Preconference Course:
Modeling Health Care Costs and Counts

By

Partha Deb

Willard Manning

Edward Norton

iHEA - Toronto Conference, 2011

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

What is the cost of interest?

1. Costs in fixed period of time (e.g., stroke costs paid in 2009)?
2. Per episode or per lifetime costs of stroke in incident cases?

Our focus is on the former

Second question requires survival methods and consideration of censoring in data

Characteristics of Health Care

Costs and Utilization

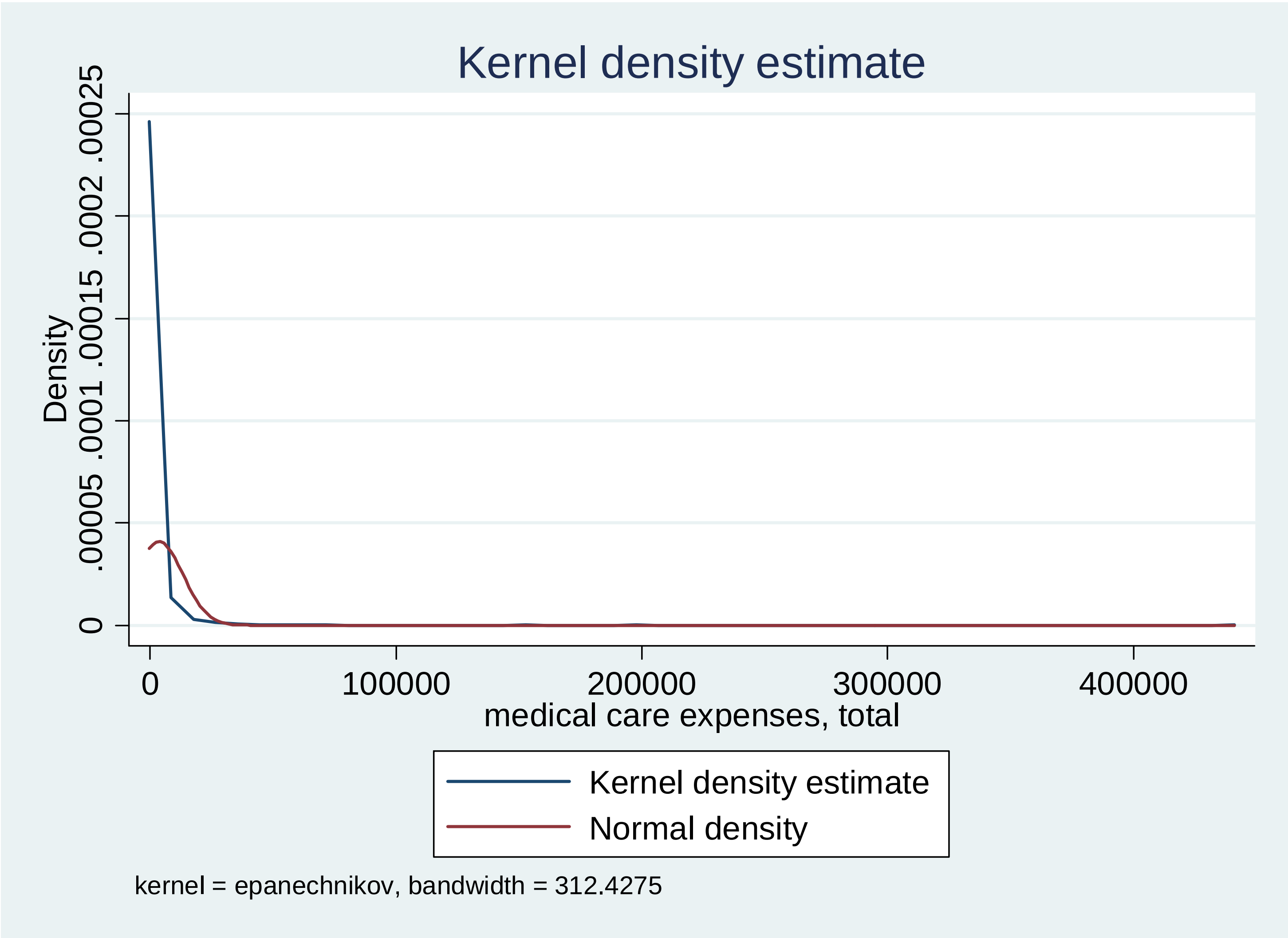
Large fraction of population without any care during period of observation

Consumption among those with any care is very skewed (visits, hospitalizations, costs)

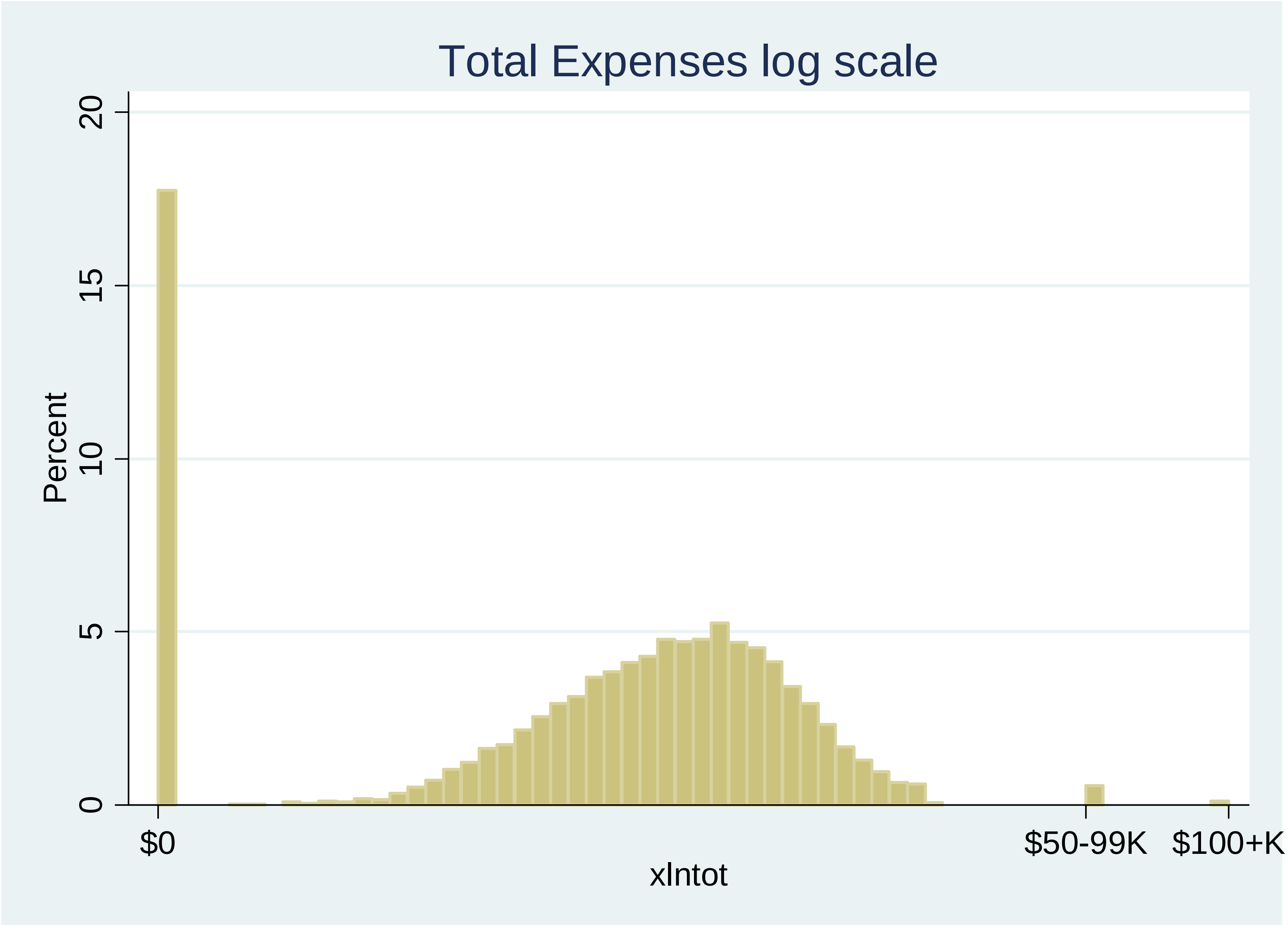
Nonlinearity in response to covariates

Cost response may change by level of consumption (e.g. outpatient versus inpatient, or low to high levels).

Density of Total Medical Expenditures, Adults, MEPS 2004

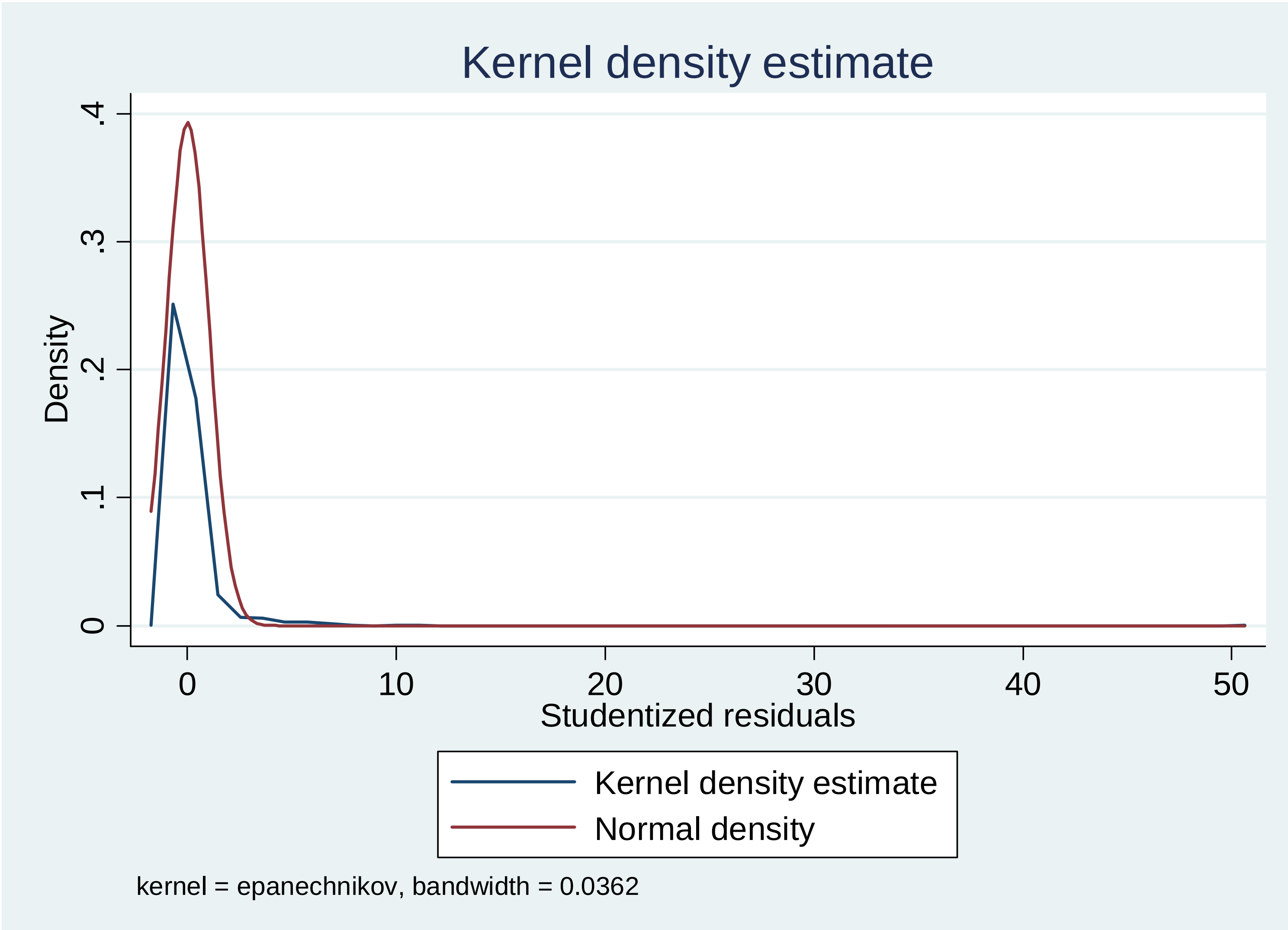


Discrete Version of Density for Total Medical Expenditures



“Bins” are $1000 wide, except upper two. Log-scaled to pull in right tail.

Density of Studentized Residuals for Total Medical Expenditures, Adults, MEPS 2004



Potential Problems from Ignoring Characteristics

Usual econometric (least squares) methods will yield less precise estimates of means and marginal effects.

If failing to deal with inherently nonlinear response may lead to biased estimates for substantial subpopulations.

Results not robust to tail problems unless very large samples.

Estimates from one subsample may forecast poorly to another subsample from same population

Need more robust methods that

Recognize distribution of data

Are less sensitive to right tail

Provide estimates of E(y|x)

Overview

Studies with skewed outcomes but no zero mass problem

Alternative models

Comparing alternative models

Assessing model fit

Interpretation

Overview

Studies with skewed outcomes but no zero mass problem

- **Alternative models**
  - **OLS on untransformed use or expenditures**
  - **OLS for log(y)**
  - **Box-Cox generalization**
  - **Generalized Linear Models (GLM)**
  - **Generalized Gamma Model**

Studies with No Zero Mass (cont'd)

Concerns

Robustness to skewness

Reduce influence of extreme cases

Good forecast performance

No systematic misfit over range of predictions or range of major covariates (e.g., price, income).

Efficiency of estimator

OLS of y on x's

Advantages

Easy

No retransformation problem

Marginal /incremental effects easy to calculate

Disadvantages

Not robust in small to medium sized data sets
or where some subgroups are small

Can produce out-of-range predictions: $\hat{y}_i = x_i' \hat{\beta} < 0$

Inefficient (ignores heteroskedasticity)

Poor forecast performance

Why are we concerned with robustness to skewness in OLS?

OLS overemphasizes extreme cases when data are very skewed right or cases have leverage. For OLS,

$$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon$$

but some ε's are extremely large, as well as x's extreme or rare

Raises the risk of influential outlier(s) that pull estimate $\hat{\beta}$ away from β.

See figures on pages 6 and 7 for potential problems.

Log(y) or Box-Cox Models

Advantages

- **Widely known, especially log(y) version.**
  - **Reduces robustness problem by focusing on symmetry.**
- **Improved precision if y skewed right.**
- **May reduce (but not eliminate) heteroscedasticity.**

Disadvantages

- **Retransformation problem could lead to bias.**
- **Some Box-Cox version's coefficients are not directly interpretable.**
- **May not achieve linearity on estimation scale.**

OLS for log(y)

OLS or MLE for $\log(y) = X\beta + \varepsilon$

where $E(\varepsilon) = 0$, $E(X'\varepsilon) = 0$.

Estimates for $E(\log(y)|x)$, not $\log(E(y)|x)$).

Usually want arithmetic mean, not geometric mean.

May be difficult to obtain unbiased estimates of mean response

$E(y|x)$ if error ε heteroscedastic in x's or other z's.

Dilemma with OLS for log(y)

Logged estimates are often far more precise and robust than direct analysis of unlogged dependent variable.

But, no one interested in log scale results *per se*.

Effect of Heteroscedasticity

Untransformed dependent variable (e.g., cost)

Need GLS for efficient estimates and to correct inference statistics (*Or* use Huber/White/Eicker with OLS to get consistent inference statistics)

Transformed dependent variable (e.g., log(cost))

Need GLS for efficient estimates & correct inference statistics (Or use Huber/White/Eicker with OLS to get consistent inference statistics)

And correction for form of hetero. to yield consistent predictions on the raw (untransformed) scale

OLS on log(y) for comparison of two treatment groups with normal errors

Assume $\log(y)_G \sim N(\mu_G, \sigma^2_G)$ where treatment G = A or B

$$E(y \mid G = A) = e^{(\mu_A + 0.5\sigma_A^2)}$$

Under heteroscedasticity by group

$$\frac{E(y_A)}{E(y_B)} = e^{((\mu_A - \mu_B) + 0.5(\sigma_A^2 - \sigma_B^2))}$$

Under homoscedasticity (σ^2 = a constant)

$$\frac{E(y_A)}{E(y_B)} = e^{(\mu_A - \mu_B)}$$

Note: Same issue applies if error not normally distributed

Retransformation with Covariate Adjustment

Suppose $y > 0$ and we run OLS regression for $\ln(y) = x\beta + \varepsilon$.

With $E[\varepsilon|x] = 0$, β & $E[\ln(y)|x]$ consistently estimated by linear regression.

Policy questions not typically focused on β *per se*, but on how $E[y]$ varies with x.

Expectations if E(ε) = 0 and E(X′ε) = 0:

$$E(y_i) = e^{x_i'\beta}\, E(e^{\varepsilon_i} \mid x_i)$$

$$E(y_i) \neq cons \bullet e^{x_i'\beta}$$ **if ε is heteroscedastic in x**

Marginal effects of a covariate x (e.g., income) on expected outcome on the raw scale:

$$\frac{\partial E(y_i)}{\partial x_k} = e^{x_i'\beta}\left(\beta_k E(e^{\varepsilon_i} \mid x_i) + \frac{\partial E(e^{\varepsilon_i} \mid x_i)}{\partial x_k}\right)$$

$$\frac{\partial E(y_i)}{\partial x_k} \neq \{E(y_i \mid x_i)\}\,\beta_k$$ **if heteroscedastic in x**

$$\frac{\partial E(y_i)}{\partial x_k} \neq \{e^{x_i'\beta}\}\beta_k$$ **as is often assumed**

More in Edward's section later this morning

Examples

Health Insurance Experiment (HIE) error variance on log scale for users increasing in cost sharing for outpatient and total medical expenses. Use of homoscedastic model overstates effect of cost sharing. (Manning, JHE, 1998).

Visits from National Health Interview Survey. Response heteroscedastic in gender and education. (Mullahy, JHE, 1998).

MEPS 2004 response heteroscedastic in income and education (see below).

Box-Cox Models

Log transform not only solution to skewness

Assume transform of y such that:

$$[(y_i^{\lambda}-1)/\lambda] = x_i'\beta + \varepsilon_i \quad \text{if } \lambda \neq 0$$

$$\log(y_i) = x_i'\beta + \varepsilon_i \quad \text{if } \lambda = 0$$

where ε_i is distributed iid as $N(0, \sigma^2)$.

Estimate by MLE

Tends to minimize skewness in residuals

Log is not always "best" transform; depends on degree of skewness.

Example: Square Root Model by OLS

Assume that $\sqrt{y}$ is linear and additive

$$\sqrt{y_i} = x_i' \beta + \varepsilon_i$$

with E(ε) = 0 and E(x'ε) = 0. Then,

$$E(\hat{\beta}_{OLS}) = \beta .$$

Thus, OLS or least squares unbiased on square root scale.

Heteroscedasticity only raises efficiency and inference problems on square root scale.

Square Root Model by OLS (cont'd)

Back to the raw scale:

$$y_i = (x_i'\beta)^2 + 2\,(x_i'\beta)\,\varepsilon_i + \varepsilon_i^2$$

Thus

$$E(y_i \mid x_i) = (x_i'\beta)^2 + \sigma_\varepsilon^2(x)$$

What is the marginal effect of x?

$$\frac{\partial E(y_i)}{\partial x_k} = 2(x_i'\beta)\beta_k + \frac{\partial \sigma_\varepsilon^2(x_i)}{\partial x_k} \neq 2\,(x_i'\beta)\,\beta_k$$

Heteroscedasticity on square root scale raises bias issues on raw scale if not properly retransformed.

More on Retransformation Issues

Normal assumption is not innocuous!

Although estimates of β's may be insensitive, the expectation of untransformed value can be quite sensitive to departures from normality, esp. in right tail.

Solutions

Use Duan's (JASA, 1983) smearing estimator by subgroup, which is non-parametric. Difficult if heteroscedastic in a continuous covariates or in multiple covariates.

Use an appropriate Generalized Linear Model (GLM).

Retransformation Issues (cont'd)

Retransforming model results for log(y) by least squares

$$\log(y) = x\beta + \varepsilon$$

Homoscedastic case

$$E(y \mid x) = e^{(x\beta + 0.5\sigma^2)} \quad \text{if } \varepsilon \text{ is normally distributed}$$

$$E(y \mid x) = \left(e^{x\beta}\right) s, \quad \text{if not normally distributed}$$

$$\hat{s} = \frac{1}{N} \sum e^{(\log(y) - x\hat{\beta})} \quad \text{smearing (Duan, 1983)}$$

Heteroscedastic by group

Different variances by group if ε normally distributed

Different smearing by group if ε not normal

Duan's Smearing Estimator

Example

```
regress lny $x
      predict double resid, residual
egen Dsmear = mean(exp(resid))
display Dsmear
```

Consistent estimate of E(exp(ε)); Duan (JASA, 1983)

The smearing factor is typically between 1 and 4

Separate smearing by group if heteroscedastic by group

Retransformation Issues (cont'd)

Generally error ε is not normally distributed, heteroscedsticity may be complex, or may be heteroscedastic in several variables.

Normal theory retransformation methods can be biased.

Heteroscedastic smearing by group is:

- **Inefficient**
- **Difficult if covariate continuous**

Alternative: model E(y|x) directly using GLM.

Generalized Linear Models (GLM)

Goal

estimate mean of y, conditional on covariates x's.

Specify

a distribution that reflects mean - variance relationship

a link function between linear index xβ and mean μ = E(y|x)

Example

Gamma regression with log link

V(y|x) proportional to $[E(y \mid x)]^2$

$$Log(E(y \mid x_i)) = x_i \beta \quad \Rightarrow \quad E(y_i \mid x_i) = e^{x_i' \beta}$$

′

GLM (cont'd)

Use data to find distributional family and link

Family "down weights" noisy high mean cases

Link can handle linearity

Note difference in roles from Box-Cox

- **Box-Cox power transform to mostly gain symmetry in error (residual)**
- **GLM with power link function addresses linearity of response on scale to be chosen**

GLM (cont'd)

GLM/GEE/GMM modeling approach's estimating equations

$$\sum_{i=1}^{N} \frac{\partial \mu(x_i'\beta)}{\partial \beta} \times V(x_i)^{-1} \times (y_i - \mu(x_i'\beta)) = 0$$

Given correct specification of E[y|x] = $\mu(x'\beta)$, the key issues relate to second-order or efficiency effects.

This requires consideration of the structure of V(y|x).

GLM Variance Structure

Accommodates skewness & related issues via variance-weighting rather than transform/retransform methods.

Assumes $\mathrm{Var}[y|x] = \alpha \times [E(y|x)]^{\delta}$

$$= \alpha \times [\exp(x\beta)]^{\delta}$$

This implies moment restriction

$$E[\{y - \exp(x'\beta)\}^{2} - \{\alpha \times [\exp(x'\beta)]^{\delta}\} \mid x] = 0$$

GLM Variance Structure (cont'd)

For GLM, can

1. Adopt alternative "standard" parametric distributional assumptions,

$\delta = 0$	**(e.g. Gaussian NLLS)**
$\delta = 1$	**(e.g. Poisson)**
$\delta = 2$	**(e.g. Gamma)**
$\delta = 3$	**(e.g. Wald or inverse Gaussian)**

Estimation and inference available in Stata's glm or xtgee procedures.

If δ not near integer, consider extended GLM (see below) or use closest parametric case and take an efficiency loss.

GLM Variance Structure (cont'd)

2. Estimate δ via:

gamma regression of (y-$\hat{y}$)^2 on [1, $x'\hat{\beta}$] (modified "Park test" estimated by GLM).

linear regression of log((y-$\hat{y}$)^2) on [1,$x'\hat{\beta}$] (modified "Park test" estimated by least squares).

nonlinear regression of (y-$\hat{y}$)^2 on $\alpha(\exp(x'\hat{\beta}))^{\delta}$

Given choice of δ, can form V(x) and conduct (more efficient) second-round estimation and inference.

