

Performance of Statistical Models to Predict Mental Health and Substance Abuse Cost

Maria Montez-Rath^{1,2}

Cindy L Christiansen^{2,3}, Susan L Ettner⁴, Susan Loveland^{2,3}, and Amy K Rosen^{2,3}

HERC Cyber Seminars
Nov 8, 2006

1 BU SPH, Department of Biostatistics

2 Center for Health Quality, Outcomes and Economic Research, Bedford VAMC

3 BUSPH, Health Services Department

4 UCLA, Division of General Internal Medicine and Health Services Research

Supported by VA HSR&D Service, Grant # IIR 20-035-2

Background

Mental Health Project

- Population in the Department of Veterans Affairs (VA) is highly comorbid
 - 29% of patients seen during fiscal year 1999 (FY99) had mental health and substance abuse (MH/SA) disorders
 - Average of 14 diagnoses (medical and psychiatric) compared to 9 in the general VA population
- No risk-adjustment model developed specifically for patients with psychiatric disorders

MH Project

- Subjects:
 - 914,225 veterans
 - who used VA healthcare services during FY'99
 - with any MH/SA diagnosis
- Goal: develop and validate a case-mix classification system: “PsyCMS”
 - for predicting concurrent and prospective MH/SA costs and utilization

Performance Paper

- Ordinary least squares with risk adjustment generally used to model cost in health services research
- Why do economists use other models?
- Make sure that case-mix system developed does not depend on statistical model used

Data

- Sample: individuals on development sample (60%) with positive MH/SA costs (n=525,620)
- Adjusters: 20 age/sex + 12 MH/SA categories
- Outcome: total MH/SA costs (does not include outpatient pharmacy)

Methods – Model Specification

- **Robust** regression to take into account patients clustered within facilities
 - each patient assigned to last facility where care was received
- **Weighted** regression to adjust costs of patients who died during the FY

Stata: glm y_wt x [aweight=wt], f(modeldist) l(modellink) robust cluster (sta3n)

- Forced mean predicted = mean observed
- Cost data characteristics guided choice of model
 - non-negative and right skewed

Model's Description

| Model Name | Dependent Variable | Model Specification | | | |
|----------------------------------|--------------------|---------------------|----------|------------|--------------------|
| | | Family | Link | V(μ) | Description |
| Normal with identity link | Cost | Normal | Identity | 1 | $E(Y) = xb$ |
| Log Normal | Log(cost) | Normal | Identity | 1 | $E(Y) = xb$ |
| Sqrt Normal | Sqrt(cost) | Normal | Identity | 1 | $E(Y) = xb$ |
| Gamma with log link | Cost | Gamma | Log | μ^2 | $\ln\{E(Y)\} = xb$ |
| Gamma with sqrt link | Cost | Gamma | Sqrt | μ^2 | $\sqrt{E(Y)} = xb$ |

Note: $\mu = E(Y)$ and $V(\mu)$ is variance function

- OLS + 2 transformation models + 2 GLMs
- Distribution family identifies variance function
 - Park test

Normal with Identity Link (OLS)

- Additive effects on original scale

$$E\{Y\}=X\beta$$

where Y iid Normal with constant variance

- Predictions given in original scale: $\hat{y} = X \hat{\beta}$
- mean predicted = mean observed
- Does not deal with cost data characteristics:
 - non-negative and right-skewed

Stata: `glm y x, f(gaussian) l(identity)`

Log Normal (1)

- Additive effects on **log** scale:

$$E\{\ln(Y)\} = X\beta$$

where $\ln(Y)$ iid Normal with constant variance

- Predictions given in log scale \Rightarrow retransformation
- Direct transformation gives biased estimates
 - assumes $E\{\ln(Y)\} = \ln\{E(Y)\}$

Log Normal (2)

- Bias correction at population level by
 - forcing mean predicted = mean observed
 - assuming homoscedastic error
- Final predictions: $\hat{y} = s * \exp(x \hat{\beta})$
where $s = \sum y / \sum \exp(x \hat{\beta})$

Stata: gen lny=ln(y)
glm lny x, f(gaussian) l(identity)

Square-root Normal

- Additive effects on **sqrt** scale

$$E\{\text{sqrt}(Y)\} = X\beta$$

where $\text{sqrt}(Y)$ iid Normal with constant variance

- Similar problems to log transformation
- Final predictions: $\hat{y} = s * (x \hat{\beta})^2$
where $s = \sum y / \sum (x \hat{\beta})^2$
- Squaring $x \hat{\beta}$ implicitly introduces two-way interactions

