Lecture 2
Basic Bayes and two stage normal normal model…

Diagnostic Testing

Diagnostic Testing

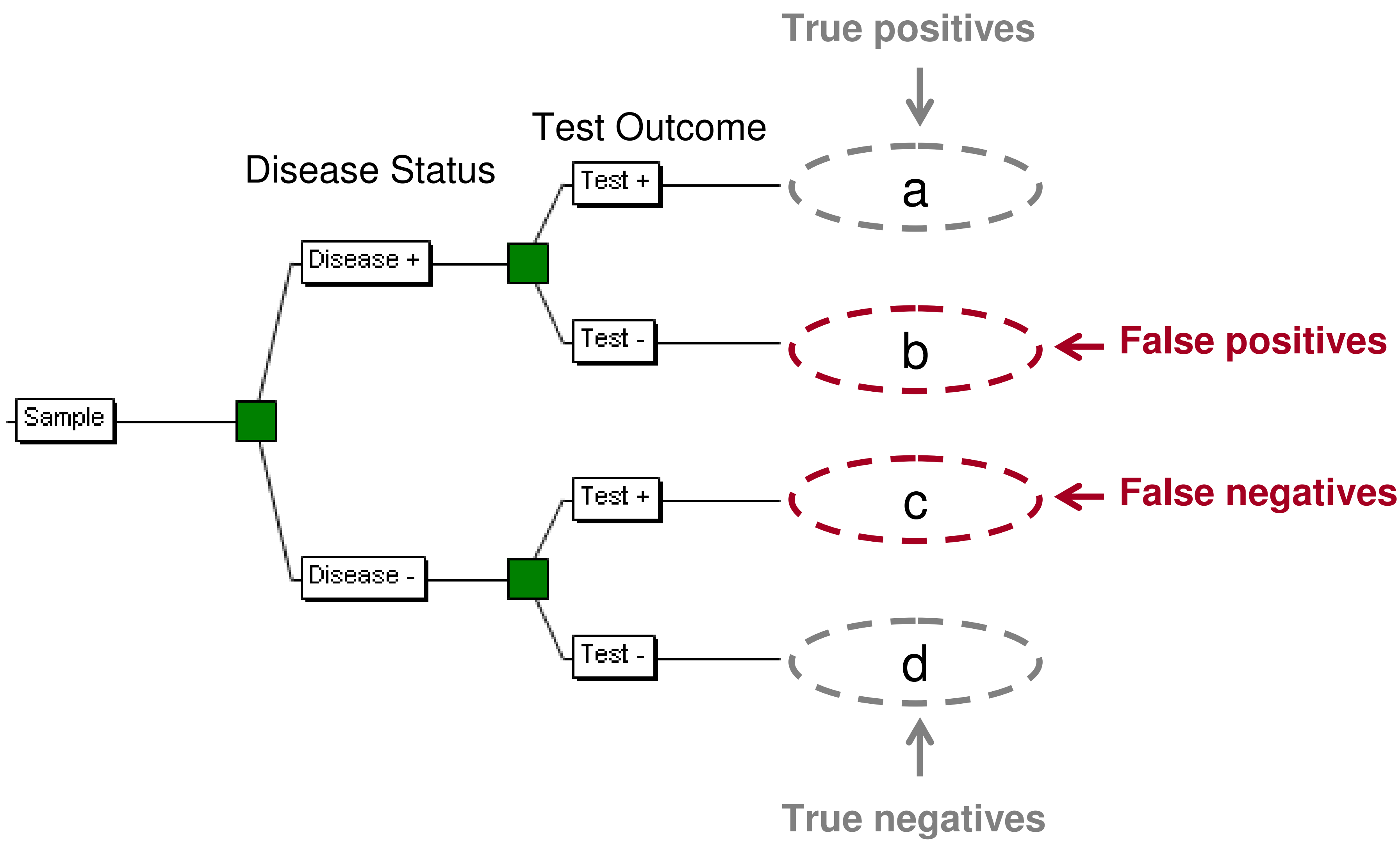
Ask

Marilyn®

BY MARILYN VOS SAVANT

A particularly interesting and important question today is that of testing for drugs. Suppose it is assumed that about 5% of the general population uses drugs. You employ a test that is 95% accurate, which we'll say means that if the individual is a user, the test will be positive 95% of the time, and if the individual is a nonuser, the test will be negative 95% of the time. A person is selected at random and is given the test. It's positive. What does such a result suggest? Would you conclude that the individual is a drug user? What is the probability that the person is a drug user?