Overview

Studies with skewed outcomes but no zero mass problem

Alternative models

Comparing alternative models

Assessing model fit

Interpretation

Performance of Alternative Estimators

Examine alternative estimators of log(E(y|x)) for consistency and precision

Determine sensitivity to common data problems in health economics applications

Skewness
Heavy tailed, even with log transform
Heteroscedasticity
Different shapes to pdf

Results: no dominant estimator

See Manning and Mullahy (JHE, 2001) for details

Monte Carlo Simulation

Data generation

Skewness in dependent measure

Log normal with variance 0.5, 1.0, 1.5, 2.0

Heavier tailed than normal on the log scale

Mixture of log normals

Heteroscedastic responses

Std. dev. proportional to x

Variance proportional to x

Alternative pdf shapes

monotonically declining or bell-shaped

Gamma with shapes 0.5, 1.0, 4.0

Estimators Considered

Log-OLS with

- **homoscedastic retransformation**
- **heteroscedastic retransformation**

Generalized Linear Models (GLM), log link

- **Nonlinear Least Squares (NLS)**
- **Poisson**
- **Gamma**

Figure 1
Effect of Skewness on the Raw Scale

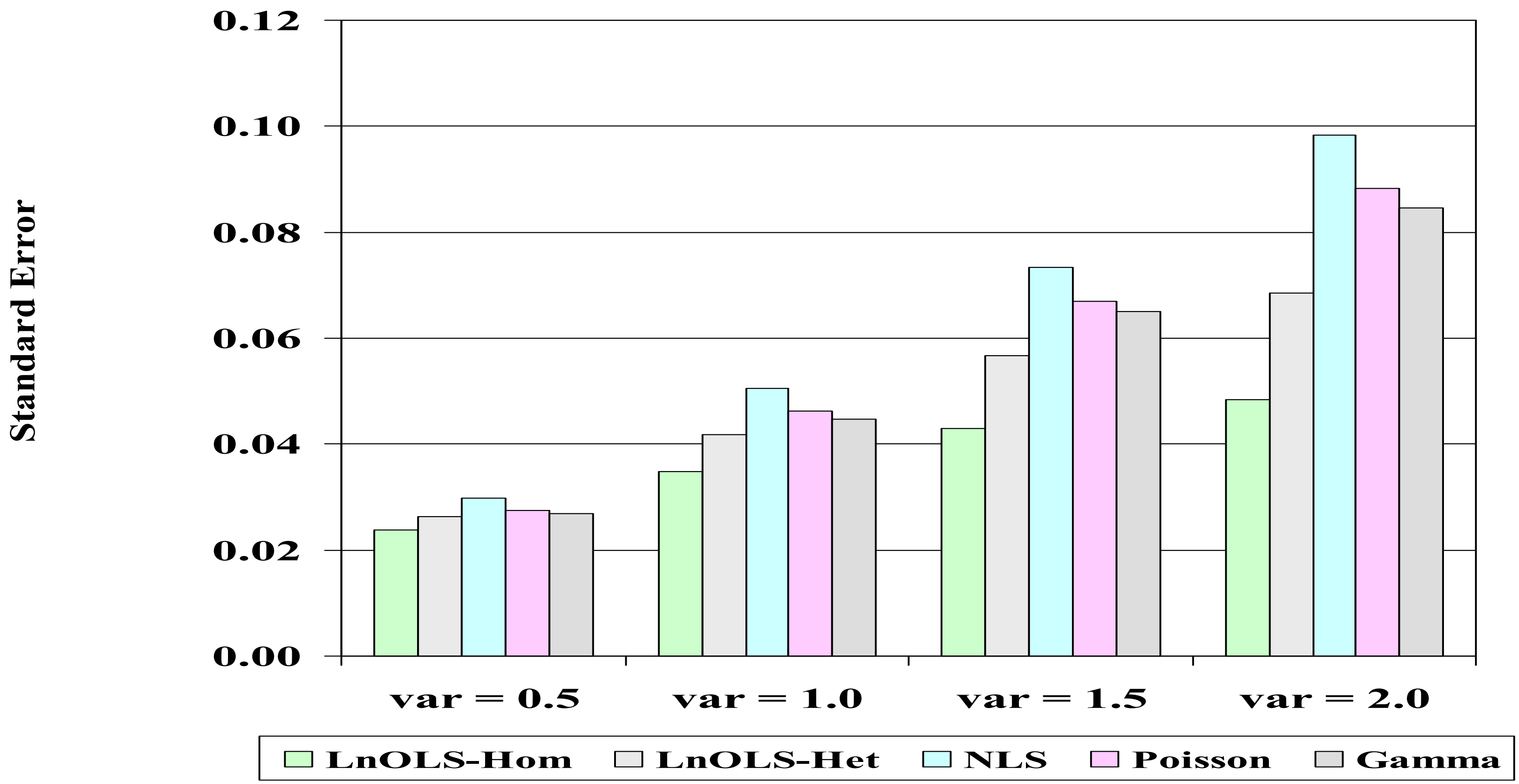


Figure 2
Effect of Heavy Tails on Log Scale

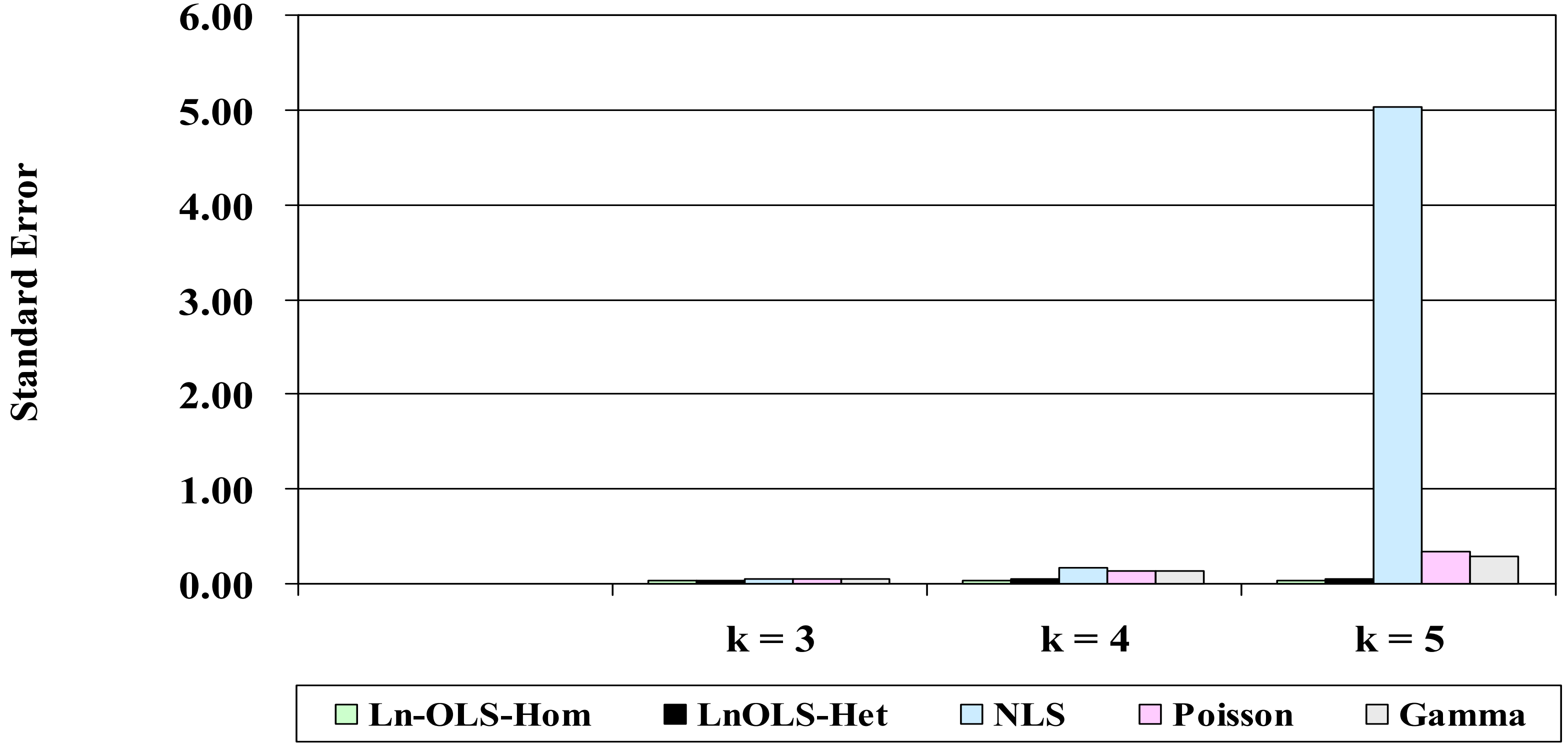


Figure 3
Effect of Shape

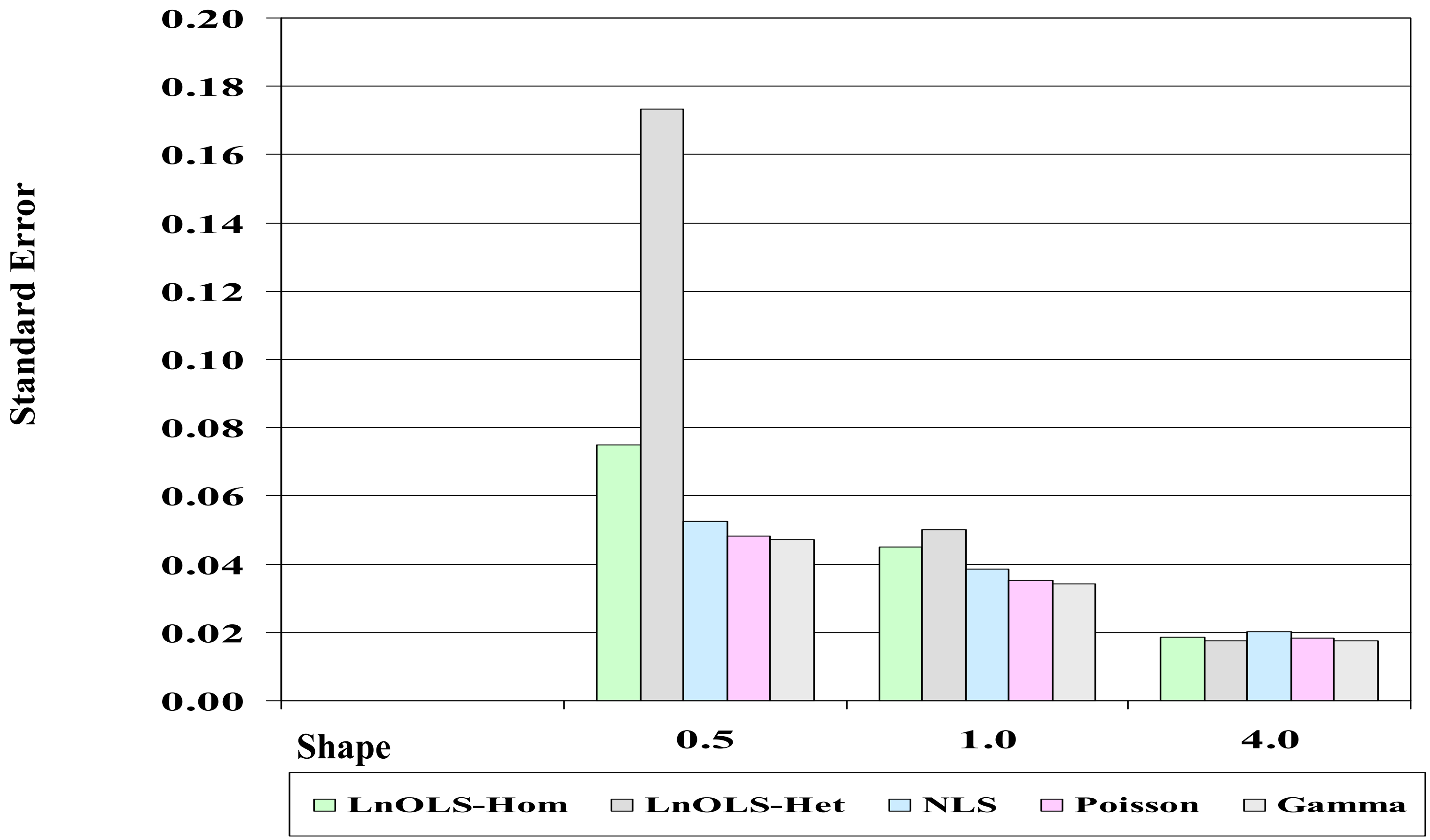
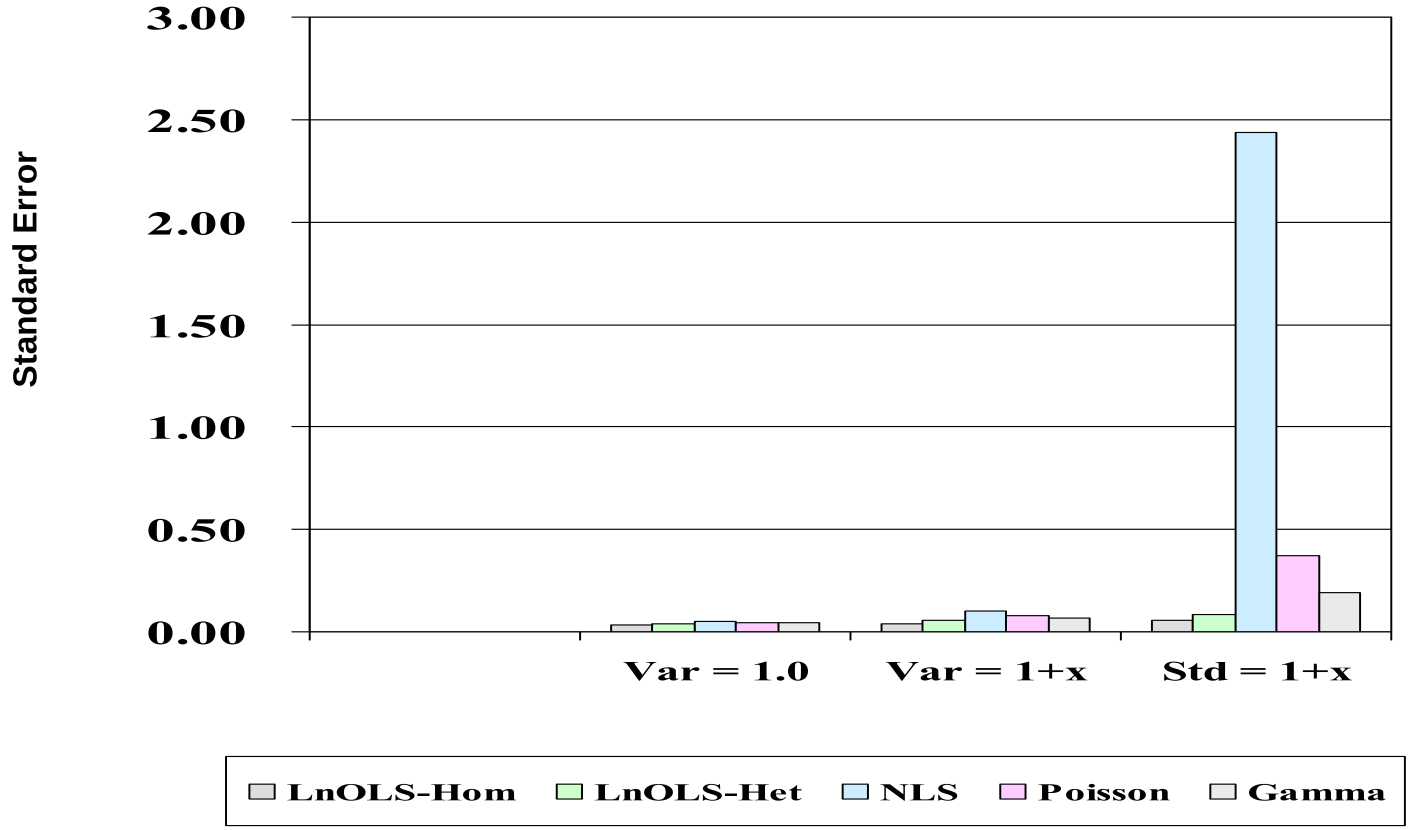


Figure 4
Effect of Heteroscedasticity on the Log Scale



Summary of Simulation Results

All are consistent, except Log-OLS with homoscedastic retransformation if the log-scale error is actually heteroscedastic.

GLM models suffer substantial precision losses in face of heavy-tailed (log) error term. If kurtosis > 3, substantial gains from least squares or robust regression.

Substantial gains in precision from estimator that matches data generating mechanism.

Overview

Studies with skewed outcomes but no zero mass problem

Alternative models

Comparing alternative models

Assessing model fit

Interpretation

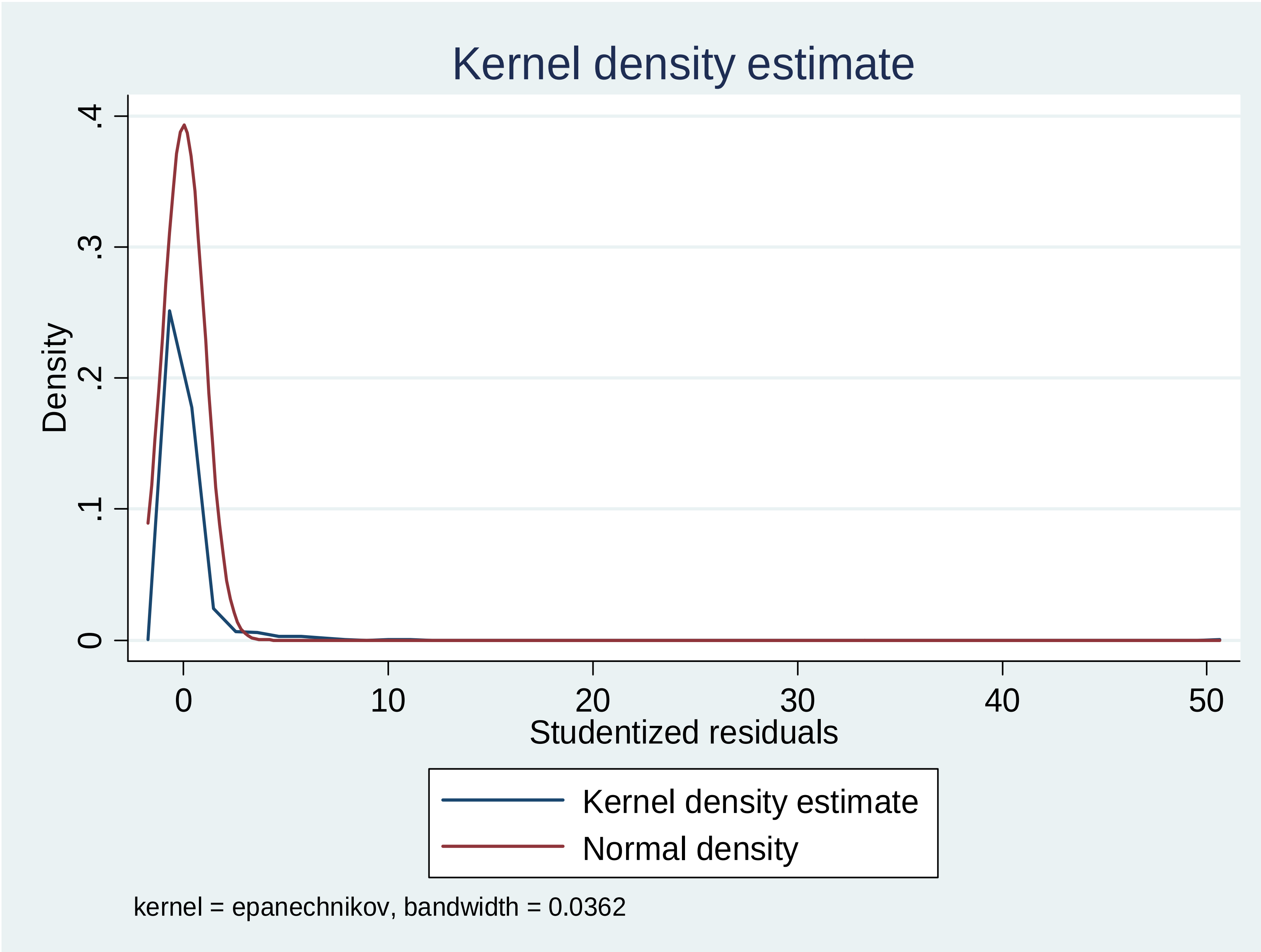
MEPS Data for Examples

Medical Expenditure Panel Survey (MEPS) data
- Representative of non-institutionalized US population
- Data available from 1996 - 2009
- Subsample of NHIS
- Available to public

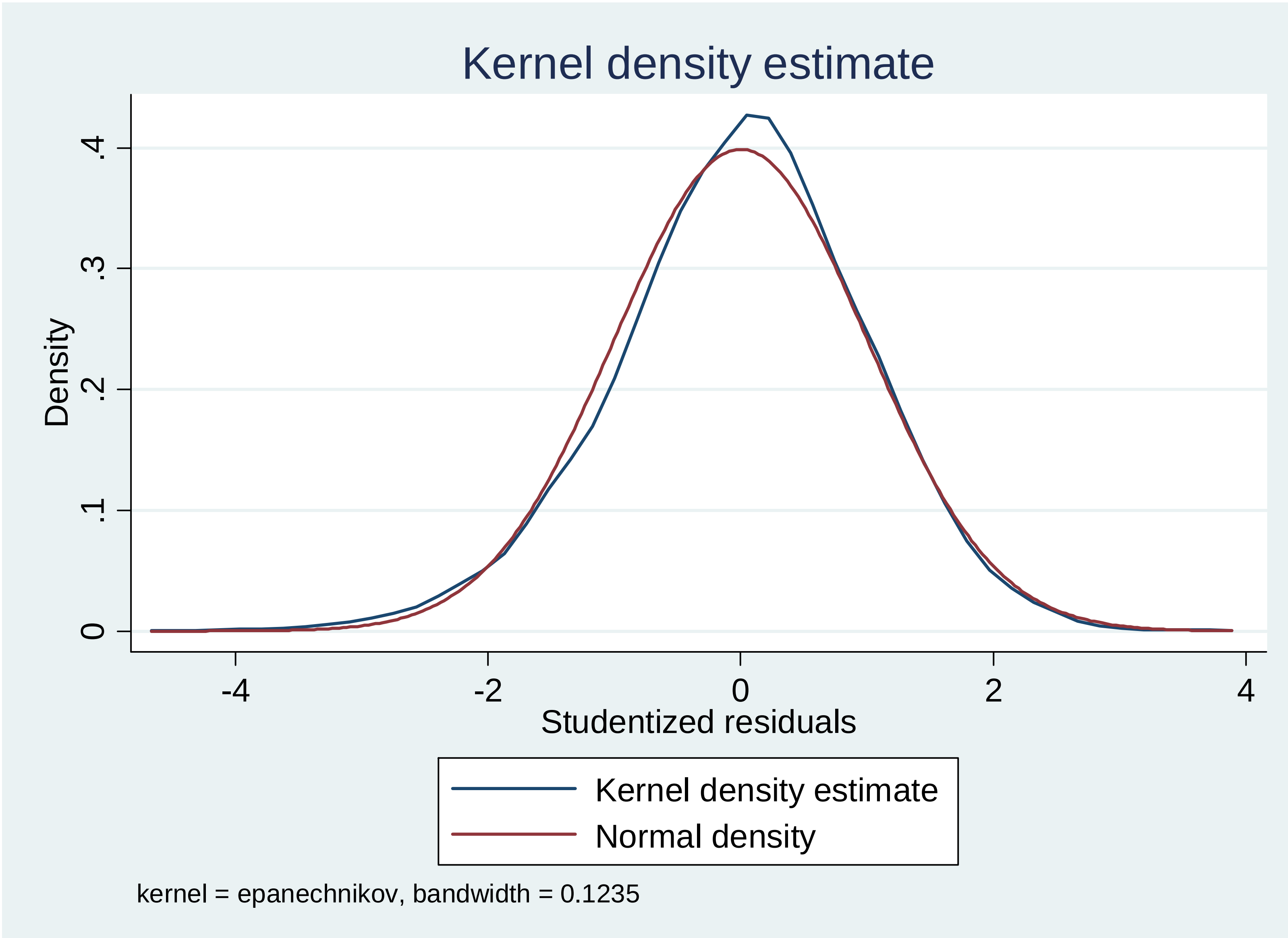
Information on
- Health expenditures and utilization
- Health status
- Insurance
- Demographics, income, education, family

MEPS sample for these examples
- Observations at person-year level, N = 19,386.
- Adults (ages 18+) without missing data.
- Year = 2009.

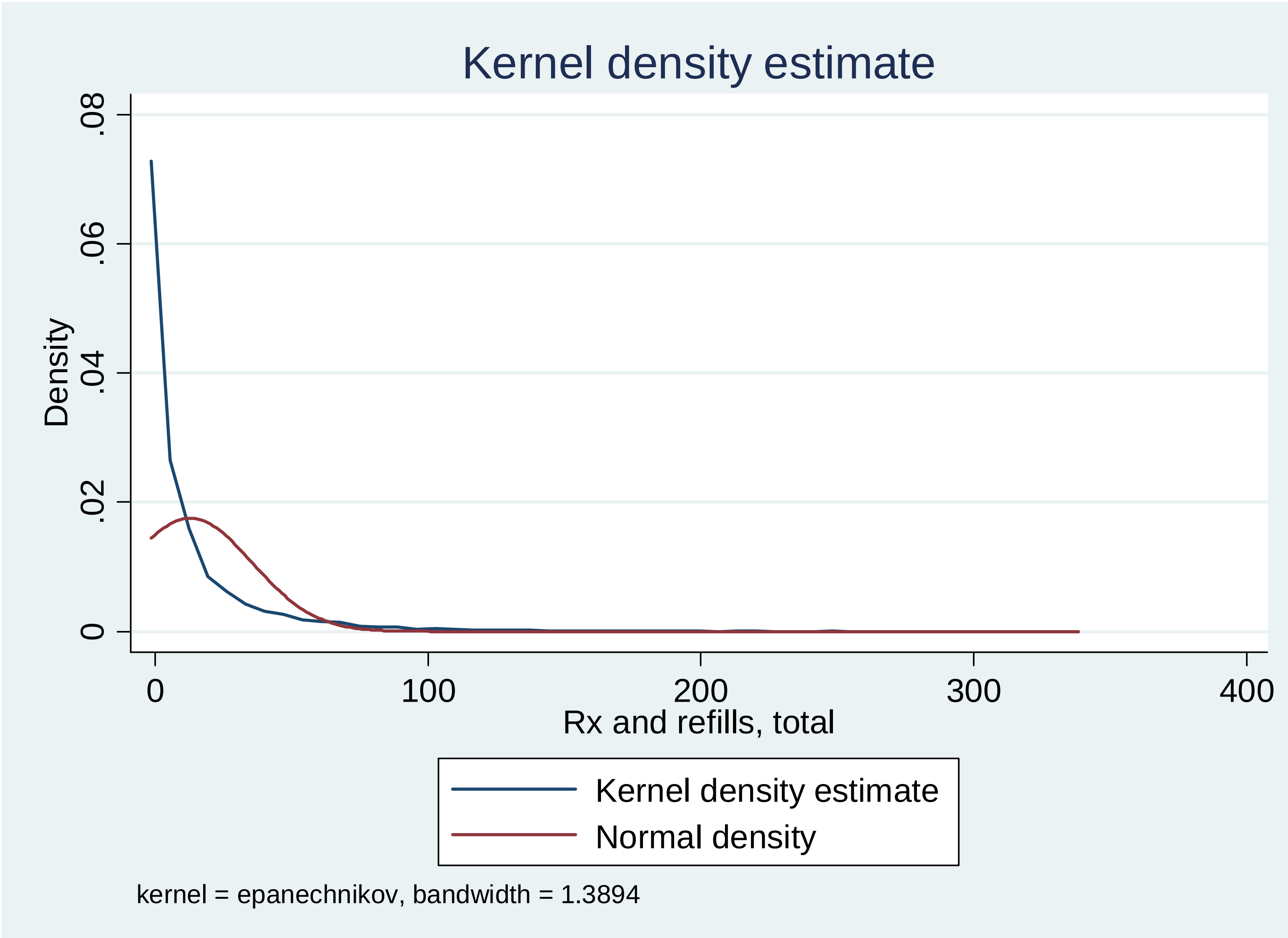
Density for OLS Studentized Residuals Total Medical



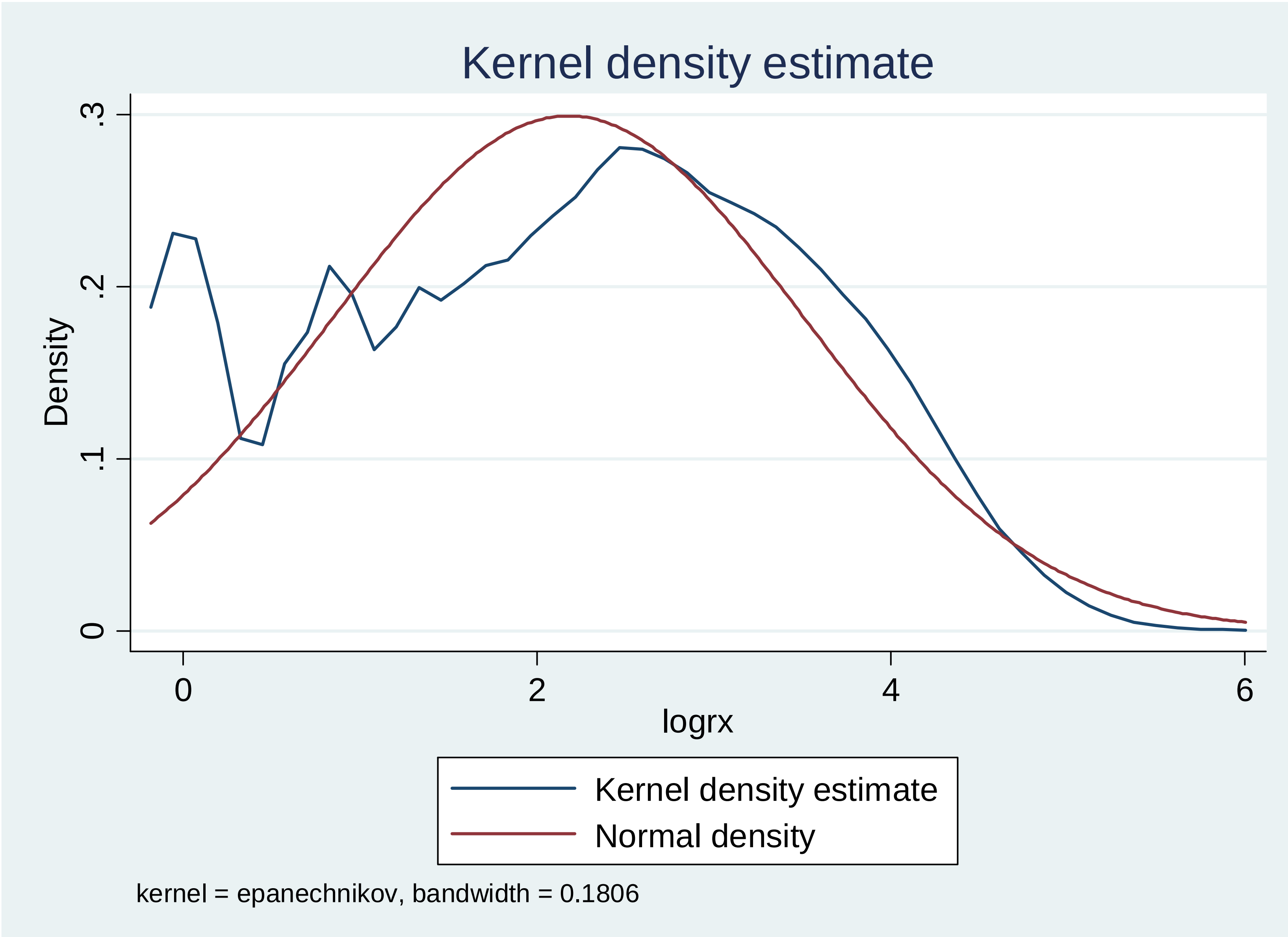
Density for OLS Studentized Residuals, Log Scale Total with $ > 0



Density for Number of Rx fills and refills



Density for log(# Rx fills | Rx > 0)



Overview

Statistical issues and potential problems

Skewness

Studies with skewed outcomes but no zero mass problems

Model checks

Studies with zero mass and skewed outcomes

Studies with count data

Conclusions

Overview

Studies with skewed outcomes but no zero mass problem

- **Assessing model fit**
  - **Picking a model**
    - **Box-Cox test**
    - **GLM family test**

  - **Checking for heteroscedasticity**

  - **Checking model fit**
    - **Pregibon's Link Test and Ramsey's RESET test**
    - **Modified Hosmer-Lemeshow test**

  - **Tests of overfitting and cross validation**
    - **Copas test**

Model Checks

Primary Concern

Systematic bias as a function of covariates *x*

Secondary Concern

Efficiency

Tertiary Concern

Ease of use

Tests can be modified for most models considered here

Most are easily implemented in Stata

MEPS Data for Examples

Dependent variables for examples

Variable	Mean	if y > 0 Mean	if y > 0 Std Dev	Pos. % %
Total medical $	3386	4480	10604	82.2
Total dental $	211	566	978	37.3
# Prescriptions	13	19	25	66.6

Box-Cox Test

Purpose

To determine relationship between $x\beta$ and E(y|x)

Box-Cox test

Find MLE value of λ $$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda}$$

Stata command `boxcox y $indv if y > 0`

Conclude

If $\hat{\lambda} = -1$ inverse $(1/y) = X\beta + \varepsilon$

If $\hat{\lambda} = 0$ ln(y) $\ln(y) = X\beta + \varepsilon$

If $\hat{\lambda} = .5$ square root $\sqrt{y} = X\beta + \varepsilon$

If $\hat{\lambda} = 1$ linear $y = X\beta + \varepsilon$

If $\hat{\lambda} = 2$ square $y^2 = X\beta + \varepsilon$ **(if skewed left)**

Box-Cox Test

Examples

Variable	$\hat{\lambda}$	Conclusion
Total medical	.0637 (.0038)	Close to log*
Total dental	-.1144 (.0080)	Close to log*
# Prescriptions	.0744 (0063)	Close to log*

*** but significantly different from zero (log) at p < 0 .05.**

Note: λ is called \theta in Stata for some versions of the test.