Generalized Linear Models (GLMs)

- Models of the form

$$g\{E(Y)\} = X\beta$$

where g is the link function and $Y \sim$ distribution

- Well defined by specifying mean and variance
- No retransformation problems
- Predictions are given directly by $\hat{y} = g^{-1}(x \hat{\beta})$

Gamma Models

- Gamma distribution:
 - always positive
 - long tail
- Two links: log and square-root

Stata: `glm y x, f(gamma) l(log)`
`glm y x, f(gamma) l("power .5")`

- Final predictions: $\hat{y} = s * g^{-1}(x \hat{\beta})$

where $s = \sum y / \sum g^{-1}(x \hat{\beta})$

g is either log or square-root function

Methods – Model Selection

- Measures of predictive ability:

1.
$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

2.
$$MAPE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

3. Predictive ratios of decile of predicted cost

$$PR_j = \frac{\bar{\hat{y}}_j}{\bar{y}_j}$$

- Bias-corrected bootstrap confidence intervals

Methods - Sample size study

- Objective: study effect of analyzing random sample
 - Run specified models on subsamples starting at 5,000 patients (~1% of total sample)
 - 100 samples for each sample size
 - Re-compute statistics
 - Generate 95% percentile intervals for each statistic at each sample size

Results

- MH/SA cost
 - mean = \$2,602
 - SD = \$11,052
 - median = 385
 - skewness = 14
- Park test gives $\hat{\gamma} = 1.6 \Rightarrow$ Gamma variance
- Box-Cox gives $\hat{\lambda} = -.12$
(hard to know what it means)

RMSE and MAPE results on total sample

| Model | RMSE | | | MAPE | | |
|-----------------------------------|----------|-----------|--------|----------|-----------|-------|
| | Estimate | 95% Conf. | Int.* | Estimate | 95% Conf. | Int.* |
| Gaussian with identity link (OLS) | 10,397 | 10,130 | 10,657 | 2,997 | 2,941 | 3,052 |
| Log Normal | 13,974 | 13,585 | 14,352 | 2,801 | 2,759 | 2,840 |
| Sqrt Normal | 9,860 | 9,644 | 10,070 | 2,554 | 2,514 | 2,592 |
| Gamma with log link | 21,374 | 20,246 | 22,552 | 3,324 | 3,249 | 3,395 |
| Gamma with square-root link | 10,434 | 10,193 | 10,708 | 2,797 | 2,744 | 2,859 |

