | -1684.131    741.9877 |
| xbt4          | -84.50918 | 702.9003  | -0.12 | 0.904 | -1462.403    1293.384 |
| xbt5          | 1116.715  | 791.1745  | 1.41  | 0.158 | -434.2219    2667.652 |
| xbt6          | 365.4751  | 872.6898  | 0.42  | 0.675 | -1345.256    2076.207 |
| xbt7          | -1006.109 | 683.5575  | -1.47 | 0.141 | -2346.085    333.8666 |
| xbt8          | -1908.943 | 719.4691  | -2.65 | 0.008 | -3319.317    -498.5697 |
| xbt9          | 1258.806  | 1145.722  | 1.10  | 0.272 | -987.151     3504.762 |
| xbt10         | 691.741   | 1227.973  | 0.56  | 0.573 | -1715.451    3098.933 |

.test xbt1 xbt2 xbt3 xbt4 xbt5 xbt6 xbt7 xbt8 xbt9 xbt10
F( 10,  7119) =  1.55
Prob > F =  0.1164

.drop xb xbt* xtile
Empirical example

- Cost attributable to Heart Failure (HF) post myocardial infarction (MI)
- Use Medstat data, extract all claims between July 1, 1997 and December 31, 2000

Prior period
6 months

Post period
12 months

HF = I(Patient who develop HF in post-period)

Time line

Index MI Date
<table>
<thead>
<tr>
<th>Exclusion Criteria</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Starting cohort</strong></td>
<td></td>
</tr>
<tr>
<td>No claims in pre-MI or enrollment information</td>
<td>1279</td>
</tr>
<tr>
<td>Patients with pre-MI HF only</td>
<td>1696</td>
</tr>
<tr>
<td>Patients with pre-MI MI only</td>
<td>806</td>
</tr>
<tr>
<td>Patients with pre-MI HF and MI</td>
<td>269</td>
</tr>
<tr>
<td>Patients with post-MI MI</td>
<td>575</td>
</tr>
<tr>
<td>Post MI HF patients with intermediate coronary syndrome and coronary occlusion</td>
<td>18</td>
</tr>
<tr>
<td>without MI</td>
<td></td>
</tr>
<tr>
<td>Patients with capitation type insurance plans</td>
<td>2258</td>
</tr>
<tr>
<td>Patients with VSD</td>
<td>16</td>
</tr>
<tr>
<td>Patients with mitral regurgitation</td>
<td>46</td>
</tr>
<tr>
<td>Patients with cost of index hospitalization &lt; $500</td>
<td>9</td>
</tr>
<tr>
<td>Patients with LOS&gt;100 and cost &lt; $10,000</td>
<td>1</td>
</tr>
<tr>
<td>Patients with age&lt;18 years or missing age</td>
<td>3</td>
</tr>
<tr>
<td>Patients with missing plan information</td>
<td>20</td>
</tr>
<tr>
<td>Patient with index date not in (1Dec1998 – 1Dec 2000)</td>
<td>4633</td>
</tr>
<tr>
<td><strong>Eligible cohort</strong></td>
<td></td>
</tr>
<tr>
<td>Patients without full 1-year follow-up data.</td>
<td>3427</td>
</tr>
<tr>
<td>Observations with missing covariate values</td>
<td>3</td>
</tr>
<tr>
<td><strong>Final sample for this study</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>7428</strong></td>
</tr>
</tbody>
</table>
Empirical Example: Cost of Post-MI HF