Checking for Heteroscedasticity

Concern is retransformation bias under log(y) or Box-Cox transformation

Use one of standard tests for heteroscedasticity on log-scale

Breusch - Pagan - Godfrey – White test

Park test – GLM version

Consider $\ln(y_i) = x_i' \beta + \varepsilon_i$

Use least square residuals on log scale to create

$$\text{logvar}_i = \left(\ln(y_i) - x_i' \hat{\beta}\right)^2$$

Estimate response logvar to x's by GLM (gamma, log link)

```
glm logvar$indv,family(gamma)link(log)robust
    test  $indv
```

Or use alternative test for heteroscedasticity

Total Medical Expenditures, if Positive MEPs 2004, Adults

Significantly heteroscedastic in

- o Decreasing variance in age and being female ($p < 0.001$)
- o Higher variance for blacks (p == 0.003)
- o Decreasing variance in income and education ($p < 0.001$)

- Borderline increasing variance for uninsured (p= 0.09)
- Not significant in health status / functioning

Complex heteroscedasticity probably rules out OLS on log(total medical expenditures) in favor of GLM

- Studentized residuals too skewed and heavy-tailed for normal theory model, bias in retranformation
- Group-wise smearing will have major precision losses.

GLM Family Test

Purpose

Determine relationship between raw-scale mean and variance functions, E(y|x), and Var (y|x)

Use a GLM family test that is modified Park test

[Requires correct link function, check link or Box-Cox first]

```
glm  y $indv, family(gamma)link(log)
     predict xbetahat, xb
gen  rawresid = y – exp(xbetahat)
gen  rawvar = rawresid^2
glm  rawvar xbetahat, f(gamma)link(log)
```

coefficient on xbetahat indicates distribution

Stata: see sample programs

GLM Family Test (cont'd)

OLS alternative:, but possible retransformation bias

1. **Regress y (raw scale) on x**
2. **Save raw scale residuals $\hat{r}$, predicted value of y**
3. **Regress ln($\hat{r}^2$) on $\ln\left(\hat{y}\right)$ and a constant**

Because use of log transform of residual squared raises a retransformation bias issue, GLM version preferred

GLM Family Test (cont'd)

Coefficient on xbetahat = ln($\hat{y}$) gives the family

If $\hat{\gamma} = 0$ Gaussian NLLS (variance unrelated to mean)
If $\hat{\gamma} = 1$ Poisson (variance equals mean)
If $\hat{\gamma} = 2$ Gamma (variance exceeds mean)
If $\hat{\gamma} = 3$ Wald or inverse Gaussian

Variable	$\hat{\gamma}$	Std. Error	Conclusion
Total medical	1.7211	0.1099	Gamma*
Total dental	1.5259	0.2799	Gamma
# Prescriptions	1.4392	0.1474	Either Gamma or Poisson*

***For total medical, # Prescriptions, Gamma is consistent, but not efficient, because specific variance function for gamma is not optimal.**

GLM Variance Structure (cont'd)

Alternative setup:

Estimate link and variance power functions simultaneously, using

$$E(y \mid x) = \mu = g^{-1}(x\beta)$$

$$g(\mu_i) = (\mu_i^{\lambda} - 1) / \lambda$$

$$V(y_i) = \theta_1 (\mu)^{\theta_2}$$

Estimate λ, θ's, and β jointly.

Allows for λ ≠ 0 (non-log) models and θ_2 ≠ integer.

Basu and Rathouz's extended estimating equation or GLM approach (Biostatistics, 2005; code available from STATA Journal 5(4))

Example: Total medical expenditures if positive.

Parameter	*coef.*	*Std. err.*	*z*	*p-value*
λ	0.1468	0.0477	3.08	0.002
θ_2	1.6540	0.1183	13.98	0.000

Reject log link ($\lambda = 0$) at $p = 0.002$
Reject Poisson family ($\theta_2 = 1$) at $p < 0.001$
Reject Gamma family($\theta_2 = 2$) at $p = 0.004$

Power link of 0.15 reduces tests of nonlinearity on scale of estimation or interest, but does not eliminate results

Major problem in fit is still specification of age and gender terms, especially for ages ≥ 65

Rejects log decisively for dental expenditures and Rx

Assessing the Model Fit for Linearity

Pregibon's Link Test (scale of estimation)

Ramsey's RESET Test (scale of estimation)

Modified Hosmer-Lemeshow (scale of estimation or scale of interest)

Link and RESET Tests

Purpose

To determine linearity of response on scale of estimation

These tests work for any model (e.g., OLS, logit, probit)

Pregibon's Link test for least squares

$$y = \delta_0 + \delta_1 (x\hat{\beta}) + \delta_2 (x\hat{\beta})^2 + \nu$$

Test $\hat{\delta}_2 = 0$

Stata: `linktest`

Ramsey's RESET test (one version, as implemented in Stata)

$$y = \delta_0 + x\delta_1 + \delta_2 (x\hat{\beta})^2 + \delta_3 (x\hat{\beta})^3 + \delta_4 (x\hat{\beta})^4 + \nu$$

Test $\hat{\delta}_2 = \hat{\delta}_3 = \hat{\delta}_4 = 0$

Stata: `ovtest`

Link and RESET Tests (cont'd)

For alternative estimators with linear index: $x'\beta$

Use original estimator with functions of $x'\hat{\beta}$, and $(x'\hat{\beta})^2$ as covariates

STATA example:

```
logit $depv $indv
      predict xbeta, xb
gen   xbeta2 = xbeta^2
logit $depv xbeta xbeta2, robust
      test xbeta2
```

Similarly for RESET.

Link and RESET Tests (cont'd)

Conclude

These tests are diagnostic, not constructive

If do not reject null, keep model the same
If reject null, there could be problem with functional form

Example for log(y) by OLS version

	***p*-values**		
Variable	**Link**	**RESET**	**Conclusion**
Total medical	**0.33**	**<0.001**	**RESET problem**
Total dental	**0.74**	**0.56**	**No problem**
# Prescriptions	**0.03**	**<0.001**	**RESET problem**

Similar conclusions for gamma GLM with log link.

Link and RESET Tests (cont'd)

Advantages

- **Easy**
- **Omnibus tests**

Disadvantages

- **Incomplete for multipart models**
- **Sensitive to influential outliers, especially RESET**

Modified Hosmer-Lemeshow Test

Purpose

To check fit on scale of interest or raw scale for systematic bias

Modified Hosmer-Lemeshow test

Estimate model (e.g., glm y or OLS $\ln(y) = x\beta + \varepsilon$)

Retransform to get $\hat{y}$ on raw scale

Compute raw-scale residual $\hat{r} = y - \hat{y}$

Create 10 groups, sorted by specific x (or by $x\hat{\beta}$)

***F*-test of whether all 10 mean residuals different from zero**

Look for systematic patterns (e.g., U-shaped pattern)

Stata: see sample program

Modified Hosmer-Lemeshow Test (cont'd)

Conclude

This test is also non-constructive

No problem if there is no systematic pattern

If reject null, there could be problem with either

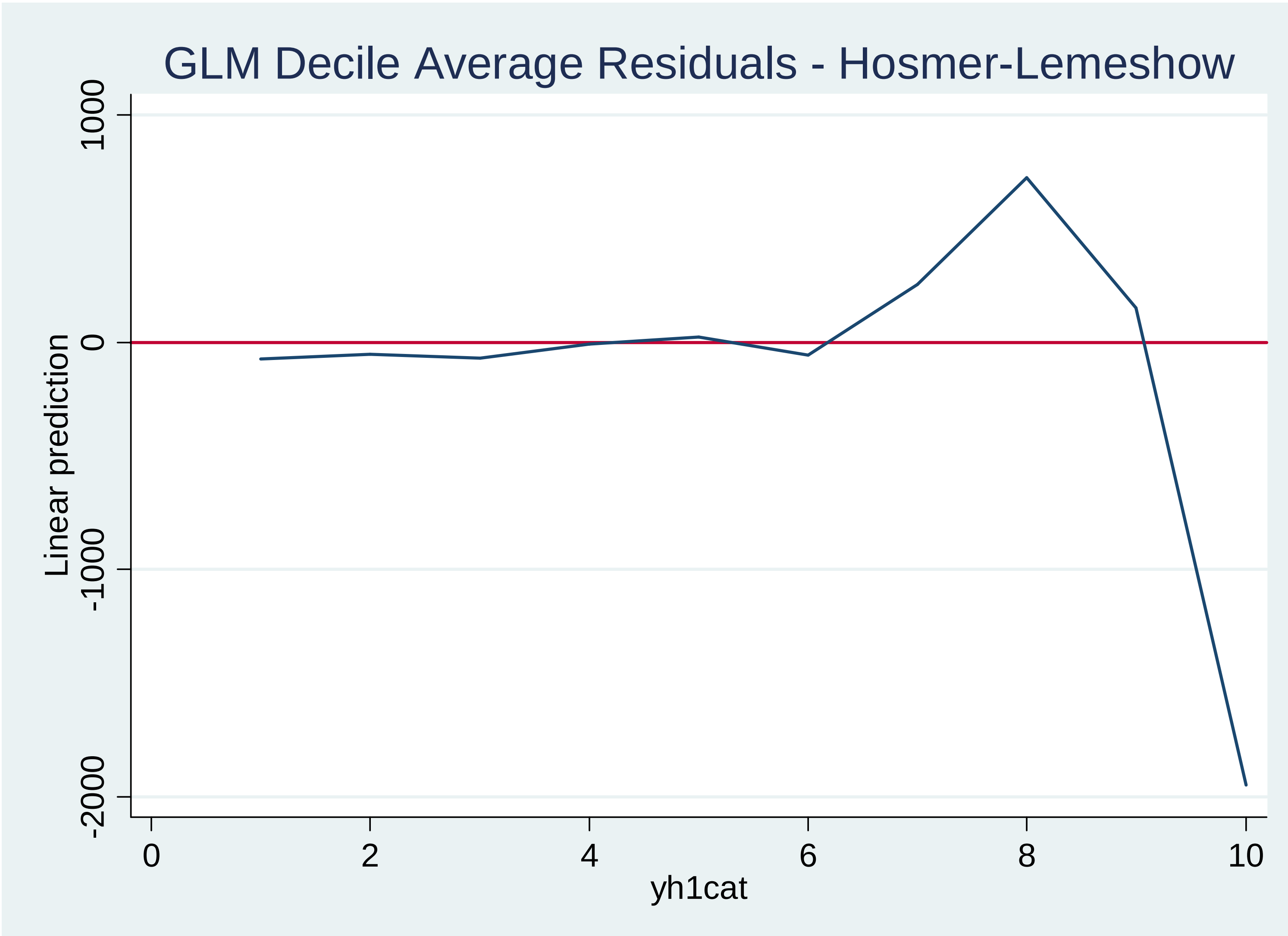
- **left side (wrong power or link function)**
- **right side (wrong functional form of *x's*)**

Example for gamma with log link

Variable	*p*-value for *F*-test	Conclusion
Total medical	< 0.001	Problem*
Total dental	0.02	Problem*
# Prescriptions	< 0.001	Problem*

*** Problem largely in age-gender specification.**

Modified Hosmer Lemeshow Estimates by Deciles of Prediction for Total Medical Expenditures



Modified Hosmer-Lemeshow Test (cont'd)

Test linearity fit for power function matters for total medical expenditures (Totexp), positive cases only.

Use modified Hosmer-Lemeshow, Pregibon's Link, and Ramsey's RESET tests on estimation scale. OLS example.

Variable	Power Function	Hosmer-Lem. *F*-test	Pregibon Link Test t	Ramsey RESET
Totexp	1.0	90.45	6.29	56.55
Totexp^0.5	0.5	15.26	7.26	51.11
Log(Totexp)	0.0	4.94	0.97 (ns)	18.24
(1/Totexp)^0.5	-0.5	35.76	5.38	81.81

All significant at $p < 0.01$ except not significant (ns)

Modified Hosmer-Lemeshow Test (cont'd)

All models considered fail specification tests, except log transform using Link Test

For all tests best fit for transforms considered is log model

Remaining specification failure is largely due to age-gender specification, not fine-tuning transformation

Modified Hosmer-Lemeshow Test (cont'd)

Advantages

- **Works on scale of ultimate interest, as well as on scale of estimation; can choose scale concerned about**
- **Works for any model (including logit, probit, 2-part, NB)**
- **Can detect problems missed by omnibus Link and RESET tests, because can look at fit for key covariates**

Disadvantage

- **Lacks power**
- **Individual coefficients sensitive to influential observations if done on scale of interest (raw scale)**

Copas Tests

***Overfitting* can be a problem**

Tailoring the model to the specific data set

But at the expense of explaining other similar data sets

Overemphasis on explaining a few outliers when data are very skewed or cases have leverage. For OLS,

$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon$ **but some ε's are extremely large as well as x's extreme – risk of influential outlier**

Overfitting is often a major problem for expenditure data esp. for small to moderate sample sizes or rare covariates.

Maximizing *R*-squared leads to overfitting.

Copas Test (cont'd)

Purpose

To test for over-fitting using split sample cross validation

Copas test (one version of it)

Randomly split sample into two equal groups A and B

Estimate model on sample A, retain coefficients $\hat{\beta}_A$

Forecast to sample B

$$\hat{y}_B = X_B \hat{\beta}_A$$

Regression model for sample B

$$y_B = \delta_0 + \delta_1 \hat{y}_B + \eta$$

Test $\hat{\delta}_1 = 1$

Repeat 999 more times to get distribution

Variation: also forecast results from sample B on A

Stata: see sample program

Copas Test (cont'd)

Examples for total medical expenditure

Estimator	δ	δ=1?	Std. err
GLM	0.989	ns	0.0674
OLS on raw	0.991	ns	0.0413

Least squares would have to have 2.65 sample size to have same precision as GLM.

Mean absolute deviation larger with least squares.

If $\hat{\delta}$ significantly different from one, consider outliers or pruning model.

If $\hat{\delta}$ quite imprecise, consider more efficient or robust methods.

Copas Test (cont'd)

Advantages

Can detect overfitting, which can be a big problem

Overfitting not detected by other within sample tests

***R*-squared envy may lead to overfitting**

Disadvantages

One influential case can cause problems

One influential case is always in sample A or in B

GLM downweights catastrophic and may appear to do poorly if such case shows up in test sample.

Summary of MEPS modeling
Positive Total Medical Expenditures

Standard OLS log($) subject to complex heteroscedasticity and error not normally distributed.

- Normal theory models will be biased on retransformation by failure of normality
- Potential bias for estimates of impact of x on E($ | x, $ > 0) due to heteroscedasticity

Log transform overcorrects in Box-Cox family and log link is not optimal for GLM

- Potential for some bias

Distribution for GLM is neither Poisson-like nor Gamma family for QMLE

- Efficiency gains from using EEE or iteratively reweighted least squares

Summary of MEPS modeling(cont'd)

Simple age, gender and age•gender specification not adequate

- **Mispredicts most expensive group – the elderly**

GLM (log link, gamma) more precise than OLS on raw dollars

OLS more susceptible to influential outliers

- **Here issue is expensive cases with any health limitation.**
- **Important but uncommon subgroup.**

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Overview

Studies with skewed outcomes but no zero mass problem

- **Alternative models**
- **Comparing alternative models**
- **Assessing model fit**
- **Interpretation**

Overview

Studies with skewed outcomes but no zero mass problem

- **Interpretation**
  - **Marginal effects**
  - **Incremental effects**
  - **Interaction effects**

Single-Equation Models for $y > 0$

Interpretation

- **$E(y \mid X)$**
- **Marginal and interaction effects**

Models

- **OLS**
- **$\ln(y)$ for normal and homoskedastic error**
- **$\ln(y)$ for non-normal and heteroskedastic error**
- **Square root of y, as example of Box-Cox**
- **GLM with log link**

Marginal and Interaction Effects (1)

Compute marginal and interaction effects

Also compute E(y | X)

Compare five different single-equation models

Use the same MEPS 2004 data

Show formulas for general model

Show basic Stata code

[Will not discuss weights, design effects]

Marginal and Interaction Effects (2)

Marginal effects

For continuous variables

Take derivative

Incremental effects

For dummy variables

Also for discrete change in continuous variables

Take discrete difference

Interaction effects

Take double derivative or double difference

Marginal and Interaction Effects (3)

Marginal effects in linear models (OLS) are easy

Marginal effects in nonlinear models are more complicated
Several ways to compute them

- **For full sample**
  - **Recycled predictions**
  - **Average-of-the-probabilities approach**
- **For a single, typical observation**
- **Can change value for subsample or whole sample**
- **Stata's `mfx` command is often misleading**

The appropriate method depends on the research question

Marginal and Interaction Effects (4)

Basic model

$$y = \beta_0 + \beta_1 age + \beta_2 female + \beta_{12} age \times female + X\beta + \varepsilon$$

Marginal effect of age operates through *two terms*

Covariates X include controls for

Race, income, education, health status, and insurance

Base case person is white & non-Hispanic,

ln(inc) = 10 (inc. = $22,026), less than high school education, mental & physical health status = 50 (mean), limitations = 0, private insurance

Marginal and Interaction Effects (5)

Mean of *y*

Averaged over full sample

45-year old woman

Marginal effect of age

Averaged over full sample

45-year old woman

Interaction effect of age and gender

Averaged over full sample

Base-case person

All calculations are conditional on covariates

OLS (1)

Model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon = X\beta + \varepsilon$$

Interpretation

$$E\left(y \mid X\right) = X\hat{\beta}$$

$$\frac{dE\left(y \mid X\right)}{dx_1} = \hat{\beta}_1 + \hat{\beta}_{12} x_2$$

this has two terms!

$$\frac{\partial^2 E\left(y \mid X\right)}{\partial x_1 \partial x_2} = \hat{\beta}_{12}$$

OLS (2)

```
reg $y $rhs, robust

predict yhat
gen me_age = _b[age] + _b[fage]*female
gen ie = _b[fage]
```

Notation

- `hat` **for predicted values**
- `me` **for marginal effect**
- `ie` **for interaction effect**

See sample code for calculations for typical observation

- **45-year old woman**

OLS (3)

Mean of *y*	
Full sample	**\$3,830**
45-year old woman	**\$3,756**
Marginal effect of age	
Full sample	**\$18**
45-year old woman	**\$6**
Interaction effect of age and gender	
Full sample	**–\$26**
Base-case person	**–\$26 (exactly same)**

ln(*y*), Normal and Homoskedastic (1)

Model

$$\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon = X\beta + \varepsilon$$

Interpretation

$$E(y \mid X) = \exp\left(X\hat{\beta} + 0.5\hat{\sigma}^2\right)$$

$$\frac{dE(y \mid X)}{dx_1} = \left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right) E(y \mid X)$$

$$\frac{d^2 E(y \mid X)}{dx_1 dx_2} = \left[\hat{\beta}_{12} + \left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right)\left(\hat{\beta}_2 + \hat{\beta}_{12} x_1\right)\right] E(y \mid X)$$

FYI: ln(*y*), Normal and Homoskedastic (2)

Incremental effects: discrete changes in age and gender

$$\frac{\Delta E(y|X)}{\Delta age} = \left(e^{X\hat{\beta}|age=50} - e^{X\hat{\beta}|age=40}\right)e^{0.5\hat{\sigma}^2}$$

$$\frac{\Delta E(y|X)}{\Delta gender} = \left(e^{X\hat{\beta}|female=1} - e^{X\hat{\beta}|female=0}\right)e^{0.5\hat{\sigma}^2}$$

$$\frac{\Delta^2 E(y|X)}{\Delta age \Delta gender} = \begin{pmatrix} e^{X\hat{\beta}|age=50,female=1} - e^{X\hat{\beta}|age=50,female=0} \\ -e^{X\hat{\beta}|age=40,female=1} + e^{X\hat{\beta}|age=40,female=0} \end{pmatrix} e^{0.5\hat{\sigma}^2}$$

ln(y), Normal and Homoskedastic (3)

```
reg $lny $rhs, robust
predict xbetahat, xb

gen yhat = exp(xbetahat + .5*e(rmse)^2)
gen me_age = (_b[age]+_b[fage]*female)*yhat
gen ie = (_b[fage]
          + (_b[age] + _b[fage]*female)
          *(_b[female] + _b[fage]*age))*yhat
```

ln(y), Normal and Homoskedastic (4)

Mean of *y*

Full sample	**\$4,733**
45-year old woman	**\$3,486**

Marginal effect of age

Full sample	**\$79**
45-year old woman	**\$40**

Interaction effect of age and gender

Full sample	**–\$54**
Base-case person	**–\$35**

ln(*y*), Non-Normal and Heteroskedastic (5)

Mean of *y*

Full sample	**$4,431**
45-year old woman	**$3,136**

Marginal effect of age

Full sample	**$76**
45-year old woman	**$36**

Interaction effect of age and gender

Full sample	**–$58**
Base-case person	**–$35**

Standard Errors (1)

Bootstrap the standard errors

Draw repeated samples, with replacement
Re-estimate the model
Re-compute all statistics of interest (means, marginal effects)
Do this 1000 times

In Stata use the `bootstrap` command

The example below would need to be modified for the specific model

Standard Errors (2)

```
program define me_lny, rclass
   tempvar xbetahat yhat me_age ie

   reg $lny $rhs if $y>0
   predict `xbetahat', xb

   gen `yhat'= exp(`xbetahat' + .5*e(rmse)^2)
   gen `me_age' = (_b[age]
     +_b[fage]*female)*`yhat'
   gen `ie' = (_b[fage] + (_b[age] +
     _b[fage]*female)*(_b[female] +
     _b[fage]*age))*`yhat'
```

Standard Errors (3)

```
   sum `me_age', meanonly
   return scalar me_age = r(mean)
   sum `ie', meanonly
   return scalar ie = r(mean)
end

bootstrap me_age = r(me_age) ie=r(ie),
   reps(1000): me_lny
```

Duan's (JASA 1983) Smearing Estimator

Corrects for non-normality in log models
Does *not* directly correct for heteroscedasticity
Can be adapted for heteroskedasticity by subgroup

Stata code

```
regress lny $x
predict double resid, residual
egen Dsmear = mean(exp(resid))
```

The smearing factor is always greater than 1.0, and is typically less than 4.0.

ln(*y*), Non-Normal and Heteroskedastic (1)

Model

$$\ln\left(y\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon = X\beta + \varepsilon$$

Interpretation

$E\left(y \mid X\right) = \exp\left(X\hat{\beta}\right)\hat{D}\left(X\right)$ **where** $\hat{D}\left(X\right)$ **is function of** X

$$\frac{dE\left(y \mid X\right)}{dx_1} = \left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right) E\left(y \mid X\right) + \exp(X\hat{\beta}) \frac{d\hat{D}\left(X\right)}{dx_1}$$

$$\frac{d^2 E\left(y \mid X\right)}{dx_1 dx_2} = \left[\hat{\beta}_{12} + \left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right)\left(\hat{\beta}_2 + \hat{\beta}_{12} x_1\right)\right] E\left(y \mid X\right)$$

$$+ \left[\left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right)\frac{d\hat{D}}{dx_2} + \left(\hat{\beta}_2 + \hat{\beta}_{12} x_1\right)\frac{d\hat{D}}{dx_1} + \frac{d^2\hat{D}}{dx_1 dx_2}\right] \exp\left(X\hat{\beta}\right)$$

FYI: ln(*y*), Non-Normal and Heteroskedastic (2)

Previous equations have derivatives of Duan smearing factor

- **Derivatives = 0 if homoskedastic (Duan smearing is scalar)**
- **Can only take derivatives if Duan smearing = f(*X*)**
- **See Ai and Norton (*JHE* 2000)**

Typically done by subgroup

- **Compute scalar Duan smearing factor by subgroup**
- **i.e., by gender, age group, insurance status, or health status**
- **See Manning (*JHE* 1998)**

ln(y), Non-Normal and Heteroskedastic (3)

```
reg $lny $rh, robust

predict double ehat, residual
gen ehat2 = ehat^2
egen Dsmear = mean(exp(ehat))

reg ehat2 $rhs, robust /* test for hetero */

egen Dm = mean(exp(ehat)) if female==0
egen Dsmearm = max(Dm)
egen Df = mean(exp(ehat)) if female==1
egen Dsmearf = max(Df)
```

ln(y), Non-Normal and Heteroskedastic (4)

```
quietly reg $lny $rhs if $y>0, robust
predict xbetahat, xb

gen yhat = exp(xbetahat)*Dsmearm if
   female==0
replace yhat = exp(xbetahat)*Dsmearf if
   female==1
gen me_age = (_b[age]+_b[fage]*female)*yhat
gen ie = (_b[fage]+(_b[age]+_b[fage]*female)
   *(_b[female] + _b[fage]*age))*yhat
   +(_b[age]+_b[fage]*female)*exp(xbetahat)
   *(Dsmearf - Dsmearm)
```

ln(y), Non-Normal and Heteroskedastic (5)