*Bias-corrected bootstrap confidence interval

Note: RMSE = root mean square error

MAPE = mean absolute prediction error

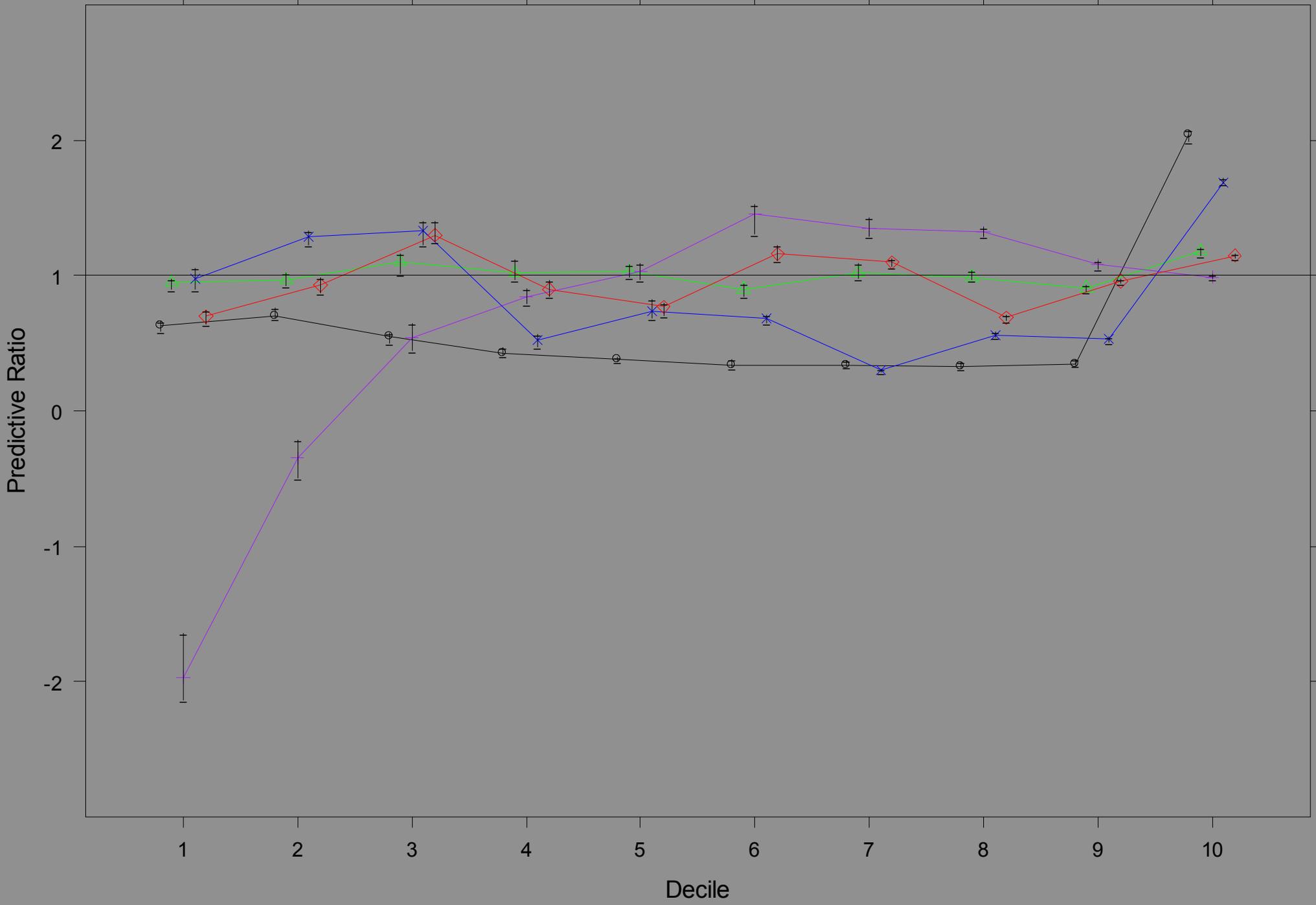
Predicted ratios and bootstrap bias-corrected CIs per decile of predicted cost in full sample

- Normal with identity link (OLS) predicted 19.6% negative values
=> First two deciles in the negative range