**With HF:** N=2259
Mean (sd) = $45,070 (57,366);
Sk. = 6.0; Kurt. = 73

**Without HF:** N=5169
Mean (sd) = $31,708 (30,110);
Sk. = 4.4; Kurt. = 50

**With HF:** N = 2259
Sk. = 0.13; Kurt. = 3.2

**Without HF:** N = 5169
Sk. = -0.03; Kurt. = 3.4
Empirical Example: Cost of Post-MI HF

- **Final Sample:** 7428 patients
  - 2259 (30.4%) with HF
- **Dependent variable:** Total expenditure
- **Independent variables:** HF, age, age-sq, insurance type, Medicare status, procedures performed, year of MI, type of MI and indicators for 30 categories of co-morbidities.
- **Compare:** OLS, log-OLS (with HF-specific smearing), GLM Gamma-Log link.
**Empirical Example:**
**Cost of Post-MI HF**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( \hat{\pi} ) Mean (std.err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>15578 (1172)</td>
</tr>
<tr>
<td>Log-OLS (Het.)</td>
<td>17491 (1159)</td>
</tr>
<tr>
<td>Gamma (log link)</td>
<td>15890 (1104)</td>
</tr>
</tbody>
</table>
Empirical Example: Cost of Post-MI HF

Deciles of Linear Predictor

1 2 3 4 5 6 7 8 9 10

Mean Residuals ($) 15000 - 12000 - 9000 - 6000 - 3000 - 0 - 3000 - 6000 - 9000 - 12000 - 15000 -

Basu et al., HE 2006
Empirical Example: Cost of Post-MI HF

- **Results from the EEE Model**
  - Mean Model (Link parameter)
    - $\hat{\lambda} = 0.201$ (95% CI: 0.04, 0.37)
    - Appropriate link $\sim \mu^{0.2}$ or $\mu^{0.25}$
  - Variance Model: $V(y) = \phi \mu^\theta$
    - $\hat{\phi} = 0.80$ (95% CI: 0.72, 0.89); $\hat{\theta} = 1.79$ (95% CI: 1.57, 2.01)
    - Appropriate variance $\sim \mu^2$

- Suggest fifth or fourth-root link with Gamma variance.
### Empirical Example: Cost of Post-MI HF

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$\hat{\pi}$ Mean (std.err.)</th>
<th>$(\hat{\pi}<em>{Row} - \hat{\pi}</em>{Col})$ Mean (std.err.)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>15578 (1172)</td>
<td>-1913 (627)</td>
</tr>
<tr>
<td>Log-OLS (Het.)</td>
<td>17491 (1159)</td>
<td>-312 (566)</td>
</tr>
<tr>
<td>Gamma (log link)</td>
<td>15890 (1104)</td>
<td>1601 (239)</td>
</tr>
<tr>
<td>EEE</td>
<td>14722 (1135)</td>
<td>2769 (560)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1168 (510)</td>
</tr>
</tbody>
</table>

* Log-OLS (Hom.) - 14619 (901) -2872 -1271 -103 (599) (574) (716)
Other models

- Multi-part models
  - Two-part: 1\textsuperscript{st} part model any costs (logit/probit) 2\textsuperscript{nd} part models costs given any
  - Multi-part: 2-part models for each type of expenditures
  - CDE: 1\textsuperscript{st} part for multinomial categories of costs 2\textsuperscript{nd} parts for levels within each category

- Mixture models
Comparison of Alternative Estimators

- Comparisons through simulations or empirical applications:
  - Duan et al., JBES 1983; Mullahy, JHE 1998; Lipscomb et al., SMDM 1999; Manning & Mullahy, JHE 2001; Deb & Trivedi, JHE 2002; Deb & Burgess, Hunter College Mimeo 2003; Basu et al., HE 2004; Buntin & Zaslavsky, JHE, 2004; Basu & Rathouz, Biostatistics 2005; Briggs et al., HE 2005; Manning et al., JHE 2006; Basu et al., HE 2006; Hill & Miller, HE 2009;

- Consistently found that no one estimator is appropriate for all expenditure data

- Theory has difficult time predicting signs of covariate effects – provides almost no guidance on functional form
Other issues with cost models

- Longitudinal costs – generalized estimating equations versus random effects model
- Right censoring
- Survival effects versus intensity effects
- Causal estimation – instrumental variables
- Predictions