Mean of y	
Full sample	**\$5,073**
45-year old woman	**\$3,136**
Marginal effect of age	
Full sample	**\$86**
45-year old woman	**\$36**
Interaction effect of age and gender	
Full sample	**–\$68**
Base-case person	**–\$35**

Square Root (1)

Model

$$\sqrt{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon = X\beta + \varepsilon$$

Interpretation

$$E(y \mid X) = \left(X\hat{\beta}\right)^2 + \hat{V} \qquad \hat{V} = \frac{1}{N-k-1}\sum \hat{\varepsilon}^2 \text{ if homosk.}$$

$$\frac{dE(y \mid X)}{dx_1} = 2\left(X\hat{\beta}\right)\left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right) + \frac{d\hat{V}}{dx_1} \quad \text{if heteroskedastic}$$

$$\frac{\partial^2 E(y \mid X)}{\partial x_1 \partial x_2} = 2\left[\left(X\hat{\beta}\right)\hat{\beta}_{12} + \left(\hat{\beta}_1 + \hat{\beta}_{12} x_2\right)\left(\hat{\beta}_2 + \hat{\beta}_{12} x_1\right)\right] + \frac{\partial^2 \hat{V}}{\partial x_1 \partial x_2}$$

Square Root (2)

```
gen sqrty = sqrt($y)
reg sqrty $rhs, robust

predict xbetahat, xb
egen Vhat = mean(e(rmse)^2) (if homosk. error)
gen  yhat = xbetahat^2 + Vhat
gen  me_age = 2*(_b[age] +
   _b[fage]*female)*xbetahat
gen ie = 2*(xbetahat*_b[fage] + (_b[age] +
 _b[fage]*female)*(_b[female]+_b[fage]*age))
```

Square Root (3)

Mean of *y*

Full sample	**$4,093**
45-year old woman	**$3,691**

Marginal effect of age

Full sample	**$23**
45-year old woman	**$14**

Interaction effect of age and gender

Full sample	**–$20**
Base-case person	**–$23**

GLM with Log Link (1)

Model

$$\ln\left[E\left(y\right)|X\right]=\beta_0+\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2=X\beta$$

Interpretation

(for all GLM models with log link (Poisson, Gamma, etc.))

$$\hat{y}=\exp\left(X\hat{\beta}\right)$$

$$\frac{dE\left(y|X\right)}{dx_1}=\left(\hat{\beta}_1+\hat{\beta}_{12}x_2\right)\hat{y}$$

$$\frac{d^2E\left(y|X\right)}{dx_1dx_2}=\left[\hat{\beta}_{12}+\left(\hat{\beta}_1+\hat{\beta}_{12}x_2\right)\left(\hat{\beta}_2+\hat{\beta}_{12}x_1\right)\right]\hat{y}$$

GLM with Log Link (2)

```
glm $y $rhs, link(log) family(gamma)

predict yhat, mu

gen me_age = (_b[age] +_b[fage]*female)*yhat
gen ie = (_b[fage]
          + (_b[age]+_b[fage]*female)
          *(_b[female] + _b[fage]*age))*yhat
```

GLM with Log Link (3)

Mean of *y*

Full sample	**$4,051**
45-year old woman	**$3,151**

Marginal effect of age

Full sample	**$46**
45-year old woman	**$18**

Interaction effect of age and gender

Full sample	**–$54**
Base-case person	**–$40**

Overview

Studies with zero mass and skewed outcomes

Zeros problem: brief overview

Interpretation, marginal and interaction effects in two-part and GLM models

The Zeroes Problem (1)

Statement of the problem

- **Often a large mass at zero**
- **These are true zeroes, not censored values**
- **Zero mass may respond differently to covariates**

Examples

- **Expenditures or use**
  - **Inpatient, outpatient, nursing home, Rx**
- **Cigarette smoking**
- **Alcohol consumption**

The Zeroes Problem (2)

Alternative estimators

OLS (ignore the problem)

ln(*y* + *c*), or Box-Cox with two parameters

GLM

Tobit (assume censored normal distribution)

Heckman selection (adjusted or generalized Tobit)

Two-part model

Conditional Density Estimation

Overview

Studies with zero mass and skewed outcomes

Zeros problem: brief overview

Interpretation, marginal and interaction effects in two-part and GLM models

Two-Part Model

Takes advantage of basic rule of probability

$$\mathrm{E}(y|x) = \Pr\left(y>0 \right) \times \mathrm{E}\left(y|y>0 \right)$$

Splits consumption into two parts

1. **Pr(any use or expenditures)**
   Full sample
   Estimate with logit or probit model

2. **Level of expenditures or use**
   Conditional on $y > 0$ (subsample with $y > 0$)
   Use appropriate continuous or count model
   (e.g., OLS, ln(*y*), GLM, truncated count)

Stata Code for Two-Part Model

Example

- `y` is dependent variable on raw scale
- `ydum` is dummy variable indicating if `y>0`
- `lny` is logarithm of `y` if `y>0`

Part 1

```
   logit  ydum $x
or probit ydum $x
```

Part 2

```
   regress y   $x if y>0
or regress lny $x if y>0
or glm     y   $x if y>0, l(log) f(gamma)
or ztnb    y   $x if y>0
```

Predictions in Two-Part Model (1)

Predictions depend on both parts of the model

$$E(y|x) = \Pr(y>0) \times E(y|y>0)$$

First part

Probit $\Pr(y) = \Phi(x\alpha)$

or Logit $\Pr(y) = \dfrac{1}{1+\exp(-X\alpha)}$

Stata code after running first part

```
predict phat
```

Predictions in Two-Part Model (2)

Second part

If the log-scale error term is not Normal

$$\hat{y} = \Phi\left(X\hat{\alpha}\right) \times \exp\left(X\hat{\beta}\right) \times \hat{D}$$

where $\hat{D}$ is Duan's (1983) smearing estimator

If GLM with log link, so ln(E(*y*))=*X*δ

$$\hat{y} = \Phi\left(X\hat{\alpha}\right) \times \exp\left(X\hat{\delta}\right)$$

Predictions in Two-Part Model (2)

Example of Stata code for ln(*y*) with non-Normal errors

```
logit $ydum $x
predict phat

reg $lny $x if $y > 0
predict lnyhat, xb
predict double resid, residual
egen Dsmear = mean(exp(resid))

gen yhat = phat*exp(lnyhat)*Dsmear
```

Predictions in Two-Part Model (3)

Example of Stata code for GLM with log link

```
probit $ydum $x
predict phat

glm $y $x if $y > 0, link(log) family(gamma)
predict yposhat, mu

gen yhat = phat*exp(yposhat)
```

Marginal Effects in Two-Part Model (1)

Continuous variable x_c

$$\frac{\partial \mathrm{E}(y)}{\partial x_c} = \frac{\partial\big(\Pr(y>0)\times \mathrm{E}(y\mid y>0)\big)}{\partial x_c}$$

$$= \left(\Pr(y>0)\ \frac{\partial \mathrm{E}(y\mid y>0)}{\partial x_c}\right) + \left(\mathrm{E}(y\mid y>0)\ \frac{\partial \Pr(y>0)}{\partial x_c}\right)$$

If there is heteroskedasticity, then $\frac{\partial \mathrm{E}\left(y\mid y>0\right)}{\partial x_c}$ may need to account for heteroskedasticity (Duan smearing by group), or use GLM

Marginal Effects in Two-Part Model (2)

Example: Logit first part, ln(*y*) second part

$$\frac{\partial E(y)}{\partial x_c} = \left(\hat{P}\,\hat{\beta}_c\,(\hat{y} \mid y > 0)\right) + \left((\hat{y} \mid y > 0)\,\hat{\alpha}_c\,\hat{P}\left(1 - \hat{P}\right)\right)$$

$$= \left(\hat{\beta}_c + \hat{\alpha}_c\left(1 - \hat{P}\right)\right)\hat{P}\,\hat{y}$$

Where $\hat{P}$ is the Pr(*y*>0), α are first-part parameters, β are second-part parameters, subscript *c* indicates the continuous variable, $(\hat{y} \mid y > 0)$ is the conditional mean, $\hat{y}$ is the unconditional mean, and there are no interaction or higher-order terms

Marginal Effects in Two-Part Model (3)

Example: Probit first part, GLM with log link second part

$$\frac{\partial E(y)}{\partial x_c} = \left(\hat{P}\hat{\beta}_c\left(\hat{y} \mid y > 0\right)\right) + \left(\left(\hat{y} \mid y > 0\right)\hat{\alpha}_c \varphi\left(X\hat{\alpha}\right)\right)$$

$$= \left(\hat{\beta}_c + \hat{\alpha}_c \frac{\varphi\left(X\hat{\alpha}\right)}{\Phi\left(X\hat{\alpha}\right)}\right)\hat{y}$$

Where all notation is as is the previous slide, ϕ is the Normal pdf, and there are no interaction or higher-order terms

Incremental Effects in Two-Part Model

Example: Logit first part, ln(*y*) second part

$$\frac{\Delta E(y)}{\Delta x_d} = \hat{P}\Big(\big(\hat{y} \mid y > 0, x_d = 1\big) - \big(\hat{y} \mid y > 0, x_d = 0\big)\Big)$$

$$+\big(\hat{y} \mid y > 0\big)\Big(\hat{P}\big(x_d = 1\big) - \hat{P}\big(x_d = 0\big)\Big)$$

$$= \hat{P} \bullet \Delta\big(\hat{y} \mid y > 0\big) + \big(\hat{y} \mid y > 0\big) \bullet \Delta(\hat{P})$$

Where all notation is as in the previous slides, the subscript *d* indicates the dichotomous variable, and there are no interaction or higher-order terms

Complications for Marginal Effects

Interaction terms in first part, nonlinear model

- **Interaction terms in nonlinear models are complicated**
- **Must take full derivative or double difference**
- **See Ai and Norton (*Economics Letters* 2003)**

Heteroskedasticity in the second part

- **Heteroskedasticity affects retransformation**
- **Alternatives to scalar smearing factor**
  - **Multiple smearing factors, by group (Manning, 1998)**
  - **Model the heteroskedasticity (Ai & Norton, 2000, 2008)**

Bootstrap the standard errors of marginal effects

BREAK FOR LUNCH

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Overview

Conditional Density Estimation

Motivation

Simple Monte Carlo example

MEPS example

Hints and comments

CDE Introduction

Conditional Density Estimation

We think CDE is an exciting new model, has much potential

Alternative way to model expenditures

- **Flexible density**
- **Allows for zeroes**
- **Can estimate E($y|x$), marginal and incremental effects**

Based on paper by Gilleskie and Mroz (*JHE* 2004)

Related papers

Efron (*JASA* 1988)

Donald, Green, Paarsch (*REStud* 2001)

CDE Motivation (1)

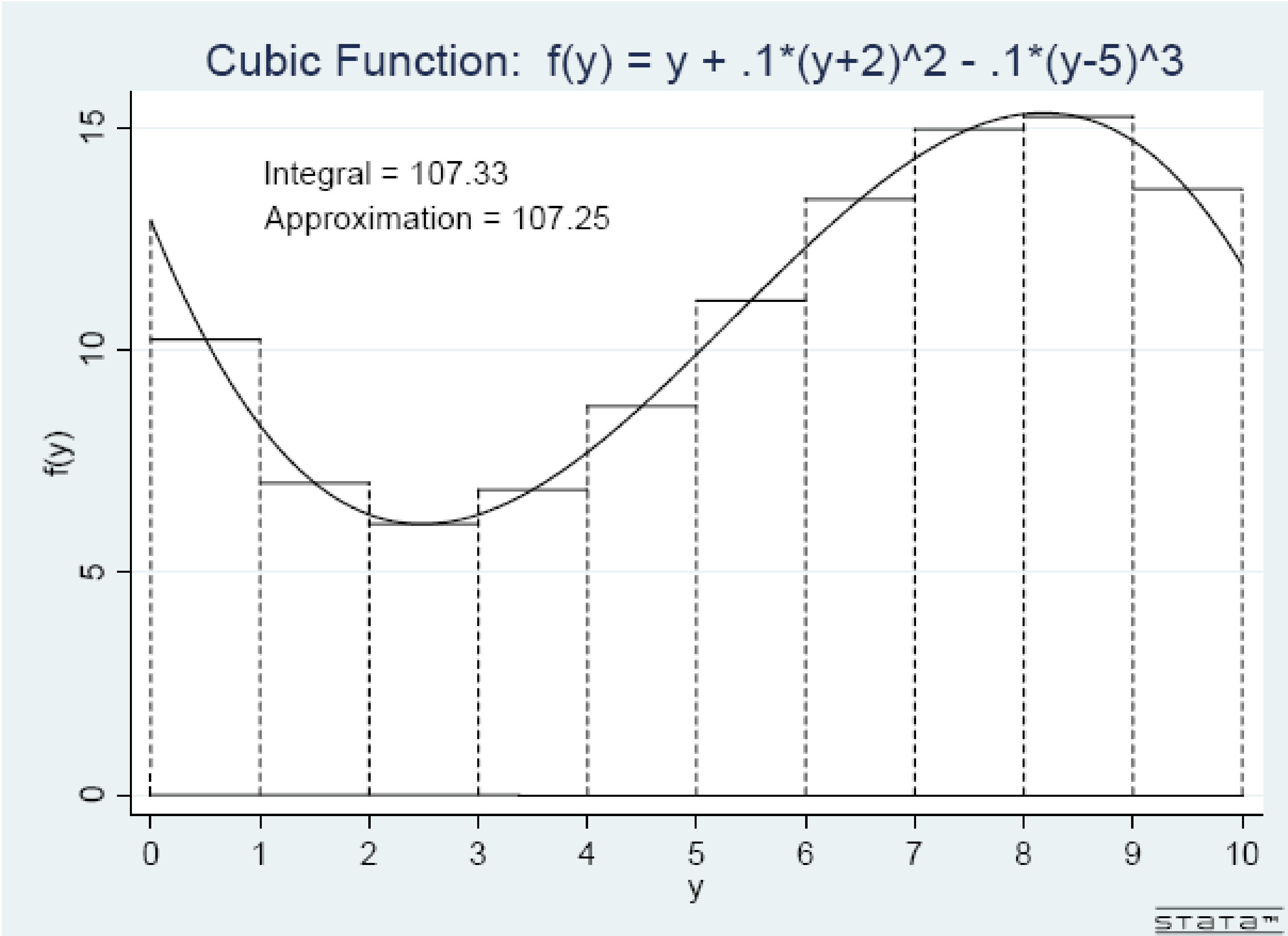
Motivation is similar to beginning calculus

How to calculate the area under a smooth curve?

Compute integral, *or*

1. **Divide curve into thin vertical slices**
2. **For each slice compute: width×height**
3. **Add over slices to get area**

More slices (each thinner) yields more precision



Cubic Function: f(y) = y + .1*(y+2)^2 - .1*(y-5)^3
Integral = 107.33
Approximation = 107.25
f(y)
0
5
10
15
0
1
2
3
4
5
6
7
8
9
10
y

CDE Motivation (2)

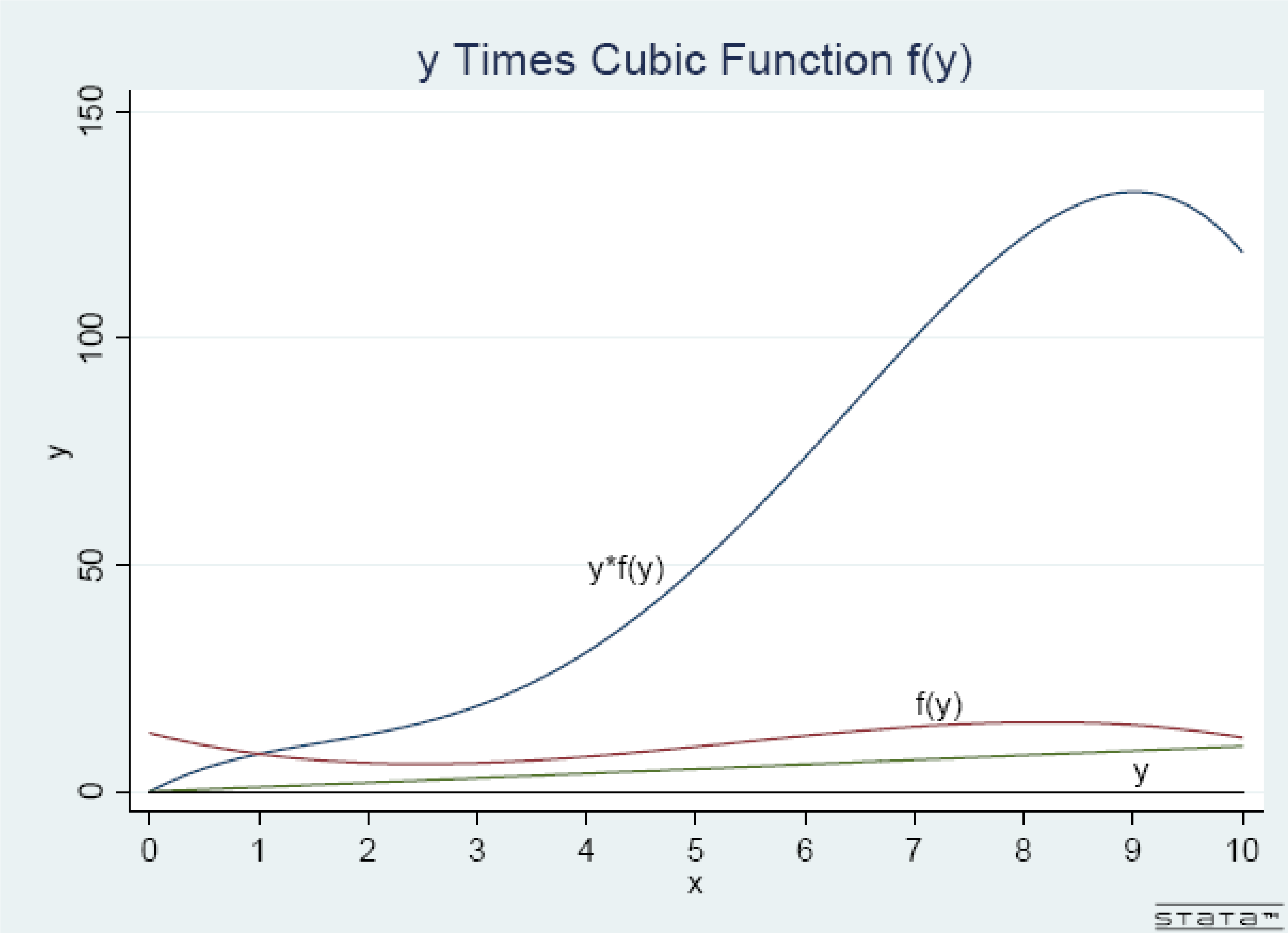
Calculus metaphor is not exact, however

Curve in question is density of expenditures *y* (histogram of *y*)

Area under density sums to 1, by definition

Want E(*y*), which is integral of *y* ×density

$$E(y) = \int_0^{\infty} y f(y)\,dy$$



y Times Cubic Function f(y)
150
100
50
0
y
y*f(y)
f(y)
y
0
1
2
3
4
5
6
7
8
9
10
x
stata™

CDE Motivation (3)

With expenditure data, there are two main estimation issues

1. **Probability of being in the bin (slice)**
   - **Histogram of *y***

2. **Mean expenditures, conditional on being in bin**
   - **Mean of *y* for *y* in a narrow range**

Advantages of CDE

Not tied to specific parametric specification

- Can estimate multi-peaked distribution
  - Data with treatment heterogeneity
  - Work hours (peaks at 20, 40 hours)
  - Ordered and count models

Incorporates zeroes easily

- One of the bins is for the zeroes

Flexible parametric specification

- Controls for covariates

Key Assumptions

Assumptions (see G&M equations 12 and 13)

1. **Probability of being in a bin depends on covariates**
2. **Mean value of *y* conditional on bin is independent of covariates**

Implications

1. **Marginal (incremental) effects operate only through bin changes**
2. **Marginal (incremental) effects do not affect the mean value of *y* conditional on the bin**

Simple Monte Carlo Example (1)

Bimodal health expenditures
All positive expenditures

Two types of patients: High expense and Low expense

Examples

- **C-section vs. normal vaginal delivery**
- **Branded vs. generic drug**
- **Inpatient vs. outpatient treatment**

High type: one-third of sample, $y \cong$ \$5,000
Low type: two-thirds of sample, $y \cong$ \$1,000

Simple Monte Carlo Example (2)

Single continuous covariate affects Pr(High type)

Examples

Age, income, co-payment, health status, ADLs

Empirical set-up

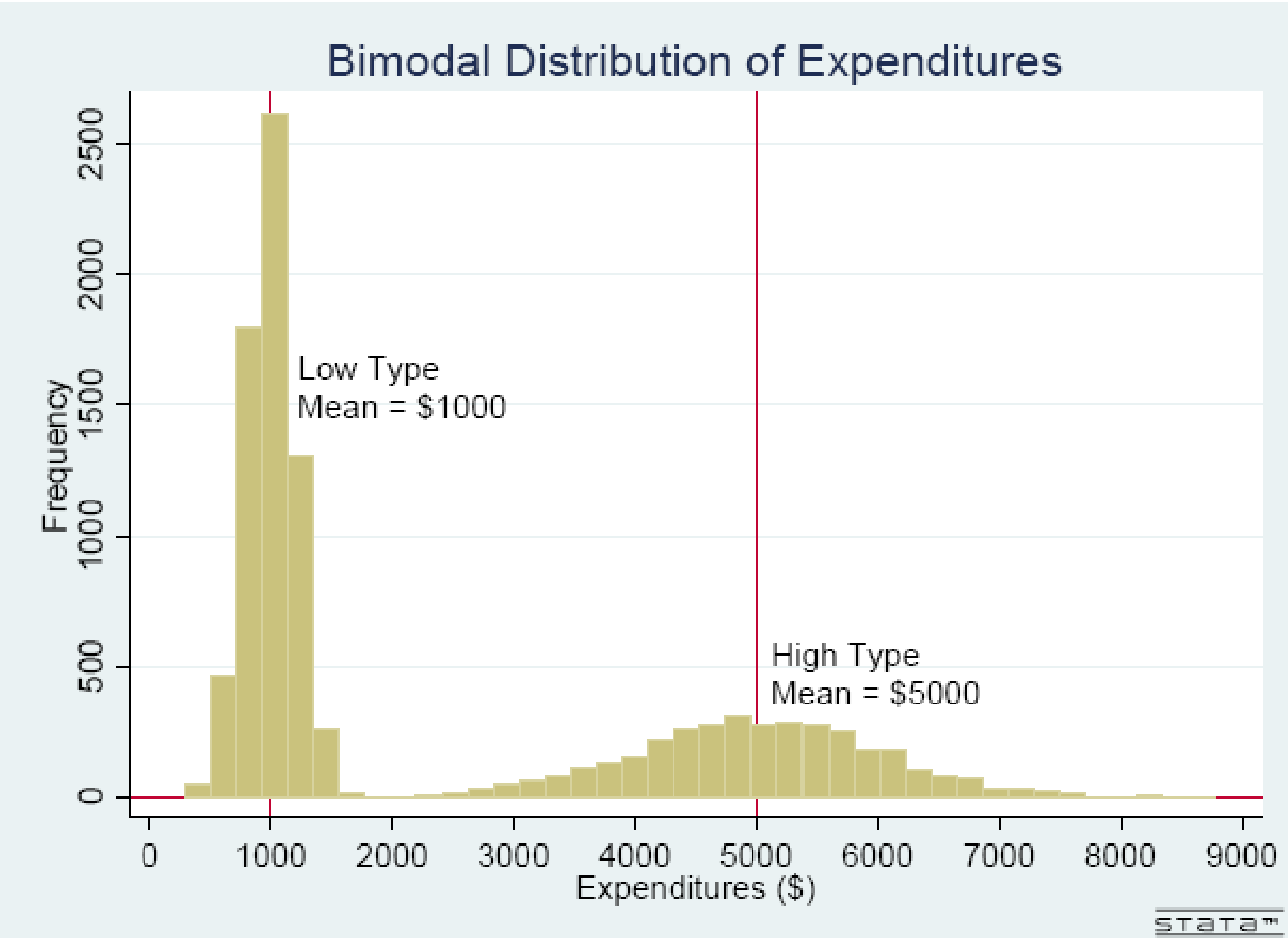
2 bins

Pr(High type) depends only on age

E(*y* |High type) = $5,000

E(*y* |Low type) = $1,000

Want to know E(*y* | *age*), marginal effect of *age*



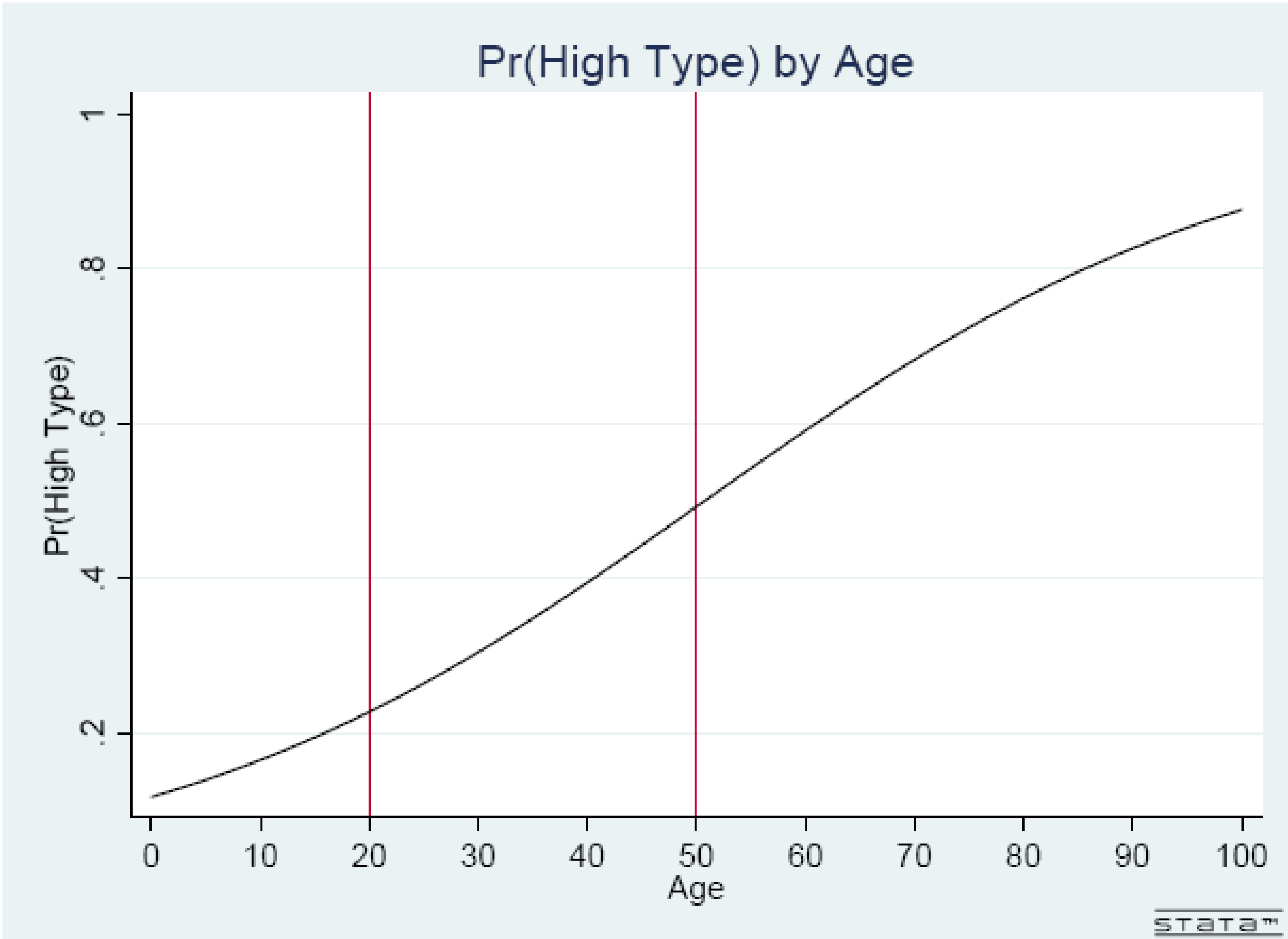
Bimodal Distribution of Expenditures
Frequency
0
500
1000
1500
2000
2500
Low Type
Mean = $1000
High Type
Mean = $5000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Expenditures ($)

Logit: Pr(High type) = f(Age) (1)

```
Logistic regression          Number of obs   =      10000
                             LR chi2(1)      =     261.66
                             Prob > chi2     =     0.0000
Log(l) = -6331.1696          Pseudo R2       =     0.0202

-----------------------------------------------------
      y_type |      Coef.   Std. Err.       z    P>|z|
-------------+---------------------------------------
         age |   .0397664   .0024918    15.96   0.000
       _cons |  -2.017876   .0907367   -22.24   0.000
-----------------------------------------------------
```

Logit: Pr(High type) = f(Age) (2)



Mean Expenditures Conditional on Age

$$E\left(y \mid age\right) = \Pr\left(LowType \mid age\right) \times E\left(y \mid LowType\right) + \Pr\left(HighType \mid age\right) \times E\left(y \mid HighType\right)$$

$$E\left(y \mid age\right) = \left(\frac{e^{-\left(-2.02+0.04age\right)}}{1+e^{-\left(-2.02+0.04age\right)}}\right) \times 1000 + \frac{1}{1+e^{-\left(-2.02+0.04age\right)}} \times 5000$$

Age	Pr(HighType)	E(y|age)
20	.227	$1,910
30	.305	$2,219
40	.395	$2,579
50	.493	$2,970

Marginal Effect of Age on Expenditures

$$\frac{dE(y \mid age)}{dage} = \frac{d\Pr(LowType \mid age)}{dage} \times E(y \mid LowType)$$

$$+ \frac{d\Pr(HighType \mid age)}{dage} \times E(y \mid HighType)$$

$$= \beta_{age} \hat{P}\left(1 - \hat{P}\right) \times 4000$$

Age	Pr(HighType)	*d*E(*y*|*age*)/*dage*
20	.227	$28
30	.305	$34
40	.395	$38
50	.493	$40

MEPS Example

Compare to results from other models

As before …

- **Use MEPS 2004 data**
- ***y* = Total annual health care expenditures**
- **Use many covariates**
- **Estimate conditional mean of *y***
- **Estimate marginal and interaction effects of two variables**

Discuss choices and estimation issues

CDE Recipe (1)

1. Choose number of bins

One for zeroes plus 10 more

2. Choose boundary values

Equal number in each positive bin

3. Estimate mean of *y* conditional on being in bin

Simple mean

4. Choose model specification for Pr(*y* in bin)

Same *X* as before

5. Estimate series of conditional logit (or probit) models

In practice, `expand` data, then estimate one big model

CDE Recipe (2)

6. Compute mean of *y* conditional on *X*

7. Compute marginal (incremental) effects

8. Bootstrap the standard errors

9. Test assumption

Is the mean of *y* constant within bin?