Gamma Log \circ —
Gamma Square Root \triangle —

Gaussian Identity $+$ —
Log Normal \times —

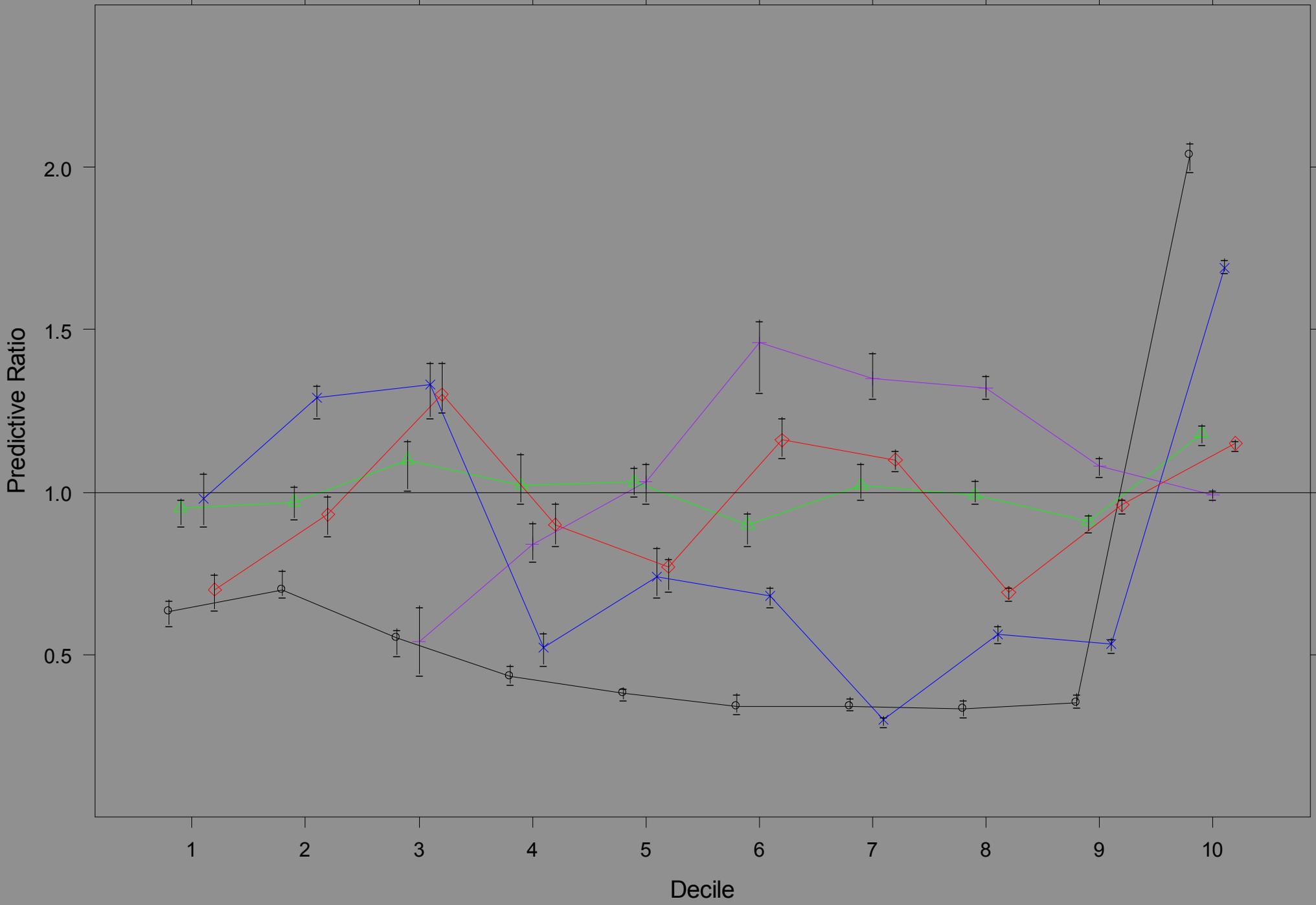
Sqrt Normal \diamond —



Previous graph without PRs for 1st two deciles of OLS

- Gamma with square-root link (range .9 to 1.2)
 - PRs close to 1.0 across all 10 deciles
 - 5/10 intervals include 1.0
- OLS extremely well at top decile (range .5 to 1.5)
 - 2/10 intervals include 1.0
- Square-root Normal (range .7 to 1.3)
 - 0/10 intervals include 1.0
- Log Normal well only at 1st decile (range .3 to 1.7)
 - either under or overpredicts by 25% in 9/10 intervals
- Gamma with log link a complete disaster!

Gamma Log ○ — Gaussian Identity + — Sqrt Normal ◇ —
 Gamma Square Root △ — Log Normal × —



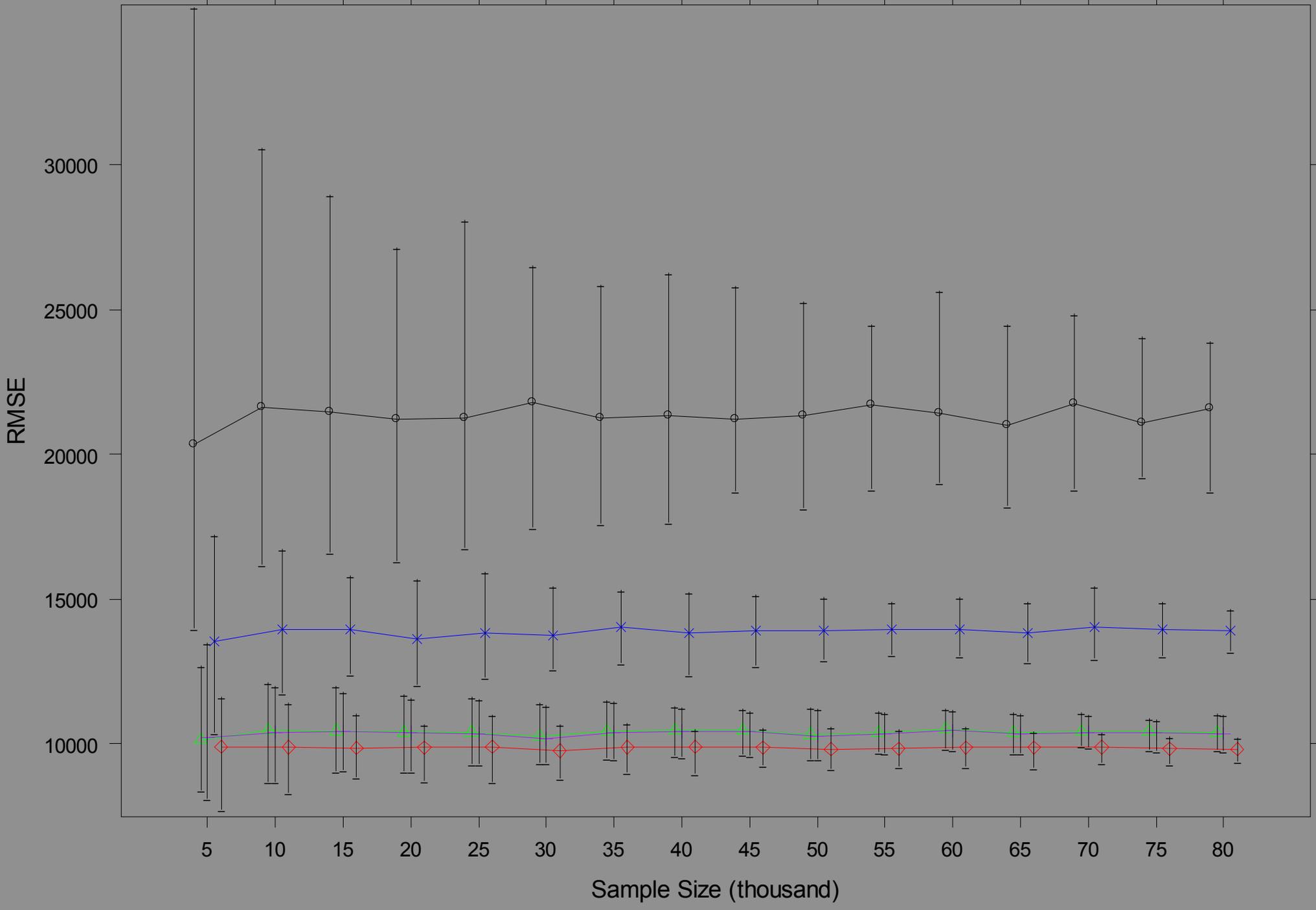
Simulation Results

- Gamma with square-root link had convergence problems for small samples
- Frequency of samples for which models did not converge ranged (out of 100):
 - 1 sample in simulation for 55,000 patients
 - 50 samples in simulation for 5,000 patients
- Problems occurred with 3 female categories
- Other models had no problems converging for those same samples

95% RMSE percentile intervals per model at each simulation of various sample sizes

- Log-Normal PI only overlaps for sample with 5,000 patients
- Gamma with log link always above mean values
- OLS and Gamma with square-root link have very similar results
- Interval overlap with Square-root Normal decreases with increasing sample size