10. Revisit choices about bins and boundaries

Consider more (or fewer) bins

Simple MEPS Example

Sample size is N = 19,386

One bin for the zeroes

17% have $y = 0$ (N = 3,440)

10 additional bins for the positives

Equal number in each positive bin

$N \cong$ 1,595

Boundaries are not equally spaced

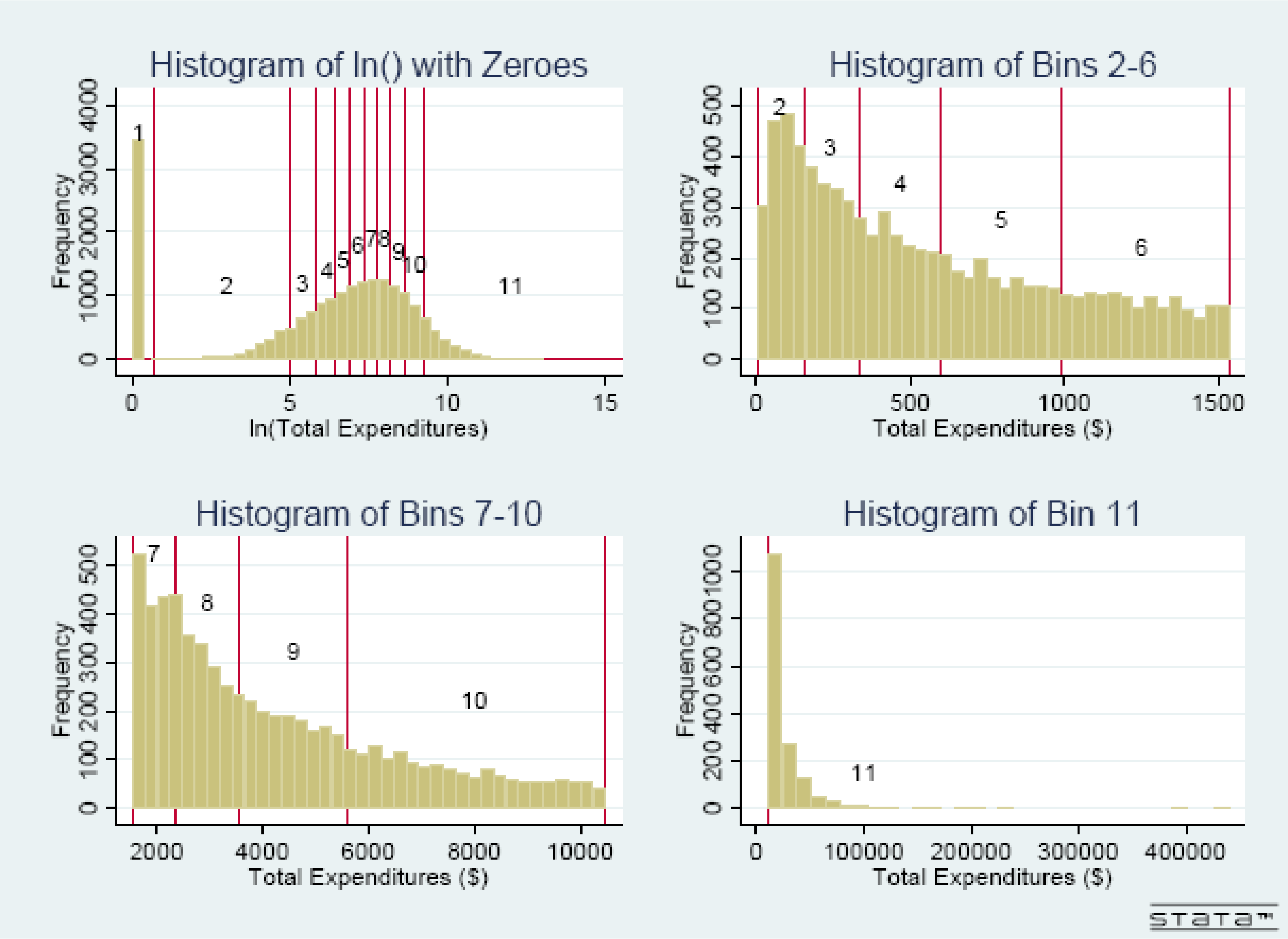
Bins and Boundaries in Stata (1)

```
egen y_k = cut($y) if $y>0, group(10)
replace y_k = y_k + 2
replace y_k = 1 if y_k == .

preserve
   gen y_min = $y
   gen y_mean = $y
   sort y_k
   collapse (min) y_min (mean) y_mean
        y_k_num, by(y_k)
   format y_min y_mean %8.0fc
   list y_k y_min y_mean
restore
```

Bins and Boundaries in Stata (2)

	y_k	y_min	y_mean	y_k_num
1.	1	0	0	3440
2.	2	2	82	1590
3.	3	153	239	1594
4.	4	338	460	1590
5.	5	600	781	1604
6.	6	989	1,247	1594
7.	7	1,537	1,931	1595
8.	8	2,365	2,896	1595
9.	9	3,559	4,507	1594
10.	10	5,621	7,625	1595
11.	11	10,476	25,025	1595



Histogram of ln() with Zeroes
Frequency
ln(Total Expenditures)
Histogram of Bins 2-6
Frequency
Total Expenditures ($)
Histogram of Bins 7-10
Frequency
Total Expenditures ($)
Histogram of Bin 11
Frequency
Total Expenditures ($)
stata™

Series of Logit Models

Estimate series of logit models

Discrete hazard models

Pr(*y* in bin 1)

Pr(*y* in bin 2 | *y* not in bin 1)

Pr(*y* in bin 3 | *y* not in bin 1 or 2)

:

Pr(*y* in bin 10 | *y* not in bin 1–9)

(No need to estimate Pr(*y* in bin 11 | *y* not in bins 1–10))

One Big Logit Model

Series of logit models potentially has enormous number of parameters

Estimate one big logit model instead

First trick is to `expand` the data

- ***k* observations for person with *y* in bin *k***
- **(e.g., 1 observation if *y* = 0, 11 observations if *y* = $100,000)**
- **In Stata: `expand y_k`**

Dependent variable equals 0 for all but highest bin

- **(e.g., if *y* is in bin 3, dependent variables equal 0, 0, 1)**

Alpha_*k*

Second trick is to create covariate α_k based on bin number (In simple models, including α_k and polynomials of α_k gets exactly same results as a series of logit models)

If there are an equal number in each bin, then

$$\alpha_k = -\ln\left(K - k\right), \qquad k < K$$

More generally, want

$$\alpha_k = \ln\left(\frac{p_k}{1 - p_k}\right) \qquad p_k = \left(\frac{\text{number in bin } k}{\text{number in bins } k - K}\right)$$

Stata Code (1)

Continue within previous `preserve` & `restore`

```
   sort y_k
   gen y_k_sum = sum(y_k_num)
   gen p_k = y_k_num/(y_k_sum[_N] - y_k_sum +
        y_k_num)
   gen alpha_k = ln(p_k/(1 - p_k))
   list
   gen y_number = y_k
   keep y_number alpha_k
   sort y_number
   save "D:alpha_k.dta", replace
restore
```

Stata Code (2)

```
expand y_k

sort dupersidx
by dupersidx: gen y_hazard = (_n == y_k)
by dupersidx: gen y_number = _n
sort y_number

merge y_number using alpha_k.dta

logit y_hazard alpha_k $rhs, robust nolog
```

Compute Conditional Mean of *y*

Same idea as in simple Monte Carlo

- **More parts to formula**
- **Put pieces together**

Sum up over all bins, the probability of being in the bin times the value of *y* conditional on being in the bin

If estimate one big logit

- **Predicted probabilities are conditional on α_k**
- **Remember that probability of being in bin 5 is conditional on not being in bins 1–4**

CDE Results for MEPS

	Men		Women	
	E[*y*]	M.E.	E[*y*]	M.E.
Age = 20	3,704	101	4,034	30
Age = 30	4,804	120	4,338	31
Age = 40	6,087	138	4,657	32
Age = 50	7,528	151	4,990	34
Age = 60	9,084	160	5,337	35
Age = 70	10,703	164	5,698	37

Marginal effects are of age

Test Hypothesis that E(y|X) is Constant in Bin

For subset of data in each bin, do joint test of whether all Coefficients are zero in model to predict y

No point in testing zero bin (y = 0 for all by construction)

In MEPS example, do 10 F-tests

For each bin regress $y = X\beta + \varepsilon$, test $\beta = 0$

MEPS Results

Does not fail in bins 3 through 6

Fails badly in bin 2 and bins 7 through 11

If fail test, consider adding more bins

Extensions

Gilleskie & Mroz (2004)

Propose a test for whether to divide each interval into *R* sub-intervals
Test based on log-likelihood

Explain that key assumption may not hold in highest interval
May need to keep *y* a function of the covariates

Can apply CDE to ordered and count models (Mroz 2008)

Two Other Issues

When reading the article and talking to Gilleskie and Mroz, two modeling issues arise that are not particular to CDE

1. **Use a flexible specification of the covariates**
   Use polynomials in the *X*'s, interactions
   Can do this for any model, not just CDE

2. **Model the unobserved heterogeneity**
   As written, the G&M paper treats the zeroes like a 2PM
   One could also add discrete factors to model unobserved heterogeneity
   Can do this for 2PM, not just CDE

Relation of CDE to Other Models

Treatment of zeroes

- **Can be like two-part model (separate model)**
- **Or in one big logit (as in the MEPS example)**
- **The correlation between the bins is not modeled**

Positive part

- **More flexible specification**
- **Can handle multi-peaked density**
- **Related to multi-part models, except for covariates in bin**

Ordered and count models

- **Translates to count models easily**
- **See Mroz (WP 2008); Liu, Dow, Norton (*JHE* 2004)**

Conclusions

We think CDE is an exciting new model, has much potential

More work to be done

- **Not trivial to implement in Stata (no canned package, yet)**
- **Optimal algorithm for choosing bins and boundaries**
- **Robustness of top bin**
- **Not clear how to control for endogeneity**
- **Monte Carlo simulations**
  - **To compare CDE to other models**
  - **To evaluate in terms of overfitting**
  - **To understand how optimal bins relate to sample size**
  - **To understand importance of key assumption**

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Overview

Alternatives for studies with zero mass and skewed outcomes

OLS on $\ln(y+c),\quad c>0$

Box-Cox regression

GLM with zeros

Tobit

Heckman Selection-like (Adjusted/Generalized Tobit) models

Two-part models

The Zeros Problem

What to do when

Substantial fraction of observations at zero

Examples

Inpatient utilization or expenditures
Nursing home utilization or expenditures
Prescription drug utilization or expenditures
Cigarette smoking

The Zeros Problem

Problems with standard OLS model

- **OLS may predict negative values, nonsensical for health**
- **Zero mass may respond differently to covariates**
  - **Access to treatment vs. extent of treatment**

Issues are more prominent when the data has more zeros

Alternative Estimators for Models with Zeros

OLS on $\ln(y + c)$, $c > 0$

Box-Cox regression

GLM with zeros

Tobit

Heckman Selection-like (Adjusted/Generalized Tobit) models

Two-part models

OLS on $\ln(y+c)$

Problem: cannot take logarithm of y when y=0

$\ln(0) = -\infty$

Solution: add positive constant c

Advantages

- **Easy**
- **Log addresses skewness, constant deals with ln(0)**

Disadvantages

- **Zero mass may respond differently to covariates**
- **Many set c=1 arbitrarily**
- **Value of c matters, need grid search for optimum**
- **Retransformation problem aggravated at low end**
- **Poorly behaved in Duan et al. (JBES, 1983)**

Box-Cox Regression

$$\tau(y;\lambda,c) = \frac{(y+c)^{\lambda}-1}{\lambda} \quad \text{if} \quad \lambda \neq 0$$

$$= \ln(y+c) \quad \text{otherwise}$$

Advantages

Transformation addresses skewness

Disadvantages

Regression parameters typically not easy to interpret
Marginal effects not easy to compute
Box-Cox parameters not always easy to identify

GLM with Zeros

Gamma regression with log link (Mullahy, 1998)

Why does it work?

In OLS models with log(y)

$$E(\ln(y_i)|x_i) = x_i'\beta$$

In GLM

$$\ln E(y_i|x_i) = x_i'\beta$$

Advantages

Easy to estimate

Disadvantages

may be numerically unstable if there are lots of zeros

Tobit Model

Classic Tobit

$$y_i^* = x_i'\beta + u$$

$$y_i = 0 \quad if\ y_i^* < 0$$

$$y_i = y_i^* \quad if\ y_i^* \geq 0$$

Stata command: `tobit`
Can also use `intreg`: more general but slower

Tobit Model

Disadvantages

Only works well if dependent variable is censored Normal

Places severe restrictions on parameters, error term

Very sensitive to minor departures from homoskedasticity and normality (Hurd, 1979)

Typically want to think of 0's as actual zeros, not indicator of a censored value.

Zeros do not have a missing data interpretation

Rarely recommended for health economics

Tobit for Right Censored Data

May be a more useful application in health economics

Example: estimating costs for episode of care, but some episodes not complete.

$$y_i^* = x_i'\beta + u$$

$$y_i = y_i^* \text{ if } y_i^* < \bar{y}_i$$

$$y_i = \bar{y}_i \text{ if } y_i^* \geq \bar{y}_i$$

In this case, the parameters of the latent equation are exactly what we want to estimate.

Heckman Selection-like Model (Adjusted Tobit)

Stata command: `heckman`

$$y_i^* = x_i'\beta + u$$
$$d_i^* = z_i'\alpha + v$$
$$y_i = 0 \quad if \; d_i^* < 0$$
$$y_i = y_i^* \; if \; d_i^* \geq 0$$

Allows for error correlation, parameters for each of 2 parts

More flexible than the Tobit
Less sensitive to distributional assumptions
It is possible to construct Heckman Selection-like models with right censoring

Two-Part Model

Takes advantage of basic rule of probability

$$\mathrm{E}(y|x) = \Pr\left(y>0\right) \times \mathrm{E}\left(y|y>0\right)$$

Splits consumption into two parts

1. **Pr(any use or expenditures)**
   Full sample
   Use logit or probit regression

   Stata commands

```
probit y x
logit y x
```

Two-Part Model

2. **Level of use or expenditures**
   Conditional on $y > 0$ (subsample with y>0)
   Use appropriate continuous or count model
   (e.g., OLS, ln(y), GLM, truncated count)

   Stata commands:

```
regress y    x if y>0
regress lny  x if y>0
glm     y    x if y>0, f(gamma) l(log)
ztnb    y    x if y>0
```

Heckman versus the Two-Part Model

In scale of estimation:

In Heckman:

$$E(y \mid y > 0) = X\beta + \rho\sigma\frac{\phi(Z\alpha)}{\Phi(Z\alpha)}$$

In TPM:

$$E(y \mid y > 0) = X\gamma$$

Makes it look like Heckman is more general

But TPM does NOT assume away correlation

They are not nested

Heckman versus the Two-Part Model

Without exclusion restrictions it is very difficult to identify ρ

Exclusion restrictions are extremely hard to find

Without exclusion restrictions, TPM fits better in many empirical applications

With exclusion restrictions, TPM is often not much worse

There is also the issue of interpretation

Are the zeros really zero or are they censored?

Do we care about the latent process or actual outcomes?

Predictions in Two-Part Model

Predictions depend on both parts of the model

$$\mathrm{E}(y|x) = \Pr\left(y>0\right) \times \mathrm{E}\left(y|y>0\right)$$

Example

First part probit $\Pr(y) = \Phi(x\alpha)$

Second part ln(*y*)

If log-scale error term is Normal and homoskedastic

$$\hat{y} = \Phi\left(x\hat{\alpha}\right) \times \exp\left(x\hat{\beta} + .5\hat{\sigma}^2\right)$$

Marginal Effects in Two-Part Model

Continuous variable x_c

$$\frac{\partial \mathrm{E}(y)}{\partial x_c} = \frac{\partial\left(\Pr(y>0)\times \mathrm{E}(y \mid y>0)\right)}{\partial x_c}$$

$$= \left(\Pr(y>0)\ \frac{\partial \mathrm{E}(y \mid y>0)}{\partial x_c}\right) + \left(\mathrm{E}(y \mid y>0)\ \frac{\partial \Pr(y>0)}{\partial x_c}\right)$$

Dummy variable x_d

$$\mathrm{E}(y \mid x_d = 1) - \mathrm{E}(y \mid x_d = 0)$$

$$= \left(\Pr(y>0 \mid x_d = 1) - \Pr(y>0 \mid x_d = 0)\right) \times \mathrm{E}(y \mid y>0 \mid x_d = 1)$$

$$+ \Pr(y>0 \mid x_d = 0) \times \left(\mathrm{E}(y \mid y>0 \mid x_d = 1) - \mathrm{E}(y \mid y>0 \mid x_d = 0)\right)$$

Example: Ln(Total Expenditures)

OLS: ln(total expenditures + 1)

TPM: Probit + OLS on ln(total expenditures>0)

Overall marginal effects (log scale)

Variable	OLS	Tobit	Heckman	TPM
lninc	0.244**	0.291**	0.259**	0.261**
	(0.026)	(0.031)	(0.027)	(0.028)
unins	-2.144**	-2.705**	-2.195**	-2.418**
	(0.059)	(0.079)	(0.076)	(0.092)

Example: Ln(Total Expenditures)

Marginal effects for P(y>0)

Variable	Tobit	Heckman	TPM
lninc	0.005**	0.026**	0.026**
	(0.001)	(0.003)	(0.003)
unins	-0.114**	-0.235**	-0.235**
	(0.007)	(0.012)	(0.012)

Marginal effects for y | y>0 (log scale)

Variable	Tobit	Heckman	TPM
lninc	0.260**	0.088**	0.087**
	(0.028)	(0.015)	(0.015)
unins	-2.126**	-0.873**	-0.856**
	(0.059)	(0.040)	(0.038)

Example: Ln(Total Expenditures)

GLM – Gamma family with log link

TPM – Probit + Gamma GLM for total expenditures >0

Marginal effects (raw scale)

Variable	GLM single equation	TPM probit	TPM GLM	TPM-overall
ln(income)	179.41**	0.026**	120.09*	176.53**
	(46.58)	(0.003)	(52.58)	(53.05)
uninsured	-1,444.06**	-0.235**	-1,394.35**	-1866.34**
	(109.25)	(0.012)	(112.32)	(128.78)

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Examples and Characteristics

Examples

- Number of visits to the doctor
- Number of Rx prescriptions filled
- Number of cigarettes smoked per day

Like expenditures / costs

- many zeros
- very skewed in non-zero range
- intrinsically heteroskedastic (variance increases with mean)

Differences

- integer valued
- concentrated on a few low values (0, 1, 2)
- prediction of event probabilities often of interest

Overview

Studies with count data

Poisson (canonical model)

Estimation

Prediction – Mean, Events

Interpretation – Marginal effects, Incremental Effects

Goodness of fit

Negative Binomial

Zero Inflated Models

Hurdle Models (Two Part Models for Counts)

Model Selection - Discriminating among nonnested models

Count Model Extensions

Extensions for "excess" zeros

- **Zero Inflated Models**
- **Hurdle Models (Two Part Models for Counts)**

Model Selection - Discriminating among nonnested models

- **In Sample**
- **Cross validation (out of sample)**

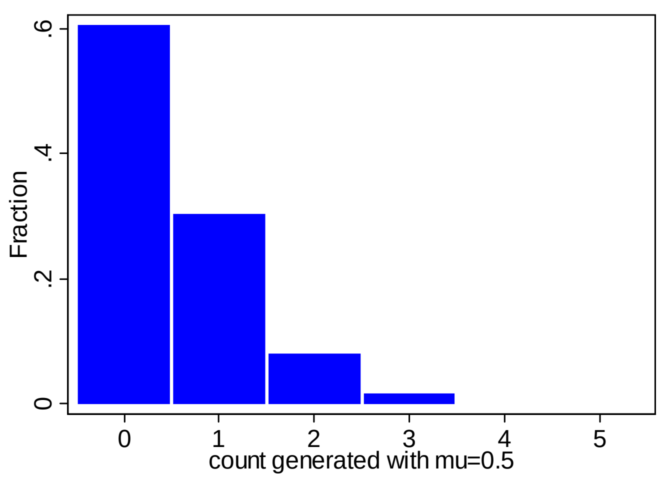
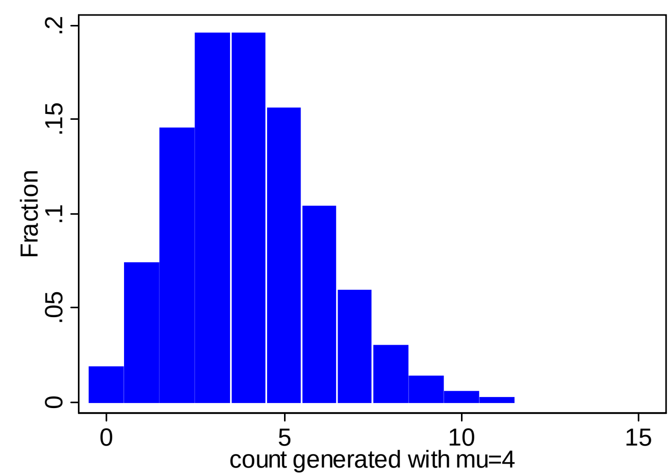
Poisson

Mean $\mu = \exp(X\beta)$

Variance $\sigma^2 = \exp(X\beta)$

Density

$$\Pr(Y = y | X) = \frac{\exp(-\mu)\,\mu^y}{\Gamma(y+1)}$$

Poisson with μ = 0.5	Poisson with μ = 4
	

FYI: Estimation

First Order Condition for MLE

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} (y_i - \mu_i) X_i = 0$$

$$MLE(\widehat{\beta}) = QMLE(\widehat{\beta})$$

Sandwich form for $Cov(\hat{\beta})$ is correct for Quasi MLE

Stata command: `poisson y x, robust`

First Order Condition for NLS Estimator

$$\sum_{i=1}^{N} (y_i - \mu_i) \mu_i = 0$$

$$MLE(\widehat{\beta}) \neq NLS(\widehat{\beta})$$

Prediction

The typical prediction of interest is the conditional mean.

But, in nonlinear models, predictions of quantities other than the conditional mean are often of interest.