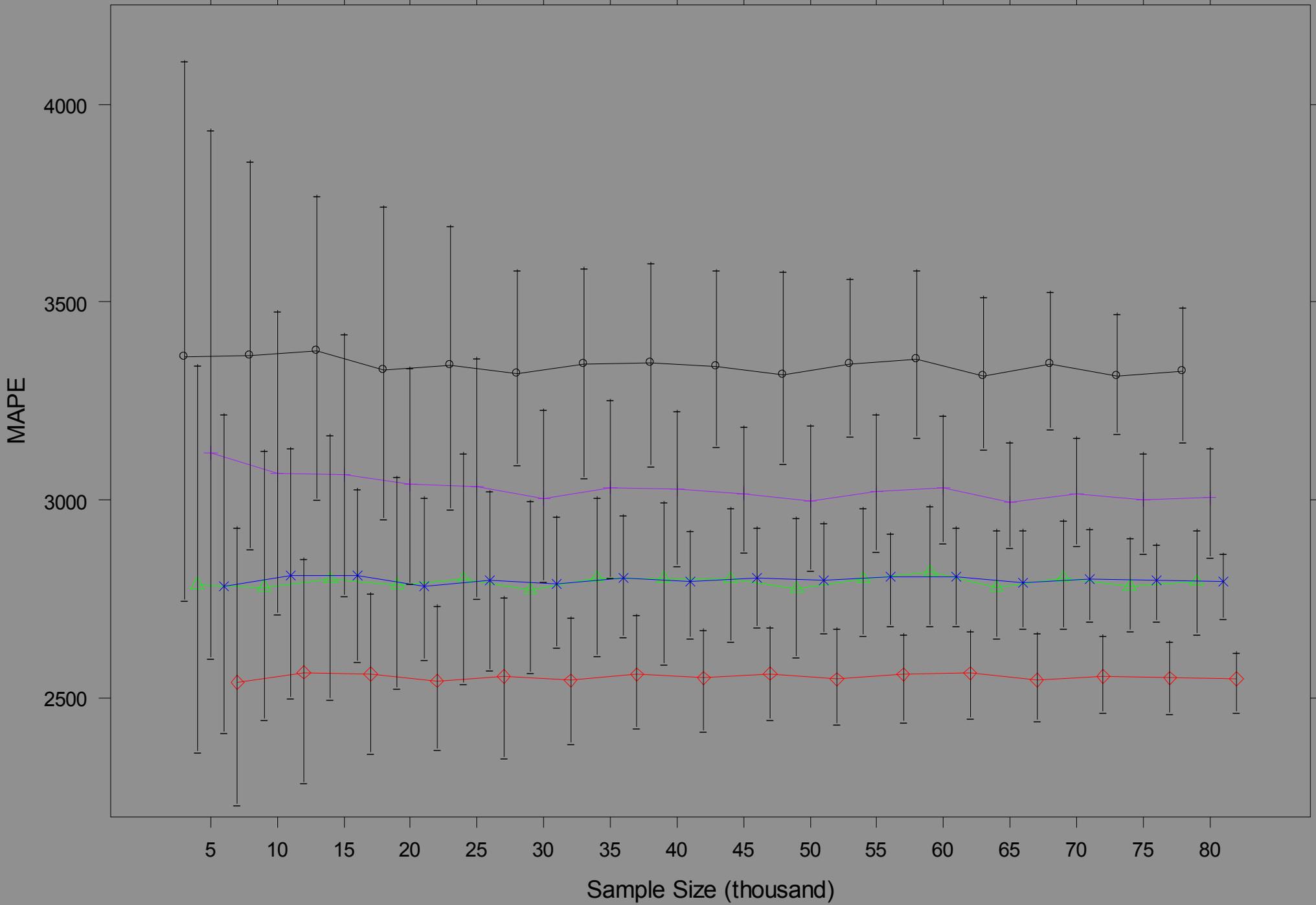
Gamma Log ○ — Gaussian Identity + — Sqrt Normal ◇ —
 Gamma Square Root △ — Log Normal × —



95% MAPE percentile intervals per model at each simulation of various sample sizes

- Bigger overlap between all models
- Square-root Normal with:
 - consistent smallest lower bound
 - upper bound lower than other mean values starting at samples with 15,000 patients
- Log Normal and Gamma with square-root link closest

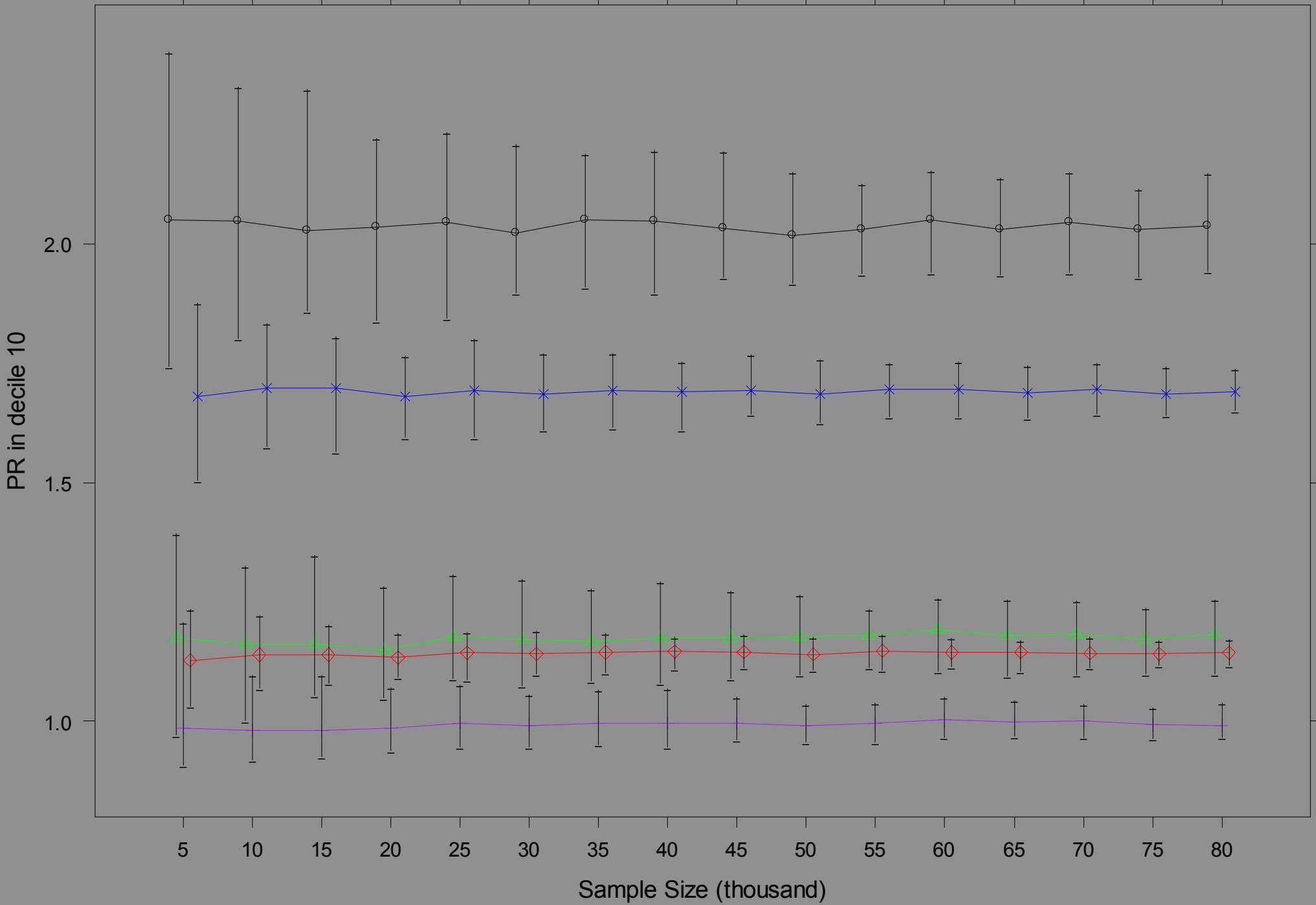
Gamma Log ○ — Gaussian Identity + — Sqrt Normal ◇ —
 Gamma Square Root △ — Log Normal × —



95% PR10 percentile intervals per model at each simulation of various sample sizes

- OLS intervals
 - always include 1.0
- Gamma with square-root link intervals
 - only include 1.0 for samples with 5,000 patients
- Square-root Normal intervals:
 - are the smallest
 - overlap only for the smallest sample sizes
- Gamma with log link and Log Normal intervals
 - are always higher

Gamma Log ○ — Gaussian Identity + — Sqrt Normal ◇ —
 Gamma Square Root △ — Log Normal × —



Discussion

- Gaussian with identity link (OLS)
 - + does well overall
 - predicts negative values 20% of time
 - + performs the best for top decile
- Gamma with log link was worse
- Log Normal model does reasonably well but overpredicts in top decile
 - multiplicative effect after retransformation

Discussion (2)

- Gamma with Square Root link performs very well in general but has convergence problems for small samples
 - + forces interactions between independent variables
 - + deals with long tail
- Sqrt Normal model performs very well
 - + sqrt transformation forces two-way interactions between independent variables
 - normal distribution does not deal with long tail

Limitations

- More advanced methods proposed:
 - Basu and Rathouz 2005 introduced method that directly estimates link in a GLM model
 - Manning, Basu, and Mullahy 2005 introduced generalized Gamma model
 - use 3 parameters instead of 2
- Only one risk-adjustment system used

Conclusion

- Models with square-root transformation or link are best in full sample
 - function helps to deal with high comorbidity of population
- Gamma distribution models variance better
- OLS suitable if correct transformation used
 - still does best for top decile
- GLMs with Gamma distribution need more data

Thank you!