In the context of count data, we might be interested in predictions of the distribution of the count variable

$\Pr(Y = 0 \mid X)$

$\Pr(Y = 12 \mid X)$

We might also be interested in predictions of certain events of interest

$\Pr(Y > 5 \mid X) = 1 - \Pr(Y \leq 5 \mid X)$

Substantively

Probability of exceeding a benefit cap (mental health)

Probability of a “drive-through delivery

Prediction

Conditional Mean: $\hat{\mu} = \exp(X\hat{\beta})$

Stata command: `predict yhat`

Distribution and events:

$$\Pr(Y = y|X) = \frac{\exp(-\hat{\mu})\,\hat{\mu}^y}{\Gamma(y+1)} \qquad \forall\, y = 0, 1, 2, 3, \ldots$$

Stata package: `prcounts`

Interpretation

Marginal Effects - for continuous variables

$$\frac{\partial E(y_i|X)}{\partial X^k} = \beta^k \ \mu_i$$

Examples: Income, Price, Health status

Incremental Effects - for binary variables

$$E(y_i|X, X^k=1) - E(y_i|X, X^k=0)$$
$$= \left[\mu_i | X^k = 0 \right] \left[\exp(\beta^k) - 1 \right]$$

Examples: Treatment/ Control, Insurance, Gender, Race

Predictions, Marginal, Incremental Effects

Approach depends on research question

1. Evaluate for hypothetical individuals

- **a. Mean (or Median, other quantiles) of X in sample**
- **b. Mean (or Median, other quantiles) of X in sub-sample of interest**
- **c. Hypothetical individual of interest**

Stata command: `mfx`

Predictions, Marginal, Incremental Effects

2.Evaluate for each individual

- **a. Average over sample**
- **b. Average over sub-samples of interest**

3. For Incremental Effects – (Treatment vs. Control)

- **a. Switch all individuals from control to treatment**
- **b. Switch controls to treatment**

Stata package: `margeff`

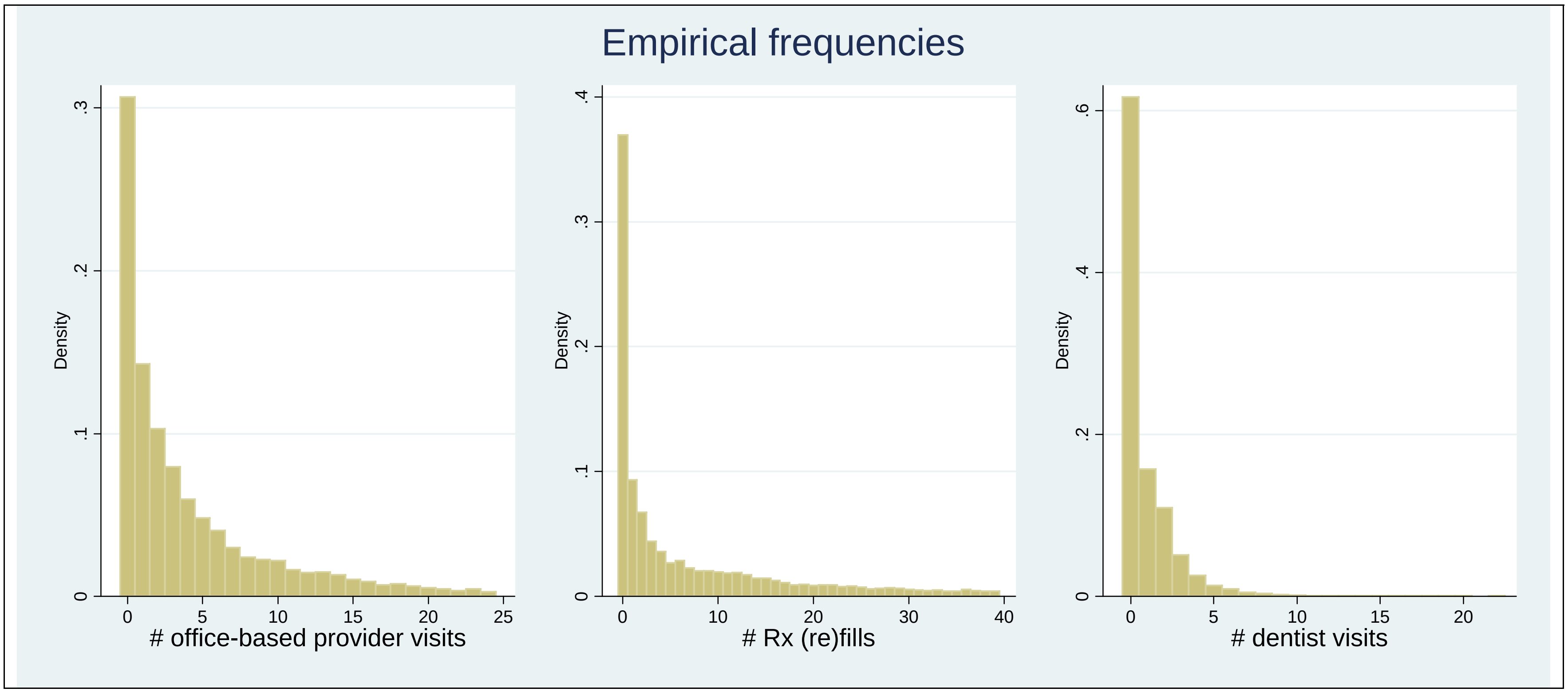
Else calculate "by hand"

How one does it can make a big difference

Examples

Data from MEPS

1. Number of office-based visits
2. Number of prescriptions, filled and refilled
3. Number of emergency room visits



Marginal Effects from Poisson

At mean of X

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.385**	-0.171	0.118**
	(0.070)	(0.097)	(0.015)
Uninsured	-2.404**	-4.871**	-0.435**
	(0.151)	(0.221)	(0.027)

At median of X

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.341**	-0.177	0.104**
	(0.065)	(0.099)	(0.016)
Uninsured	-1.883**	-4.384**	-0.336**
	(0.151)	(0.248)	(0.030)

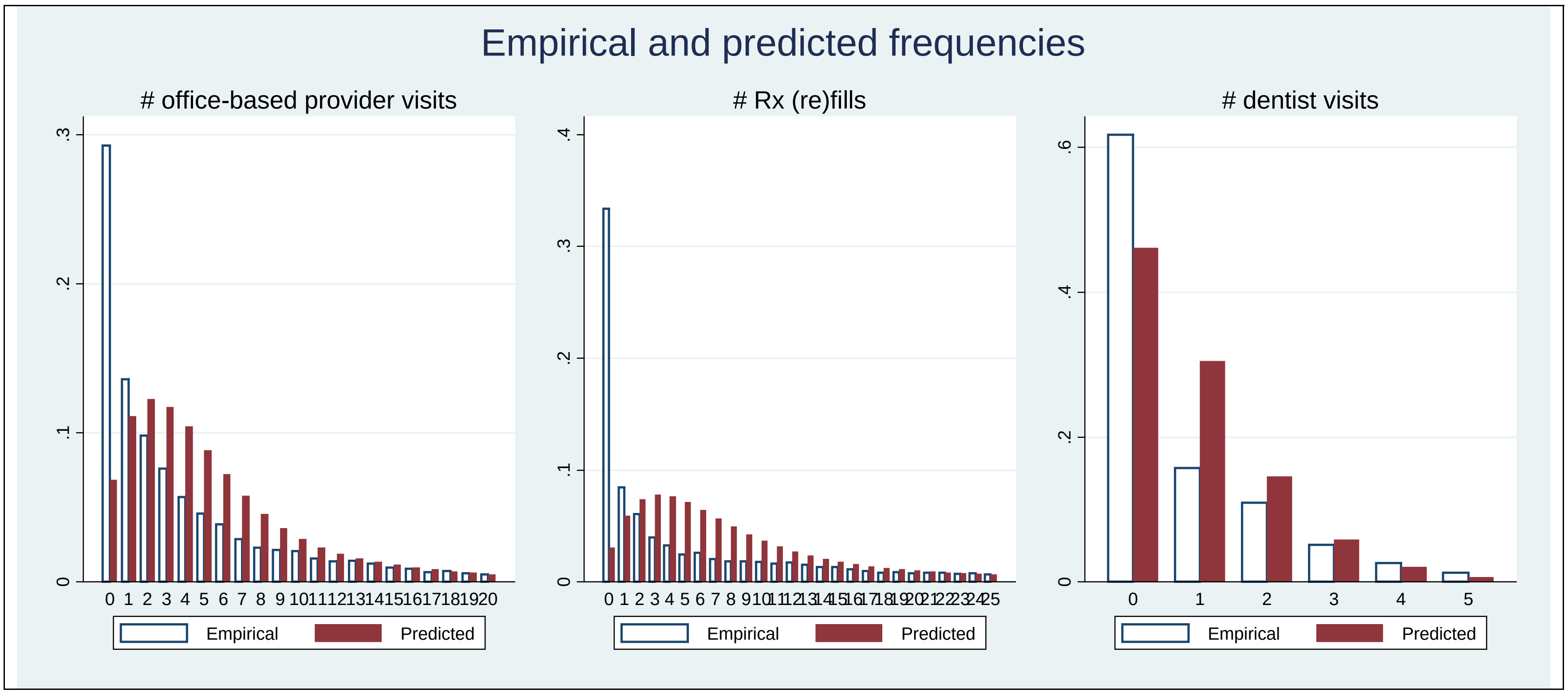
Average of effects

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.524**	-0.280	0.147**
	(0.096)	(0.159)	(0. 019)
Uninsured	-3.072**	-7.396**	-0.506**
	(0.184)	(0.315)	(0.030)

In-sample Goodness of fit

Informal / Graphical - compare empirical distribution of *y* to predicted distribution

Stata command: `prcounts`



In-sample Goodness of Fit

Mean Prediction (of distribution) Error

$$MPE = \frac{1}{J}\sum_{j=0}^{J}(f_j - \hat{f}_j)$$

Mean Square Prediction (of distribution) Error

$$MSPE = \frac{1}{J}\sum_{j=0}^{J}(f_j - \hat{f}_j)^2$$

J should be chosen to cover most (but not all the values taken by the count variable)

	Coverage	MPE	MSPE
Office visits	0-20	0.158	28.602
Rx (re) fills	0-25	0.101	39.761
Dentist visits	0-5	0.374	79.733

Poisson - Summary

Advantages

Robust (asymptotic) to misspecification of variance

Easy to compute marginal effects and predictions

Disadvantages

Not robust in finite samples

Possibly sensitive to influential observations and outliers

Not efficient if variance is misspecified

Overview

Studies with count data

Poisson (canonical model)

Negative Binomial

Estimation
Prediction – Mean, Events
Interpretation – Marginal effects, Incremental Effects
Goodness of fit

Zero Inflated Models

Hurdle Models (Two Part Models for Counts)

Model Selection - Discriminating among nonnested models

Negative Binomial

Canonical model for overdispersed data

Mean $\mu = \exp(X\beta)$

Overdispersion – variance exceeds the mean

$$Var(y|X) = \mu + \alpha\, g(\mu) > \mu$$

Negative Binomial-1 $Var(y|X) = \mu + \alpha\,\mu$

Negative Binomial-2 $Var(y|X) = \mu + \alpha\,\mu^2$

FYI: Estimation

Maximum Likelihood

Stata command: `nbreg`

GLM (based on first 2 moments)

Stata command: `glm y x, family(gamma)link(log)`

Choosing between NB-1 and NB-2

These are non-nested models

Use model selection criteria – discussed later under model comparisons

Prediction and Interpretation

Prediction

1. **Conditional Mean:** $\hat{\mu} = \exp(X\hat{\beta})$
   Stata command: `predict`

2. **Distribution:** **NB density**
   Stata package: `prcounts`

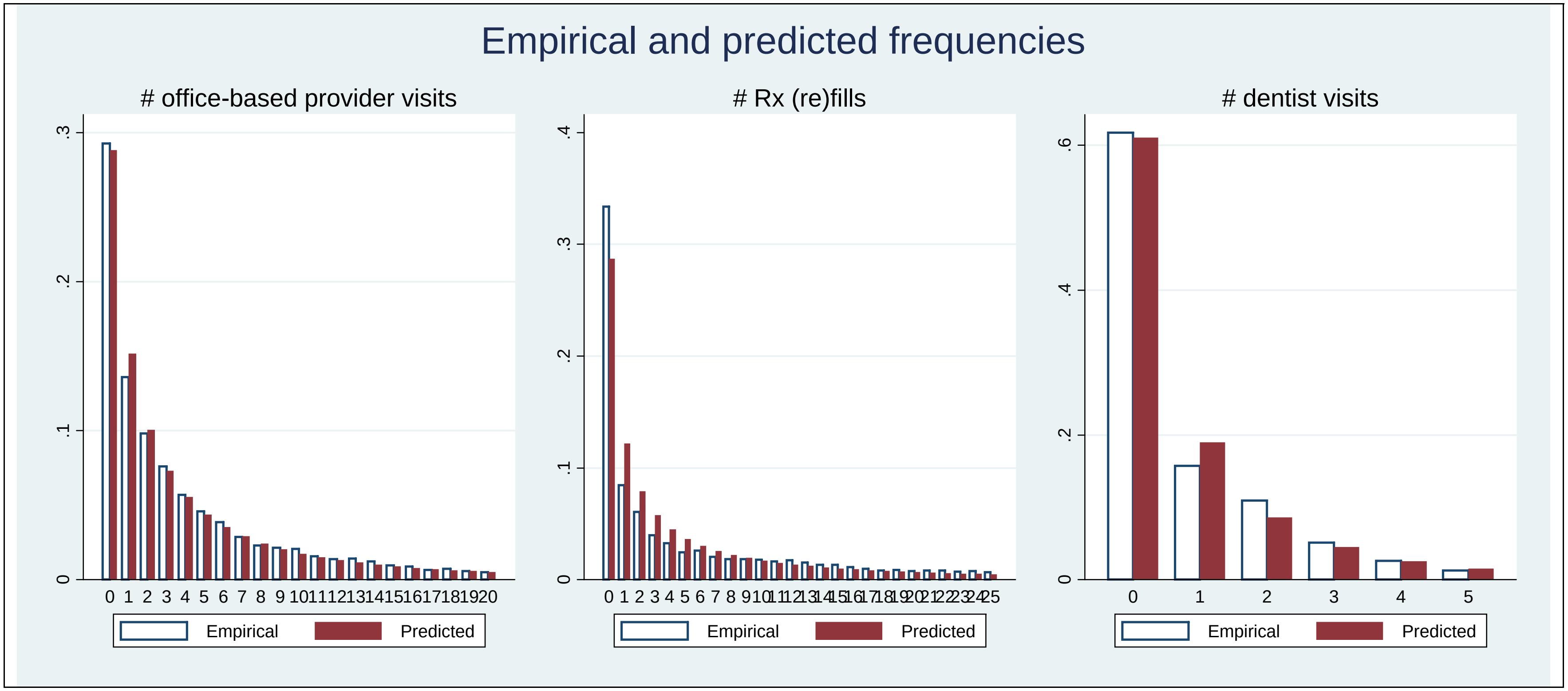
3. **Events**
   Stata package: `prcounts`

Interpretation

Marginal / Incremental Effects

Stata command: `mfx`
Stata package: `margeff`

In-sample Goodness of Fit



	Coverage	MPE	MSPE
Office visits	0-20	-0.043	0.155
Rx (re) fills	0-25	0.122	1.798
Dentist visits	0-5	-0.054	2.832

Negative Binomial - Summary

Advantages

Much less sensitive to influential observations and outliers

Mean is robust in finite samples

Disadvantages

Distribution is not robust to misspecification of variance

Not efficient if variance is misspecified

Overview

Studies with count data

Poisson (canonical model)

Negative Binomial

Zero Inflated Models

- **Estimation**
- **Prediction – Mean, Events**
- **Interpretation – Marginal effects, Incremental Effects**
- **Goodness of fit**

Hurdle Models (Two Part Models for Counts)

- **Estimation**
- **Prediction – Mean, Events**
- **Interpretation – Marginal effects, Incremental Effects**
- **Goodness of fit**

Model Selection - Discriminating among nonnested models

Zero Inflated Models

Zeros are from two processes

No demand / No demand in sample period

$$\Pr(Y = 0 \mid X) = f_1(0 \mid \theta_1) + (1 - f_1(0 \mid \theta_1))\, f_2(0 \mid \theta_2)$$

$$\Pr(Y = y > 0 \mid X) = (1 - f_1(0 \mid \theta_1))\, f_2(y \mid \theta_2)$$

where

$f_1(. \mid \theta_1)$ **is a Logit / Probit Model**

$f_2(. \mid \theta_2)$ **is a Poisson / NB Model**

Stata command: `zip / zinb`

Usually, same covariates used in $f_1(. \mid \theta_1)$ **and** $f_2(. \mid \theta_2)$

Sometimes $f_1(. \mid \theta_1)$ **specified as a constant**

Prediction and Interpretation

Prediction

1. **Conditional Mean**
   Stata command: `predict`

2. **Distribution**
   Stata package: `prcounts`

3. **Events:**
   Stata package: `prcounts`

Interpretation

Marginal / Incremental Effects

Stata command: `mfx`

Examples

Marginal effects for P(y>0)

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.010**	0.018**	0.061**
	(0.003)	(0.005)	(0.010)
Uninsured	-0.173**	-0.253**	-0.341**
	(0.030)	(0.017)	(0.026)

Overall marginal effects

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.330**	-0.028	0.150**
	(0.061)	(0.119)	(0.022)
Uninsured	-1.704**	-4.389**	-0.491**
	(0.159)	(0.275)	(0.048)

Zero-Inflated Models

Advantages

Natural way to introduce extra zeros

Disadvantages (Especially if both parts have same covariates)

Computationally complex – likelihood function can have plateaus and multiple maxima

Weak identification of Binary and Count Model parameters in finite samples

Even when marginal / incremental effects from each processes are "sensible", overall effects may be "odd" (composition effects)

Hurdle Model

Two Part Model for counts - Zeros are from different process
No demand / No demand in sample period

$$\Pr(Y = 0 \mid X) = f_1(0 \mid \theta_1)$$

$$\Pr(Y = y > 0 \mid X) = \frac{(1 - f_1(0 \mid \theta_1))}{(1 - f_2(0 \mid \theta_2))} f_2(y \mid \theta_2)$$

where

$f_1(. \mid \theta_1)$ **is a Logit / Probit Model**

$f_2(. \mid \theta_2)$ **is a Poisson / NB Model**

$\frac{1}{(1 - f_2(0 \mid \theta_2))} f_2(y \mid \theta_2)$ **is a Truncated Count Density**

Stata command:

```
logit(probit) / ztp(ztnb)
```

Prediction and Interpretation

Because model is constructed "manually", `predict` does not work directly

1. **Conditional Mean**
   Stata command: `sample program`

2. **Distribution**
   Stata command: `sample program`

3. **Events**
   Stata command: `sample program`

4. **Marginal / Incremental Effects**
   `mfx` does not work directly
   Stata command: `sample program`

Example

Marginal effects for P(y>0)

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.027**	0.015**	0.053**
	(0.004)	(0.005)	(0.006)
Uninsured	-0.273**	-0.261**	-0.195**
	(0.011)	(0.011)	(0.010)

Overall marginal effects

Variable	Office visits	Rx (re)fills	Dentist visits
ln(income)	0.350**	-0.040	0.090**
	(0.057)	(0.130)	(0.017)
Uninsured	-2.035**	-6.519**	-0.261**
	(0.124)	(0.368)	(0.056)

Hurdle Models

Advantages

Estimation in 2 parts

Same variables in both parts not a problem

Numerically well behaved

Disadvantages

Strong prior belief that zeros from different process than positives

Even when marginal / incremental effects from each processes are "sensible", overall effects may be "odd" (composition effects)

Overview

Studies with count data

- **Poisson (canonical model)**
- **Negative Binomial**
- **Zero Inflated Models**
- **Hurdle Models (Two Part Models for Counts)**

Model Selection - Discriminating among nonnested models

Model Selection

In Sample

Akaike Information Criterion

$$AIC = -2\log(L) + 2k$$

Bayesian Information Criterion

$$BIC = -2\log(L) + \log(N)k$$

Cross-validation

Estimation – 70% random sample

Prediction – Remaining 30% sample

Then use log(L) to evaluate models

Examples

Office visits

	Poisson	NB-2	ZINB	HNB
K	19	20	39	39
AIC	196661	98102	97297	96850*
BIC	196811	98259	97596	97157*

Rx (re)fills

	Poisson	NB-2	ZINB	HNB
K	19	20	39	39
AIC	324166	115404	112983	112956*
BIC	324315	115562	113290	113263*

Dentist visits

	Poisson	NB-2	ZINB	HNB
K	19	20	39	39
AIC	55373	47262	46396	46310*
BIC	55523	47419	46703	46617*

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Overview

Finite mixture models

- **Motivation**
- **Estimation**
- **Prediction – Mean, Events**
- **Interpretation – Marginal effects, Incremental Effects**

Motivation for Finite Mixture Models

1. Natural representation of heterogeneity in a finite number of latent classes

2. Heterogeneity of effects / responses

   a. for different "classes" of observations
      i. Healthy and sick
      ii. Normal and complicated pregnancies
   b. for different "classes" of outcomes
      i. visits to GPs and visits to specialists
   c. in different parts of the distribution of the outcome
      i. For low and high values of the outcome

Interested in distribution and behavior in different parts of the distribution

Finite Mixture Models

Flexible extension to basic parametric models

The models accommodate

- **Skewness**
- **Excess Kurtosis**
- **Multimodality**
- **+ in count-data contexts**
- **Overdispersion**
- **Excess Zeros**

Semiparametric / nonparametric estimator -- with enough components (Lindsay, 1995)

Experience suggests that usually only few latent classes are needed to approximate density well

Canonical Example

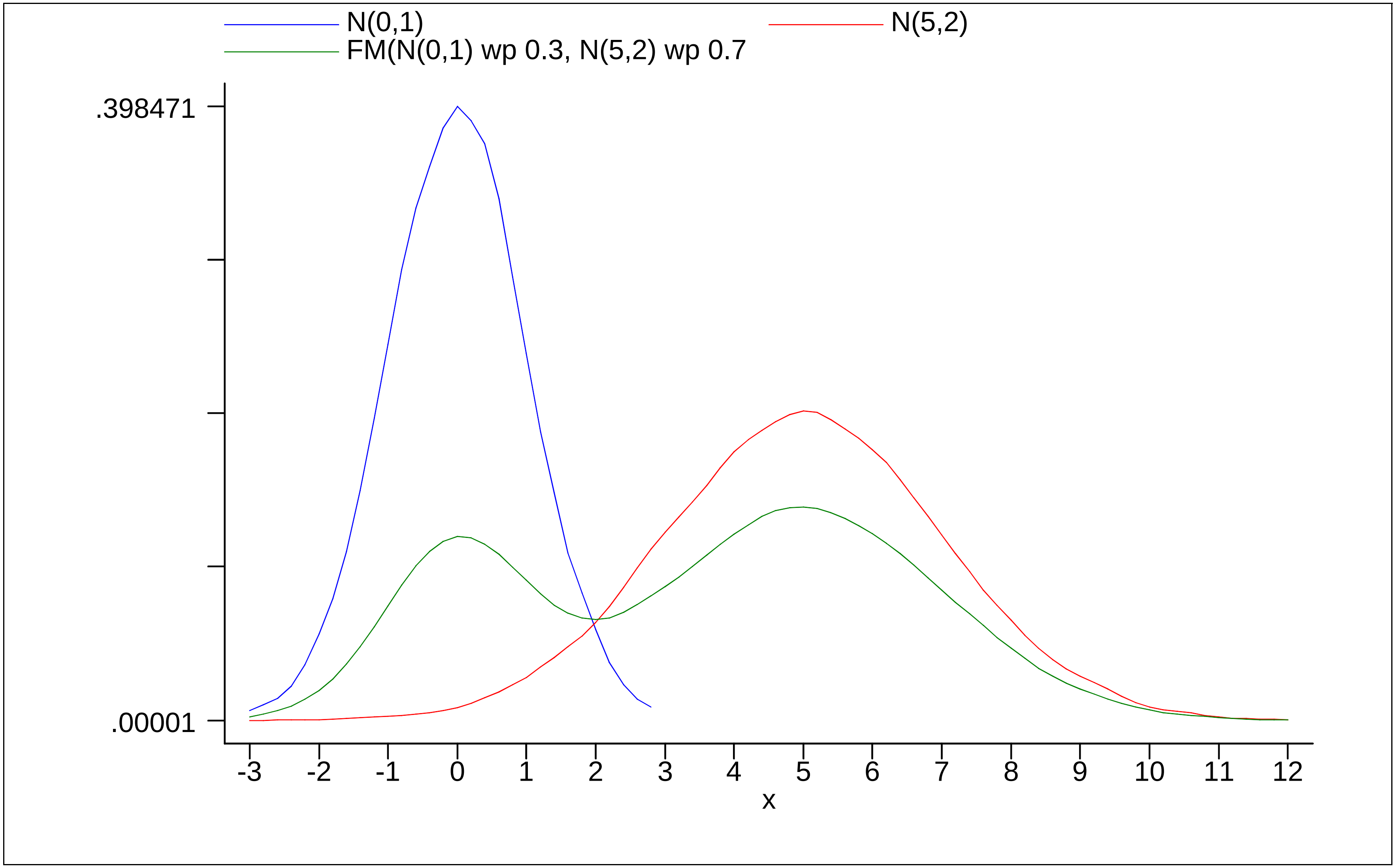
Estimating parameters of the distribution of halibut

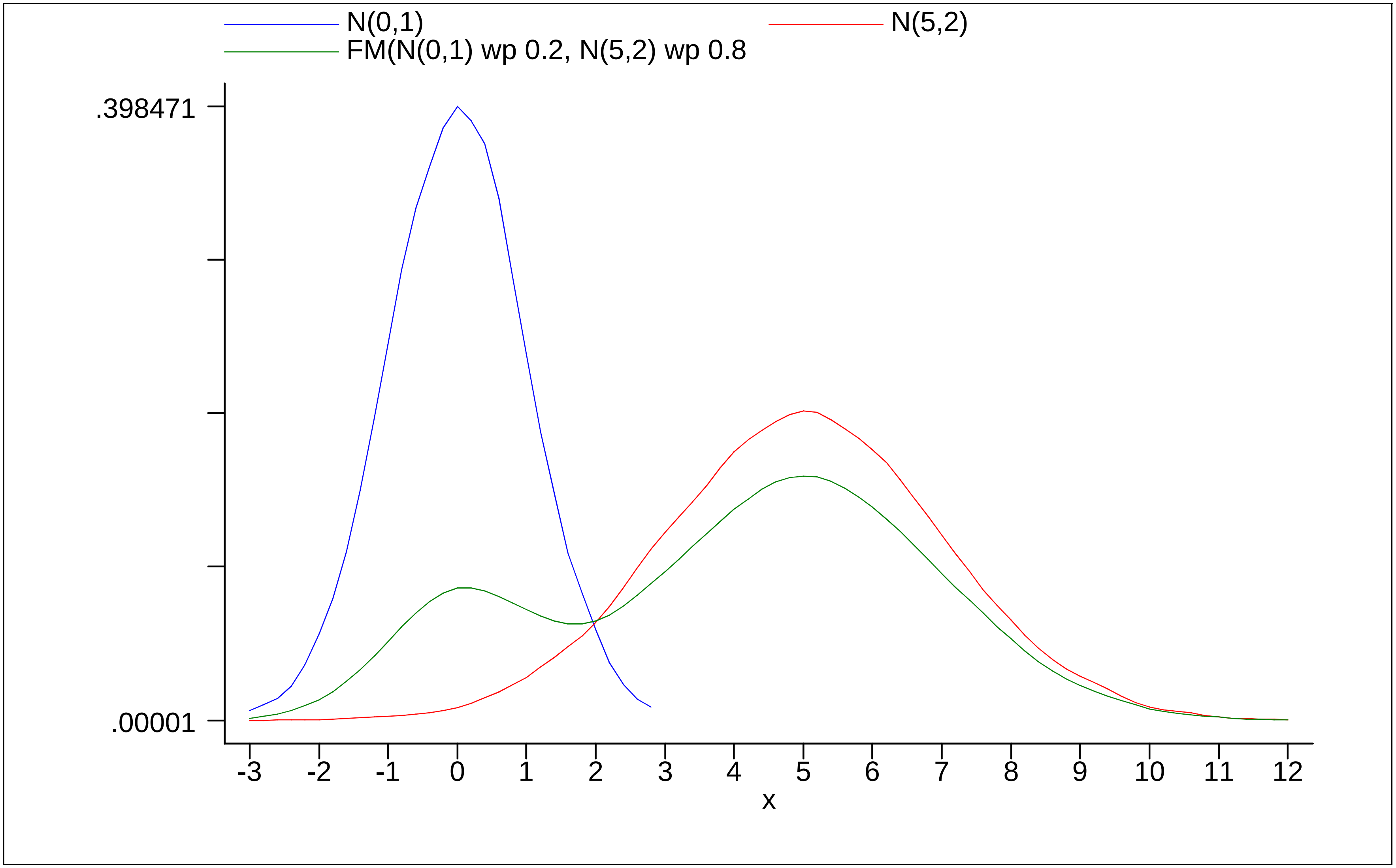
It is known that female halibut is longer than male halibut and that the distribution of lengths is normal