**for additional information email
montez@bu.edu**

Main References

Maria Montez-Rath, Cindy L Christiansen, Susan L Ettner, Susan Loveland, Amy K Rosen. **Performance of statistical models to predict mental health and substance abuse cost.** BMC Medical Research Methodology 2006, 6:53

on the web at: <http://www.biomedcentral.com/bmcmedresmethodol/>

Rosen AK, Christiansen CL, Montez ME, Loveland S, Shokeen P, Sloan KL, Ettner SL: **Evaluating risk-adjustment methodologies for patients with mental health and substance abuse disorders in the Veterans Health Administration.** International Journal of Healthcare Technology and Management 2006, 7:43-81.

on the web at: http://www.inderscience.com/search/index.php?action=record&rec_id=7872

Sloan KL, Montez-Rath ME, Spiro A 3rd, Christiansen CL, Loveland S, Shokeen P, Herz L, Eisen S, Breckenridge JN, Rosen AK. **Development and validation of a psychiatric case-mix system.** Medical Care 2006, 44(6):568-80.

on the web at:

<http://www.lww-medicalcare.com/pt/re/medcare/abstract.00005650-200606000-00012.htm;jsessionid=Fl>

Basu A, Rathouz PJ: **Estimating marginal and incremental effects on health outcomes using flexible link and variance function models.** Biostatistics 2005, 6:93-109.

pubmed:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15618530&dopt=A

Manning WG, Basu A, Mullahy J: **Generalized modeling approaches to risk adjustment of skewed outcomes data.** Journal of Health Economics 2005, 24:465-488.

pubmed:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=15811539&dopt=A

Bootstrap code Stata v8.2

```
program drop _all
set more off
capture log close
version 8

*****
** MAIN PROGRAM **
*****

capture program drop main_pgm
program define main_pgm

    xi: glm y_an i.age_sex mh01-mh12 [aweight=factor99], robust
        cluster(sta3n) f($modeldist) l($modellink)

    predict yhat1
    quietly summarize y_an [aweight=factor99]
    gen y_an_mean=r(mean)
    quietly summarize yhat1 [aweight=factor99]
    gen yhat1_mean=r(mean)
    gen sm=y_an_mean/yhat1_mean
    gen yhat2=sm*yhat1

    *RMSE after smearing
    gen diff_sq=(y_an-yhat2)^2
    quietly summarize diff_sq [aweight=factor99]
    gen rmse=sqrt(r(mean))

    *MAPE after smearing
    gen adiff=abs(y_an-yhat2)
    quietly summarize adiff [aweight=factor99]
    gen mape=r(mean)

    *PR
    *create deciles
    sort yhat2
    gen precat1 = int(10*_n/(_N+1))
    gen precat=precat1+1
```

```
quietly summarize y if precat==1
gen y_pred_m=r(mean)
quietly summarize yhat2 if precat==1
gen yhat2_pred_m=r(mean)
gen pr1=yhat2_pred_m/y_pred_m

...

quietly summarize y if precat==10
replace y_pred_m=r(mean)
quietly summarize yhat2 if precat==10
replace yhat2_pred_m=r(mean)
gen pr10=yhat2_pred_m/y_pred_m

end

*****
*** BOOTSTRAPPING ***
*****

capture program drop mainboot
program define mainboot, rclass

    *CALL THE PROGRAM TO GET POINT ESTIMATES
    main_pgm

    tempname y1
    sum rmse, meanonly
    scalar `y1'=r(mean)
    return scalar y1=`y1'

    ...

    tempname y12
    sum pr10, meanonly
    scalar `y12'=r(mean)
    return scalar y12=`y12'

end
```

Bootstrap code (2)

```
*****
*** FULL ANALYSIS ***
*****
capture program drop full_analysis
program define full_analysis

***Enter the name of your dataset here***
global dataset develop_set

*** RUN BOOTSTRAPPING TO GET CI FOR THE PREDICTIVE MARGINS
notes drop _all
note: TS Starting boot
notes

*SET THE SEED
set seed 374585038

use $dataset, clear

*CALL THE BOOTSTRAP PROGRAM, SPECIFYING THE NUMBER OF REPETITIONS
bootstrap "mainboot" rmse=r(y1) mape=r(y2) pr1=r(y3) pr2=r(y4) pr3=r(y5) pr4=r(y6) pr5=r(y7) pr6=r(y8) pr7=r(y9) pr8=r(y10) pr9=r
(y11) pr10=r(y12), reps(1000) level(95) dots strata(sta3n)
drop _all

note: TS End boot
notes

end

global modeldist gamma
global modellink "power 0.5"
full_analysis

global modeldist gaussian
global modellink identity
full_analysis
```