If gender cannot be determined at measurement

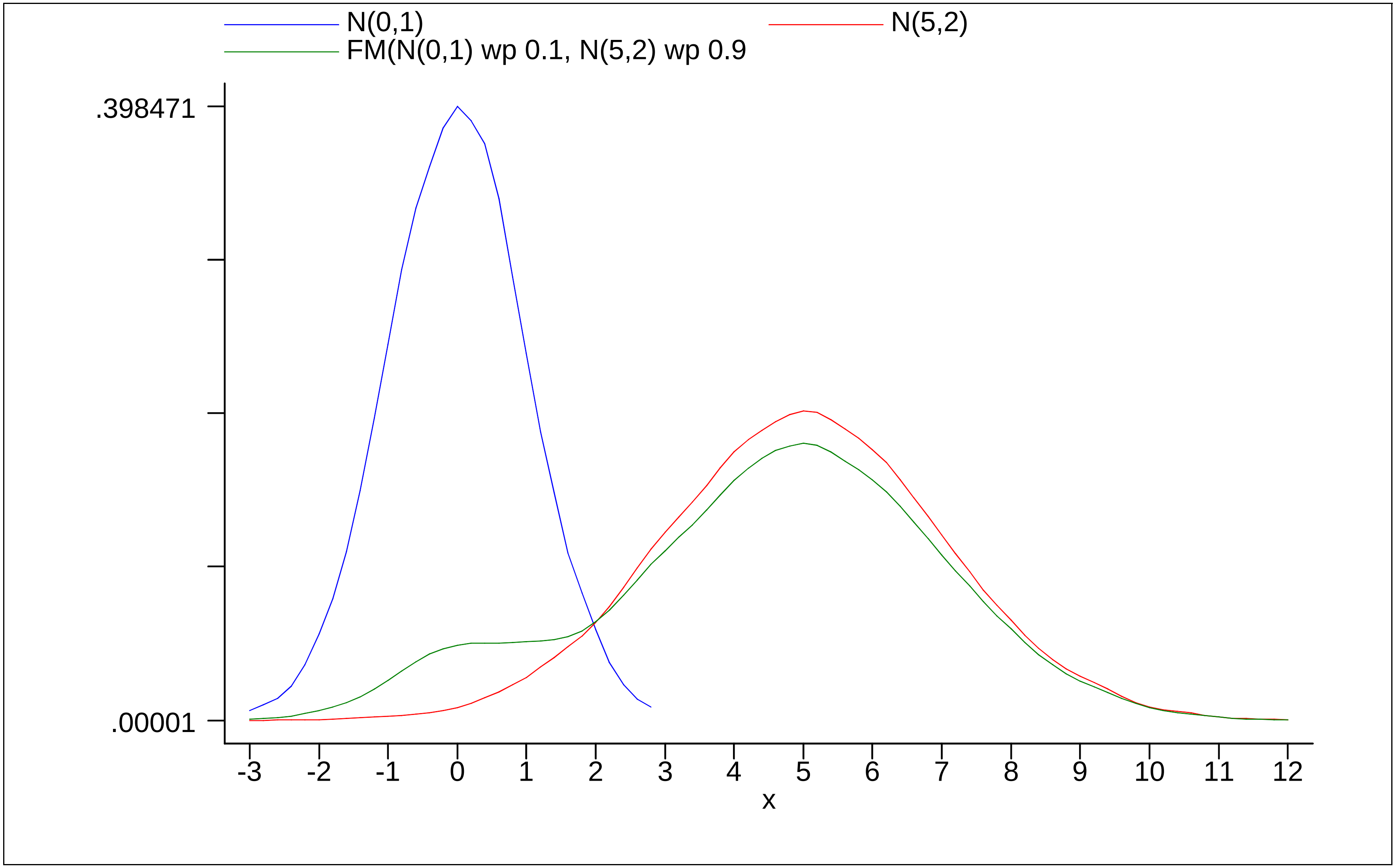
Then distribution is a 2-component finite mixture of normals

A finite mixture allows one to estimate:
1. mean lengths of male and female halibut
2. mixing probability


N(0,1)
N(5,2)
FM(N(0,1) wp 0.3, N(5,2) wp 0.7)
.398471
.00001
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11
12
x



N(0,1)
N(5,2)
FM(N(0,1) wp 0.2, N(5,2) wp 0.8
.398471
.00001
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11
12
x



N(0,1)
N(5,2)
FM(N(0,1) wp 0.1, N(5,2) wp 0.9
.398471
.00001
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11
12
x

Finite Mixture Models

More generally

Finite mixtures

- **Need not be bi (multi) modal**
- **Need not have Gaussian component distributions**
- **Are not restricted to continuous random variables**

Finite Mixture Models

With Known Separation $\delta = (0,1)$

$$\Pr(Y = y|X) = \delta\, f_1(y|\theta_1) + (1-\delta)\, f_2(y|\theta_2)$$

where

> $f_1(.|\theta_1)$ **and** $f_2(.|\theta_2)$ **are component densities (usually assumed to be from the same family)**

Finite Mixture Models

With Unknown Separation $\delta = (0,1)$ is not observed

Assume $\Pr(\delta = 0) = \pi$

$$\Pr(Y = y \mid X) = \pi \;\; f_1(y \mid \theta_1) + (1 - \pi) \;\; f_2(y \mid \theta_2)$$

where

$f_1(. \mid \theta_1)$ and $f_2(. \mid \theta_2)$ are component densities (usually assumed to be from the same family)

π can be a function of X

But this raises issues of identification and is computationally much more complex

Popular mixture component densities

Normal (Gaussian)

Poisson

Gamma

Negative Binomial

Student-t

Weibull

Properties of finite mixtures

The mean of a finite mixture density is

$$E(y_i \mid x_i) = \sum_{c=1}^{C} \pi_c E(y_i \mid x_i; c)$$

Even when π_c, which can be thought of as a prior probability, is a constant the implied posterior probability is not a constant

Applying Bayes Theorem

$$\Pr[y_i \in \text{population } c] = \frac{\pi_c f_c(y_i \mid x_i, \theta_c)}{\pi_1 f_1(y_i \mid x_i, \theta_1) + \pi_2 f_2(y_i \mid x_i, \theta_2)}$$

C = 1, 2

Estimation, Prediction and Interpretation

Estimation

Maximum Likelihood

Stata package: `fmm`

Prediction

1. **Conditional Means**
   Stata command: `predict`

2. **Posterior probabilities**
   Stata command: `predict`

Interpretation

Marginal / Incremental Effects

Stata command: `mfx`

Estimation challenges

1. **The number of components has to be specified - we usually have little theoretical guidance**

2. **Even if prior theory suggests a particular number of components we may not be able to reliably distinguish between some of the components**

3. **In some cases additional components may simply reflect the presence of outliers in the data**

4. **Likelihood function may have multiple local maxima**

Model Selection (picking number of components)

AIC=-2log(L)+2K

BIC=-2log(L)+ log(N)K

where L, K and N are the maximized log likelihood, number of parameters and observations respectively

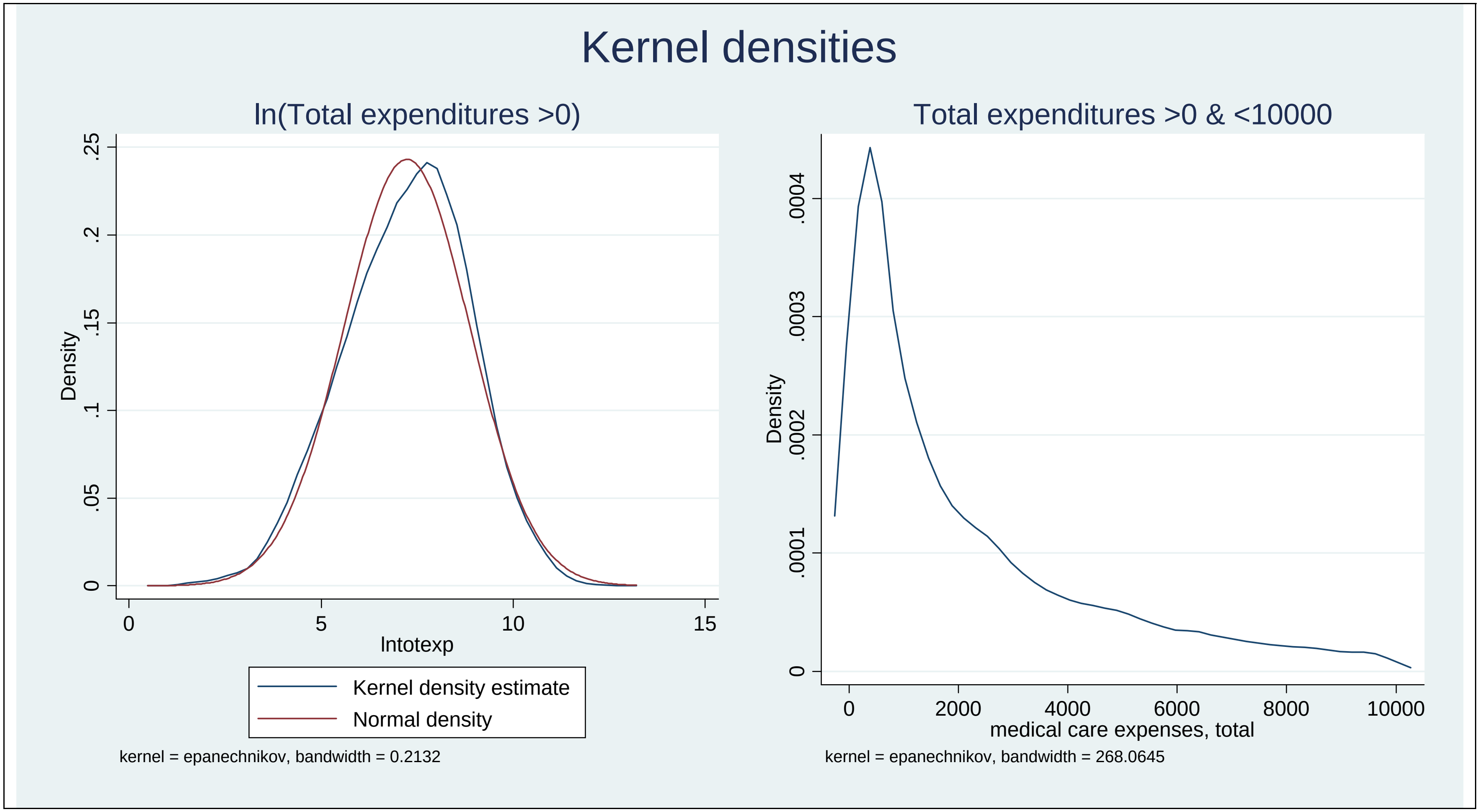
Pick the model with the smallest AIC, BIC

Examples

1. **Estimating distribution of Expenditures – mixture of Gamma densities**

2. **Estimating distribution of Office visits – mixture of Negative Binomial densities**

Distribution of Expenditures

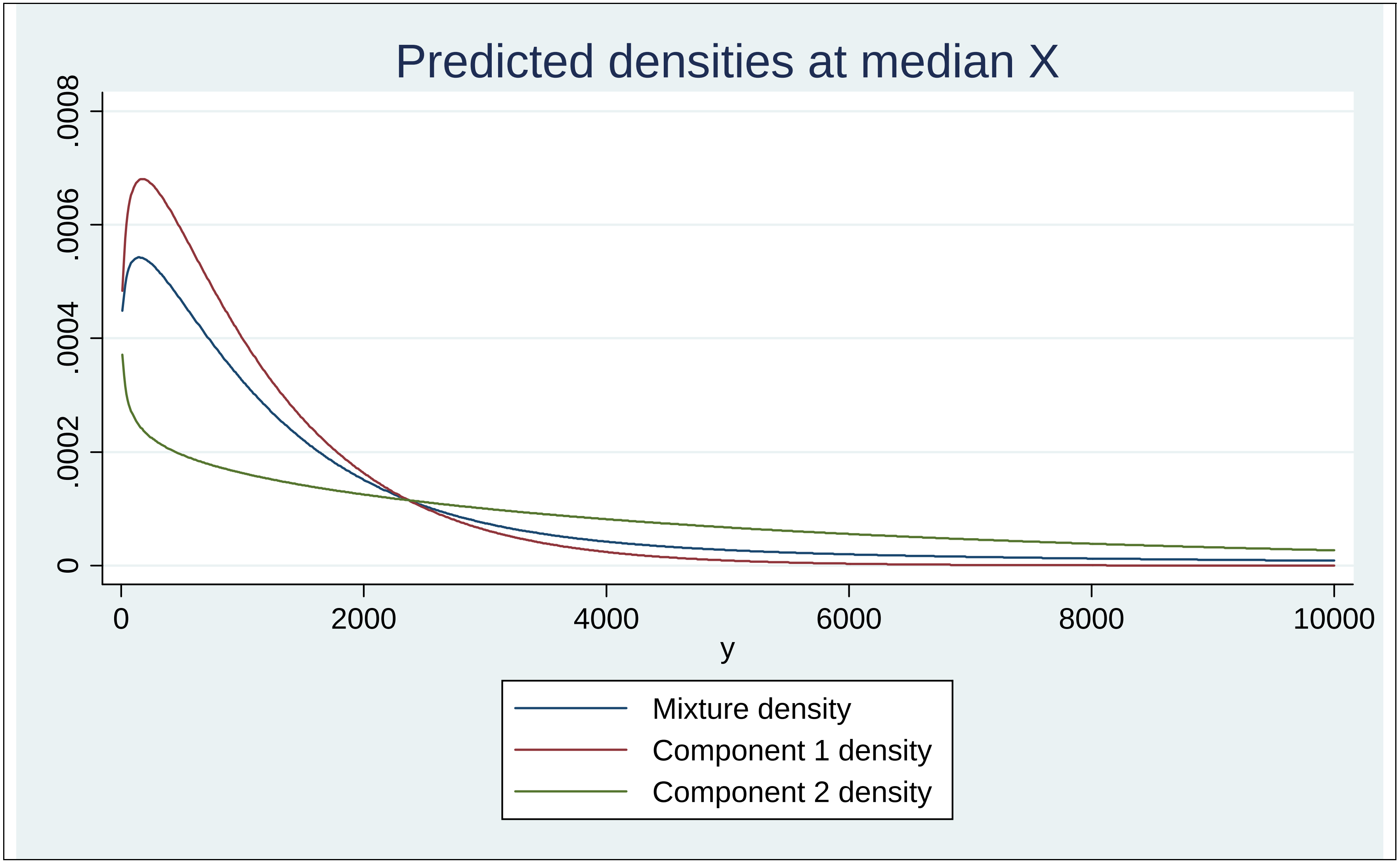


Estimating Expenditures: Marginal Effects

Variable	GLM Gamma	FMM Overall ME	FMM Class 1 ME	FMM Class 2 ME
black	-445.462**	-456.796**	-301.390**	-794.755
	(115.013)	(134.312)	(41.472)	(416.883)
hispanic	-744.610**	-775.710**	-364.473**	-1,670.015**
	(98.538)	(110.537)	(42.988)	(350.614)
other race	-1,060.426**	-1,127.010**	-369.595**	-2,774.138**
	(105.457)	(125.938)	(53.722)	(430.531)
ln(income)	120.091*	140.059*	128.929**	164.263
	(52.582)	(59.739)	(26.319)	(179.935)
uninsured	-1,394.353**	-1,427.735**	-664.879**	-3,086.698**
	(112.325)	(130.955)	(55.895)	(414.964)
class prob			0.685	
			(0.021)	
predicted mean	2686.72	2818.29	1115.37	6521.59

GLM: log link and gamma family
FMM: 2-component mixture of Gamma densities
Test of equality of class-specific coefficients: chi2(19) = 173.00

Estimating Expenditures – Component Gamma Densities

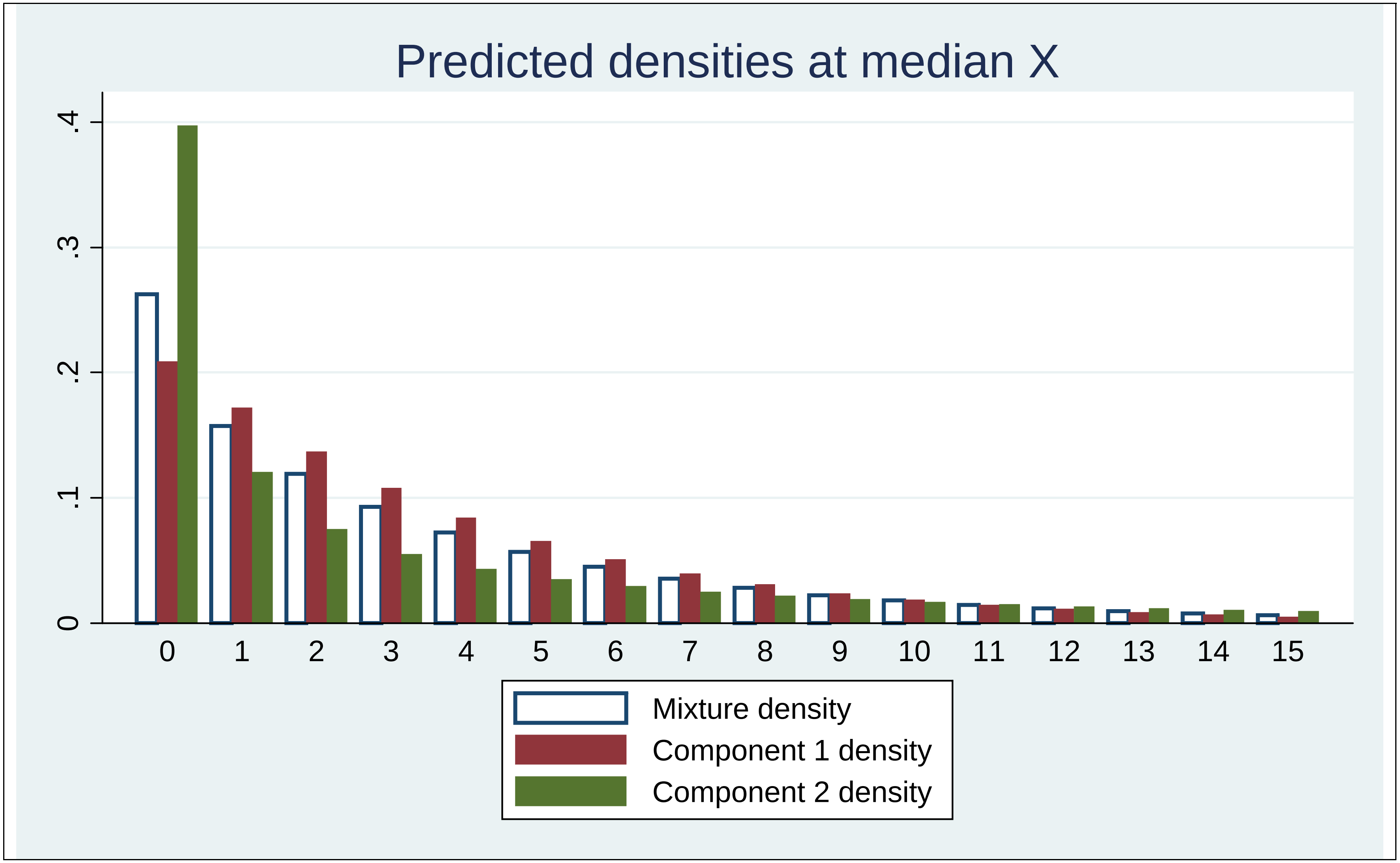


Estimating Office Visits – Marginal Effects

Variable	Negbin1	Overall mfx	Class 1 mfx	Class 2 mfx
black	-1.127**	-1.044**	-1.008**	-1.135**
	(0.090)	(0.095)	(0.111)	(0.363)
hispanic	-0.911**	-0.885**	-0.775**	-1.162**
	(0.093)	(0.090)	(0.121)	(0.357)
other race	-1.009**	-0.941**	-1.026**	-0.728
	(0.120)	(0.111)	(0.145)	(0.380)
ln(income)	0.352**	0.350**	0.195**	0.740**
	(0.048)	(0.052)	(0.065)	(0.206)
uninsured	-2.393**	-2.135**	-2.105**	-2.211**
	(0.098)	(0.096)	(0.132)	(0.374)
class prob.			0.716	
			(0.042)	
predicted mean	4.57	4.07	3.55	5.37

Negbin1: Negative Binomial-1 regression
FMM: 2-component mixture of NB-1 densities
Test of equality of class-specific coefficients: chi2(19) = 247.11

Estimating Expenditures – Component NB-1 Densities



Model Selection -- In Sample

Akaike Information Criterion

$$AIC = -\log(L) + k$$

Bayesian Information Criterion

$$BIC = -\log(L) + k\log(N)$$

Mean Square Prediction (of distribution) Error

$$MSPE = \frac{1}{N}(f - \hat{f})'(f - \hat{f})$$

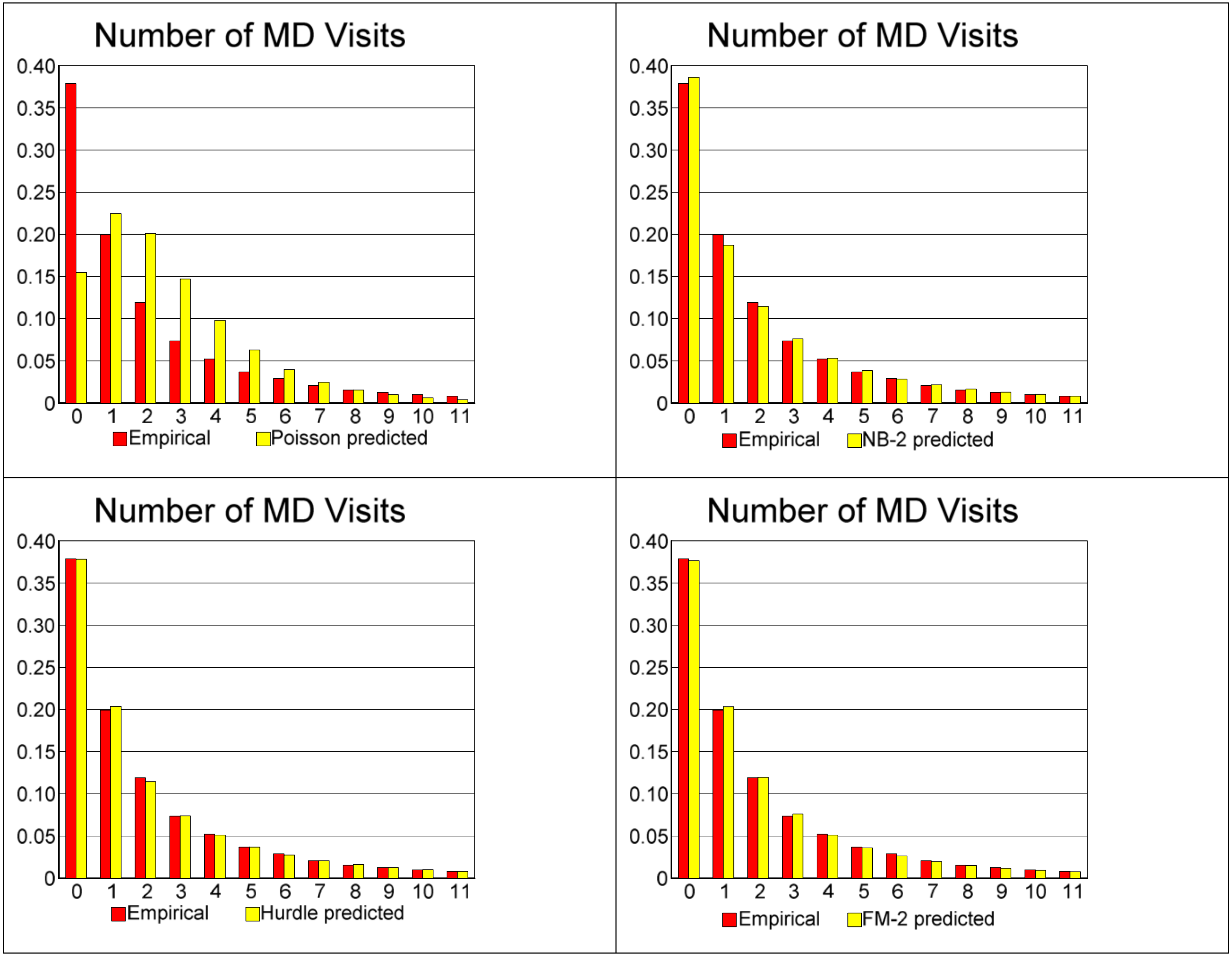
Graphical Check of Distribution

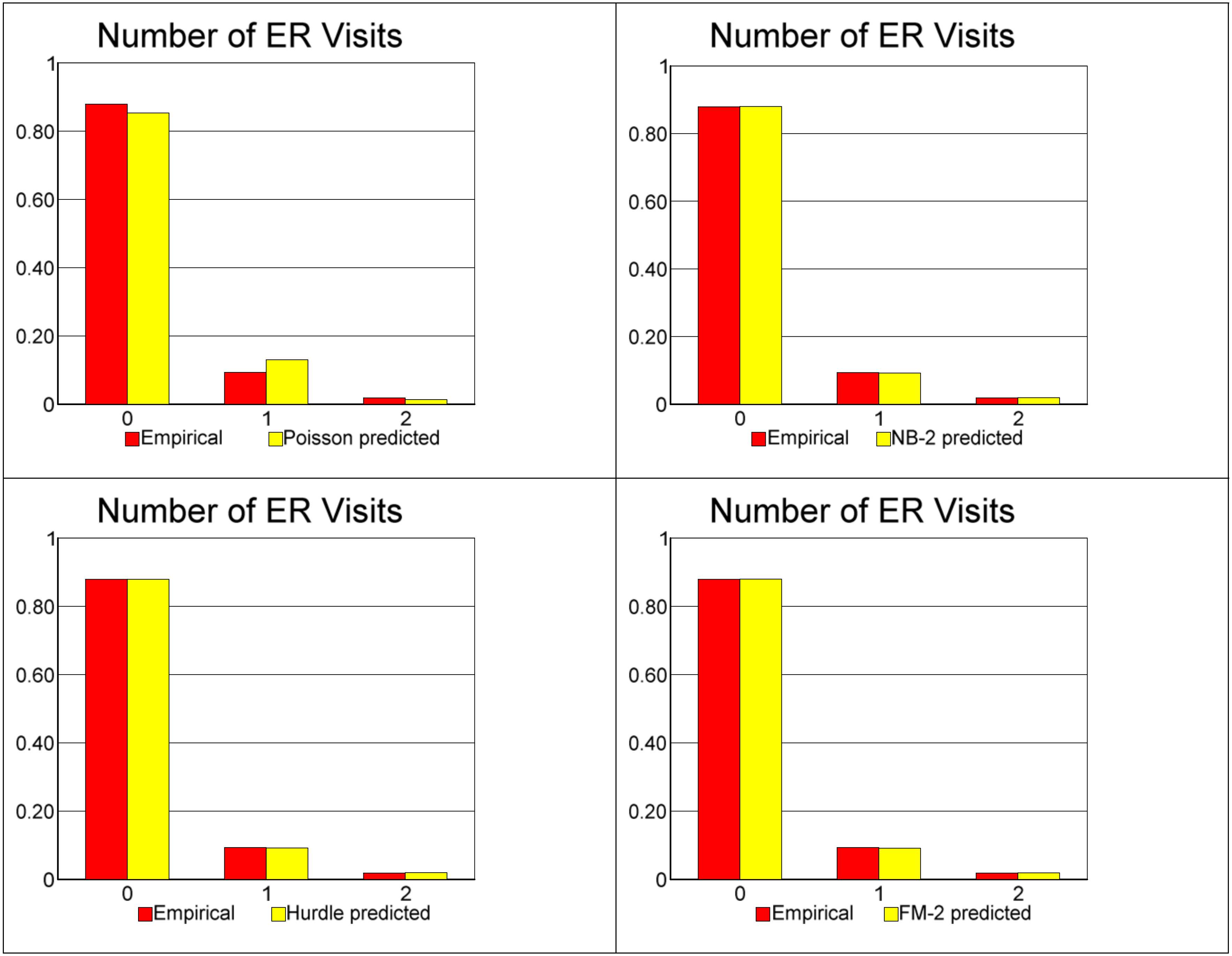
MD Visits

	Poisson	NB-2	Hurdle	FM-2
K	26	27	53	55
LogL	-50744.2	-33792.5	-33392.8	-33252.3
AIC	50770.2	33819.5	33445.8	33307.3
BIC	50870.8	33924.0	33650.9	33520.2
MSPE	1116	4.031	0.801	0.647

ER Visits

	Poisson	NB-2	Hurdle	FM-2
LogL	-8014.6	-7539.8	-7511.2	-7484.2
AIC	8040.6	7566.8	7564.2	7539.2
BIC	8141.2	7671.3	7769.3	7752.1
MSPE	34.86	0.053	0.041	0.067


Number of MD Visits
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
0 1 2 3 4 5 6 7 8 9 10 11
Empirical
Poisson predicted
Number of MD Visits
Empirical
NB-2 predicted
Number of MD Visits
Empirical
Hurdle predicted
Number of MD Visits
Empirical
FM-2 predicted

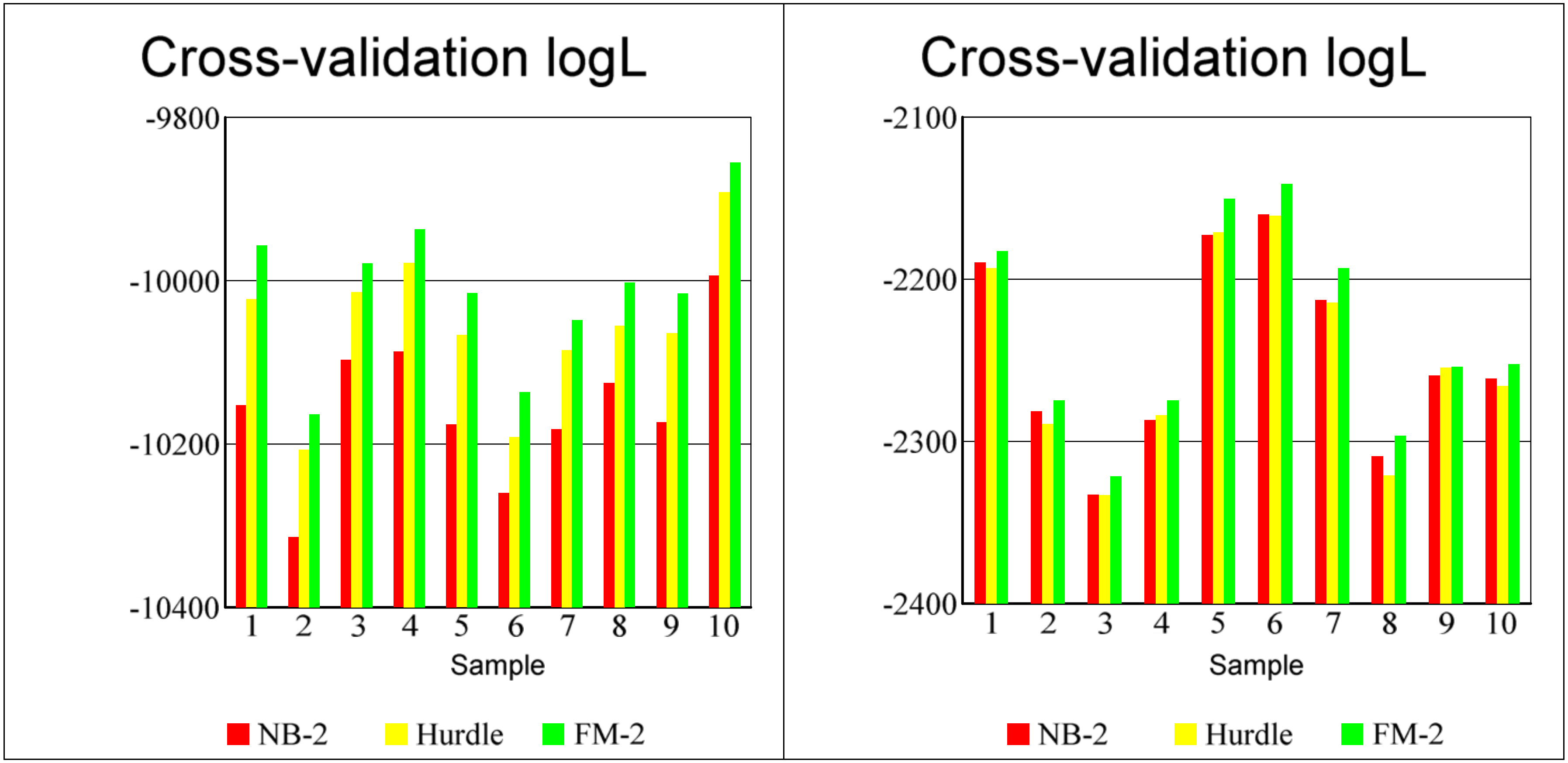


Number of ER Visits
1
0.80
0.60
0.40
0.20
0
0
1
2
Empirical
Poisson predicted
Number of ER Visits
1
0.80
0.60
0.40
0.20
0
0
1
2
Empirical
NB-2 predicted
Number of ER Visits
1
0.80
0.60
0.40
0.20
0
0
1
2
Empirical
Hurdle predicted
Number of ER Visits
1
0.80
0.60
0.40
0.20
0
0
1
2
Empirical
FM-2 predicted

Model Selection -- Cross-validation

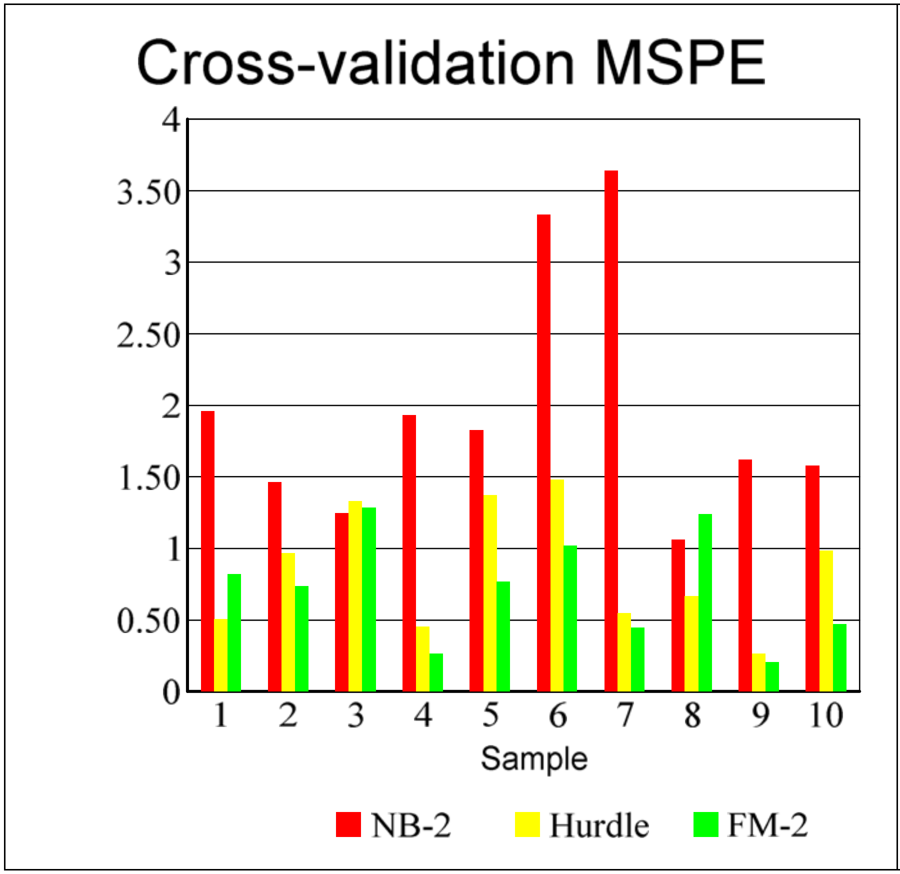
Estimation – 70% random sample
Prediction – Remaining 30% sample

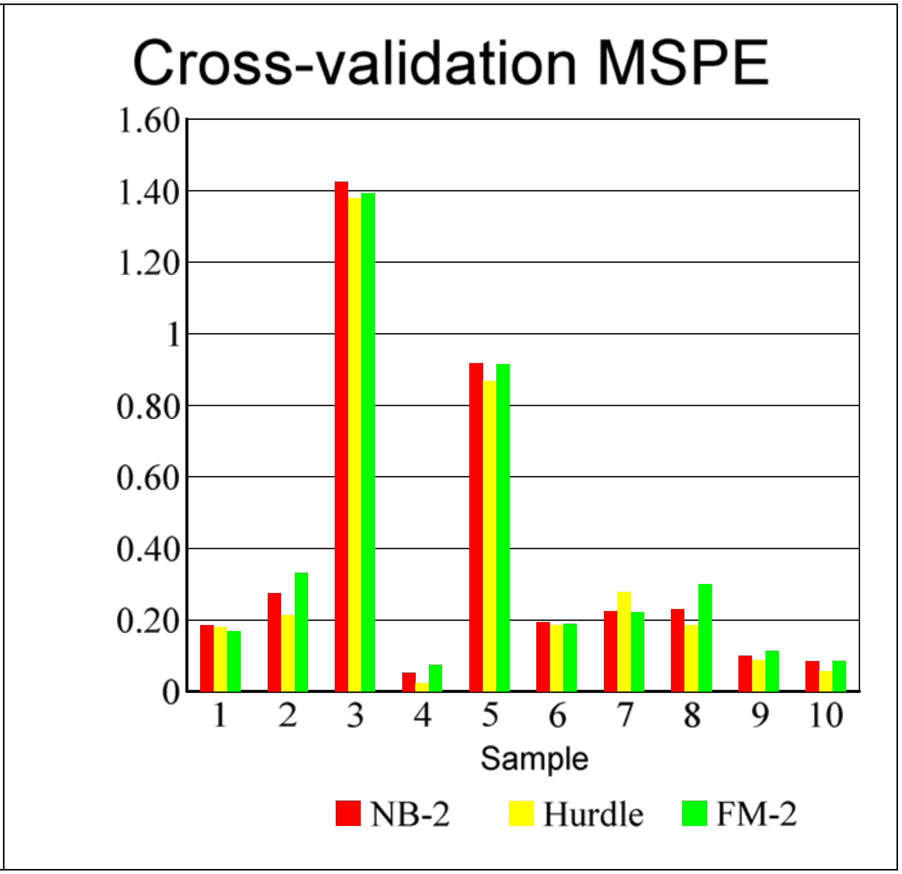
Log Likelihood



Model Selection -- Cross-validation

Mean Square Prediction (of distribution) Error





Finite Mixture Models - Summary

Advantages

- **Strong suspicion that “small” values from different process than “large” values. E.g., sick less price responsive**
- **Marginal effects vary by latent class**
- **Flexible (semiparametric) with sufficient number of classes**
- **Works for discrete data where quantile regression does not**

Disadvantages (in the case of unknown separation)

- **Computationally complex - Likelihood may have multiple maxima**
- **Deciding number of classes is non-trivial and complex**

Overview

Statistical issues -- skewness and the zero mass

Studies with skewed outcomes but no zero mass problem

Studies with zero mass and skewed outcomes

Conditional Density Estimation

Alternatives for studies with zero mass and skewed outcomes

Studies with count data

Finite mixture models

Conclusions

Conclusions

Health Care Outcomes

- **Are pervasively skewed to the right with long right tails**
- **Have substantial zeros**
- **Display heteroskedasticity even after transformation**
- **Display different responsiveness to covariates at different**
- **parts of the distribution**

No single model is "best" for all cases!

Log transform is not the only nor the best solution to skewness

Retransformation is more complicated than meets the eye

Comprehensive model checking is recommended

In-sample checks not always reliable

- **Overfitting is a very real danger**
- **Cross-validation checks recommended**

Conclusions

But it is not all bad news

We have outlined a variety of methods that

- **Work in many disparate situations**
- **Are easy to estimate (generally)**
- **Often provide a better fit**
  - **Are less sensitive to outliers**
  - **Can result in large efficiency gains vis-à-vis linear models**

Also outlined approaches to making decisions about models

Promising methodological work continues in the literature

Websites for handouts, recommended reading and programming code

Will Manning
`mailto:w-manning@uchicago.edu`
`http://harrisschool.uchicago.edu/faculty/web-pages/willard-manning.asp`

follow link to ASHE materials

Edward Norton
`mailto:ecnorton@umich.edu`
`http://www.sph.umich.edu/iscr/faculty/profile.cfm?uniqname=ecnorton`

follow link to Health Econometrics

Partha Deb
`mailto:partha.deb@hunter.cuny.edu`
`http://urban.hunter.cuny.edu/~deb/`

follow link to Teaching and then to Minicourse on Modeling Health Care Costs and Use

References

Afifi, A.A., J.B. Kotlerman, S.L. Ettner, et al. "Methods for improving regression analysis for skewed continuous or counted responses," *Annual Review of Public Health* 28:95-111, 2007.

Ai, C. and E.C. Norton. "Standard Errors for the Retransformation Problem with Heteroscedasticity," *Journal of Health Economics* 19(5):697−718, 2000.

Ai, C. and E.C. Norton. "Interaction Terms in Logit and Probit Models," *Economics Letters* 80(1):123−129, 2003.

Ai, C. and E.C. Norton. "A Semiparametric Derivative Estimator in Log Transformation Models." *The Econometrics Journal* Forthcoming 2009.

Andrews, D.W.K. "Chi-square Diagonostic Tests for Econometric Models," *Journal of Econometrics* 37: 135-156, 1988.

Basu A, B.V. Arondekar, and P.J.Rathouz. "Scale of interest versus scale of estimation: comparing alternative estimators for the incremental costs of a comorbidity," *Health Economics* 15(10):1091-107, 2006

Basu, A., W.G. Manning, and J. Mullahy. "Comparing Alternative Models: Log vs. Cox Proportional Hazard?" *Health Economics* 13(8): 749-765, 2004

Basu, A.and P.J. Rathouz. "Estimating marginal and incremental effects on health outcomes using flexible link and variance function models," *Biostatistics* 6(1): 93-109, 2005.

Blough, D.K., C.W. Madden, and M.C. Hornbrook. "Modeling risk using generalized linear models," *Journal of Health Economics* 18: 153-171, 1999.

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

Burgess, J.F. and P. Deb. "A Quasi-experimental Comparison of Statistical Models for Health Care Expenditures," working paper.

Cameron, C., and P.K. Trivedi. "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics* 1: 29-53, 1986.

Cameron, A.C., and P.K. Trivedi. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.

Cameron, A.C., P.K. Trivedi, F. Milne and J. Piggott. "A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia," *Review of Economic Studies* 55: 85-106, 1988.

Cantoni, E., and E. Ronchetti. "A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures," *Journal of Health Economics* 25: 198-213, 2006.

Cameron, A.C. and F.A.G. Windmeijer. "R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization," *Journal of Business and Economic Statistics* 14: 209-220, 1996.

Chen Y.H**.,** and X.H. Zhou "Interval estimates for the ratio and difference of two lognormal means," *Statistics in Medicine*, 25(23):4099-113, 2006.

Conway, K. and P. Deb (2005), "Is Prenatal Care Really Ineffective? Or, is the 'Devil' in the Distribution?" *Journal of Health Economics*, 24, 489-513.

Cragg, J. "Some Statistical Models for Limited Dependent Variable with Application to the Demand for Durable Goods," *Econometrica* 39: 829-844, 1971.

D'Agostino, R.B., and E.S.Pearson. "Tests for departure from normality: empirical results for the distribution of b_2 and $\sqrt{b_1}$," *Biometrika* 60: 613-622.

Dalton, K., and E.C. Norton. "Revisiting Rogowski and Newhouse on the indirect costs of teaching: A note on functional form and retransformation in Medicare's payment formulas," *Journal of Health Economics* 19(6):1027–1046, 2000.

Deb, P. and A.M. Holmes. "Estimates of Use and Costs of Behavioral Health Care: A Comparison of Standard and Finite Mixture Models," *Health Economics* 9: 475-489, 2000.

Deb, P. and P.K. Trivedi. "Demand for Medical Care by the Elderly: A Finite Mixture Approach," *Journal of Applied Econometrics* 12: 313-326, 1997.

Deb, P. and P. K. Trivedi. "The Structure of Demand for Health Care: Latent Class versus Two-part Models, " *Journal of Health Economics* 21: 601-625, 2002.

Donald, S., D. Green, and H. Paarsch. "Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates," *Review of Economic Studies* 67(4):609-633.

Dow, W.H. and E.C. Norton. "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions," *Health Services & Outcomes Research Methodology* 4: 5-18, 2003.

Duan, N. "Smearing Estimate: a Nonparametric Retransformation Method," *Journal of the American Statistical Association* 78: 605-610, 1983.

Duan, N., W.G. Manning, et al. "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economics Statistics* 1:115-126, 1983.

Duan, N., W.G. Manning, et al. "Choosing Between the Sample-Selection Model and the Multi-Part Model," *Journal of Business and Economic Statistics* 2(3): pp. 283-289, 1984.

Duan, N., W.G. Manning, et al. "Comments on Selectivity Bias," in R.M. Scheffler and L.F. Rossiter, eds., *Advances in Health Economics and Health Services Research, Vol. 6, Biased Selection in Health Care Markets*. Greenwich, Conn.: JAI Press, Inc., 1985.

Efron, B. "Logistic regression, survival analysis, and the Kaplan-Meier curve" *Journal of the American Statistical Association* 83:414-425.

Gerdtham, U.-G. “Equity in Health Care Utilization: Further Tests Based on Hurdle Models and Swedish Micro Data,” *Health Economics* 6: 303-319, 1997.

Gerdtham, U.-G. and P.K. Trivedi. “Equity in Swedish Health Care Reconsidered: New Results Based on the Finite Mixture Model,” *Health Economics* 10: 565-572, 2001.

Gilleskie, D.B. and T.A. Mroz. “A Flexible Approach for Estimating the Effects of Covariates on Health Expenditures,” *Journal of Health Economics* 23(2): 391-418, 2004.

Gourieroux, C. A. Montfort, and A. Trognon. “Pseudo-maximum Likelihood Methods: Applications to Poisson Models,”. *Econometrica* 52: 701-720, 1984.

Gurmu, S., and P.K. Trivedi. “Excess Zeros in Count Models of Recreational Trips,” *Journal of Business and Economic Statistics* 14: 469-477, 1996.

Guo, J.Q. and P.K. Trivedi. “Flexible Parametric Models for Long-Tailed Patent Count Distributions,” *Oxford Bulletin of Economics and Statistics* 64: 63-82, 2002.

Hosmer, D.W., and S. Lemeshow. *Applied Logistic Regression,* 2nd Edition. New York, John Wiley & Sons, 1995.

Jimenez-Martin, S., J.M. Labeaga and M. Martinez-Granado. “Latent Class Versus Two-part Models in the Demand for Physician Services Across the European Union”, *Health Economics* 11, 301-321, 2002.

Jones, A. “Health Econometrics,” in A. Culyer and J. Newhouse, (Eds*.), Handbook of Health Economics*. Amsterdam: Elsevier, 2000.

Koenker, R.W and Bassett, G.W. "Regression Quantiles," *Econometrica* 46:33-50, 1978.

Koenker, R.W and Bassett, G.W. "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica* 50:43-61, 1982.

Lane, P.W. Generalized linear models in soil science,” *European Journal of Soil Science* 53: 241-251, 2002.

Leung, S.F., Shihti Yu. “On the Choice between Sample Selection and Two-part Models,” *Journal of Econometrics* 72: 197-229, 1996.

Lindsay, B.J. *Mixture Models: Theory, Geometry, and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, IMS-ASA, 1995.

Liu, Z., W.H. Dow, and E.C. Norton. “Effect of drive-through delivery laws on postpartum length of stay and hospital charges,” *Journal of Health Economics* 23(1):129–155, 2004.

Lourenço, O.D. and P.L. Ferreira, "The Impact of Non-monetary Factors on the Primary Care Utilization in Portugal: Finite Mixture Models Applied to On-site and Truncated Samples, *Working Paper*, 2004.

McCullagh, P. and J.A. Nelder. 1989. *Generalized linear models*, 2nd Edition. London: Chapman and Hall.

McLachlan, G.J., and D. Peel. *Finite Mixture Models*, New York: John Wiley, 2000.

Manning, W.G. “The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem,” *Journal of Health Economics* 17: 283-295, 1998.

Manning, W.G., A. Basu, and J. Mullahy. "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," *Journal of Health Economics* 24: 465-488, 2005.

Manning, W.G., N. Duan, and W.H. Rogers. "Monte Carlo Evidence on the Choice between Sample Selection and Two-part Models," *Journal of Econometrics* 35: 59-82, 1987a.

Manning, W.G., J.P. Newhouse, et al. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment,". *American Economic Review* 77(3):251-277, 1987b.

Manning, W.G., and J. Mullahy. "Estimating Log Models: To transform or not to Transform?" *Journal of Health Economics* 20(4): 461-494, 2001.

Mroz, T.A. "A Simple, Flexible Estimator for Count and other Ordered Discrete Data," Clemson University Working Paper, 2008.

Mullahy, J. "Specification and Testing of Some Modifed Count Data Models," *Journal of Econometrics* 33: 341-365, 1986.

Mullahy, J. "Heterogeneity, Excess Zeros, and the Structure of Count Data Models," *Journal of Applied Econometrics* 12: 337-350, 1997.

Mullahy, J. "Much Ado about Two: Reconsidering Tetransformation and the Two-part Model in Health Econometrics," *Journal of Health Economics* 17: 247-281, 1998.

Newhouse, J.P., et al. *Free-for-all: Health Insurance, Medical Costs, and Health Outcomes: the Results of the Health Insurance Experiment.* Cambridge: Harvard University Press, 1993.

Norton, E.C., G.S. Bieler, S.T. Ennett, and G.A. Zarkin. "Analysis of Prevention Program Effectiveness with Clustered Data Using Generalized Estimating Equations," *Journal of Consulting and Clinical Psychology* 64(5):919–926, 1996.

Norton, E.C., W.H. Dow, and Y. Do. "Specification tests for the sample selection and two-part models," *Health Services and Outcomes Research Methodology* Forthcoming 2008.

Norton, E.C., R.C. Lindrooth, and S.L. Ennett. "How Measures of Perception from Survey Data Lead to Inconsistent Regression Results: Evidence from Adolescent and Peer Substance Use," *Health Economics* 12(2):139–148, 2003.

Norton, E.C., H. Wang, and C. Ai. "Computing interaction effects and standard errors in logit and probit models," *The Stata Journal* 4(2):154-167.

Park, R. "Estimation with Heteroscedastic Error Terms," *Econometrica* 34: 888, 1966.

Pohlmeier,W. and V.Ulrich. "An Econometric Model of the Two-Part Decision-making Process in the Demand for Health Care," *Journal of Human Resources* 30: 339-361, 1995.

Pregibon, D. "Goodness of Link Tests for Generalized Linear Models," *Applied Statistics* 29: 15-24, 1980.

Pregibon, D. "Logistic Regression Diagnostics," *Annals of Statistics* 9: 705-724, 1981.

Schellhorn, M. "The Effect of Variable Health Insurance Deductibles on the Demand for Physician Visits," *Health Economics*, 10, 441-456, 2001.

Taylor, J. M. "The Retransformed Mean after a Fitted Power Transformation." *Journal of the American Statistical Association* 81: 114-118, 1986.

Tobin, J. "Estimation of Relationships for Limited Dependent Variables," *Econometrica* 26: 24-36, 1958.

Tu, W., and X.H. Zhou, "A Wald Test Comparing Medical Costs Based on Log-normal Distributions with Zero Valued Costs," *Statistics in Medicine* 30;18(20):2749-61, 1999.

van de Ven, W.P., B.M. van Praag. "Risk Aversions of Deductibles in Private Health Insurance: Application of an Adjusted Tobit Model to Family Health Care Expenditures," in J. van der Gaag and M. Perlman, eds., *Health, Economics, and Health Economics*. Amsterdam: North Holland, 1981.

Wang, P., I.M. Cockburn, and M.L. Puterman. "Analysis of Patent Data - A Mixed Poisson Regression Model Approach," *Journal of Business and Economic Statistics* 16, 27-36, 1998.

Wedel, M., W.S. Desarbo, J.R. Bult and V. Ramaswamy. "A Latent Class Poisson Regression Model for Heterogeneous Count Data," *Journal of Applied Econometrics* 8, 397-411, 1993.

Winkelmann, R. "Duration Dependence and Dispersion in Count-Data Models," *Journal of Business and Economic Statistics* 13, 467-474, 1995.

Wooldridge, J.M. "On the Application of Robust, Regression-based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47: 5-46, 1991.

Zhou XH, S. Gao, S.L. Hui. "Methods for comparing the means of two independent log-normal samples," *Biometrics* 53(3):1129-35, 1997.

Zhou XH, and W. Tu. "Confidence intervals for the mean of diagnostic test charge data containing zeros." Biometrics. 56(4):1118-25, 2000.

Survival and Censored Cost Papers

Bang, H., and A.S. Tsiatis. "Estimating Medical Costs with Censored Data," *Biometrika* 87(2); 329-343, 2000.

Basu, A., W.G. Manning, and J. Mullahy. "Comparing Alternative Models: Log vs. Cox Proportional Hazard?" *Health Economics* 13(8): 749-765, 2004.

Dudley RA, Harrell Jr. FE, Smith LR. *et al*. "Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery," *Journal of Clinical Epidemiology* 46(3): 261-271, 1993.

Fenn, P, McGuire A, Phillips V. *et al*. "The analysis of censored treatment cost data in economic evaluation," *Medical Care* 33(8): 851-863, 1995.

Etzioni RD, Feuer EJ, Sullivan SD, e*t al.* "On the use of survival analysis techniques to estimate medical care costs," *Journal of Health Economics* 18: 365-380, 1999.

Hallstrom A, Sullivan SD. "On estimating costs for economic evaluation in failure time studies," *Medical Care* 36(3): 433-436, 1998.

Jain, A.K., and R.L. Strawderman. "Flexible hazard regression modeling for medical cost data." *Biostatistics* 3:101-118, 2002.

Jiang, H., and X.H. Zhou. "Bootstrap confidence intervals for medical costs with censored observations." Statistics in Medicine. 23(21):3365-76, 2004.

Lin, D.Y. "Linear regression analysis of censored medical costs," *Biostatistics* 1(1):35-47, 2000.

Lin, D.Y. "Proportional means regression for censored medical costs," *Biometrics* 56(3):775-8, 2000.

Lin, D.Y. "Regression analysis of incomplete medical cost data," *Statistics in Medicine* 22(7):1181-200, 2003.

Lipscomb J, Ancukiewicz M, Parmigiani G., *et al*. "Predicting the cost of illness: A comparison of alternative models applied to stroke," *Medical Decision Making* 18(2): S39-S56, 1998.

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

O'Hagan, A., and J.W. Stevens. "On Estimators of Medical Costs with Censored Data," *Journal of Health Economics* 23: 615-625, 2004.

Raikou, M., and A.McGuire. "Estimating Medical Costs Under Conditions of Censoring," *Journal of Health Economics* 23: 443-470, 2004.

Willan, AR, D.Y, Lin, and A. Manca. "Regression methods for cost-effectiveness analysis with censored data," *Statistics in Medicine* 24:131-145, 2005