Diagnostic Testing



Diagnostic Testing

- "The workhorse of Epi": The 2×2 table

	Disease +	**Disease -**	**Total**
Test +	***a***	***b***	***a + b***
Test -	***c***	***d***	***c + d***
Total	***a + c***	***b + d***	***a + b + c + d***

Diagnostic Testing

- "The workhorse of Epi": The 2×2 table

	Disease +	Disease -	Total
Test +	***a***	***b***	***a + b***
Test -	***c***	***d***	***c + d***
Total	***a + c***	***b + d***	***a + b + c + d***

$$Sens = P(+ \mid D) = \frac{a}{a+c} \qquad Spec = P(- \mid \overline{D}) = \frac{d}{b+d}$$

Diagnostic Testing

- "The workhorse of Epi": The 2 × 2 table

	Disease +	Disease -	Total
Test +	***a***	***b***	***a + b***
Test -	***c***	***d***	***c + d***
Total	***a + c***	***b + d***	***a + b + c + d***

positive predicted value

$$PPV = P(D \mid +) = \frac{a}{a+b}$$

$$NPV = P(\overline{D} \mid -) = \frac{d}{c+d}$$

negative predicted value

$$Sens = P(+ \mid D) = \frac{a}{a+c} \qquad Spec = P(- \mid \overline{D}) = \frac{d}{b+d}$$

Diagnostic Testing

- Marilyn's Example

Sens = 0.95
Spec = 0.95

	Disease +	Disease -	Total
Test +	***48***	***47***	***95***
Test -	***2***	***903***	***905***
Total	***50***	***950***	***1000***

PPV = 51%
NPV = 99%

P(D) = 0.05

Diagnostic Testing

- Marilyn's Example

Sens = 0.95 = 190/200
Spec = 0.95 = 760/800

	Disease +	Disease -	Total
Test +	***190***	***40***	***230***
Test -	***10***	***760***	***770***
Total	***200***	***800***	***1000***

PPV = 83%

NPV = 99%

P(D) = 0.20

Point: PPV depends on prior probability of disease in the population

Diagnostic Testing & Bayes Theorem

- P(D): prior distribution, that is, the prevalence of disease in the population
- P(+|D): likelihood function, that is, the probability of positive test given that the person has the disease (specificity)
- P(D|+): positive predicted value, that is, the probability that, given that the test is positive, the person has the disease (posterior probability)

$$P(D \mid +) = \frac{\overset{0.95}{P(+ \mid D)}\,\overset{0.20}{P(D)}}{\underset{0.23}{P(+)}}$$

$$P(+) = P(+ \mid D)P(D) + P(+ \mid \overline{D})P(\overline{D})$$

Bayes & MLMs…

A Two-stage normal normal model

cluster-specific random effect

$$y_{ij} = \theta_j + \varepsilon_{ij}$$

cluster

unit within the cluster

$$i = 1, \ldots, n_j, \; j = 1, .., J$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

within cluster variance

$$\theta_j \sim N(\theta, \tau^2)$$

overall mean

between clusters variance of the “true” cluster specific means (heterogeneity parameter)

Terminology

- **Two stage normal normal model**
- **Variance component model**
- **Two-way random effects ANOVA**
- **Hierarchical model with a random intercept and no covariates**

Are all the same thing!

Testing in Schools

- Goldstein and Spiegelhalter JRSS (1996)
- Goal: differentiate between `good' and `bad‘ schools
- Outcome: Standardized Test Scores
- Sample: 1978 students from 38 schools
  - MLM: students (obs) within schools (cluster)
- Possible Analyses:
  1. Calculate each school’s observed average score (approach A)
  2. Calculate an overall average for all schools (approach B)
  3. Borrow strength across schools to improve individual school estimates (Approach C)

Shrinkage estimation

- Goal: estimate the school-specific average score θ_j
- Two simple approaches:
  - A) No shrinkage $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$
  - B) Total shrinkage $\bar{y} = \frac{\sum_{j=1}^{J} \frac{n_j}{\sigma^2} \bar{y}_j}{\sum_{j=1}^{J} \frac{n_j}{\sigma^2}}$

ANOVA and the F test

- To decide which estimate to use, a traditional approach is to perform a classic F test for differences among means
- if the group-means appear significant variable then use A
- If the variance between groups is not significant greater that what could be explained by individual variability within groups, then use B

Shrinkage Estimation: Approach C

- We are not forced to choose between A and B
- An alternative is to use the a weighted combination between A and B

$$\hat{\theta}_j = \lambda_j \bar{y}_j + (1 - \lambda_j)\bar{y}$$

Empirical Bayes estimate

$$\lambda_j = \frac{\tau^2}{\tau^2 + \sigma_j^2}; \sigma_j^2 = \sigma^2 / n_j$$

Shrinkage estimation

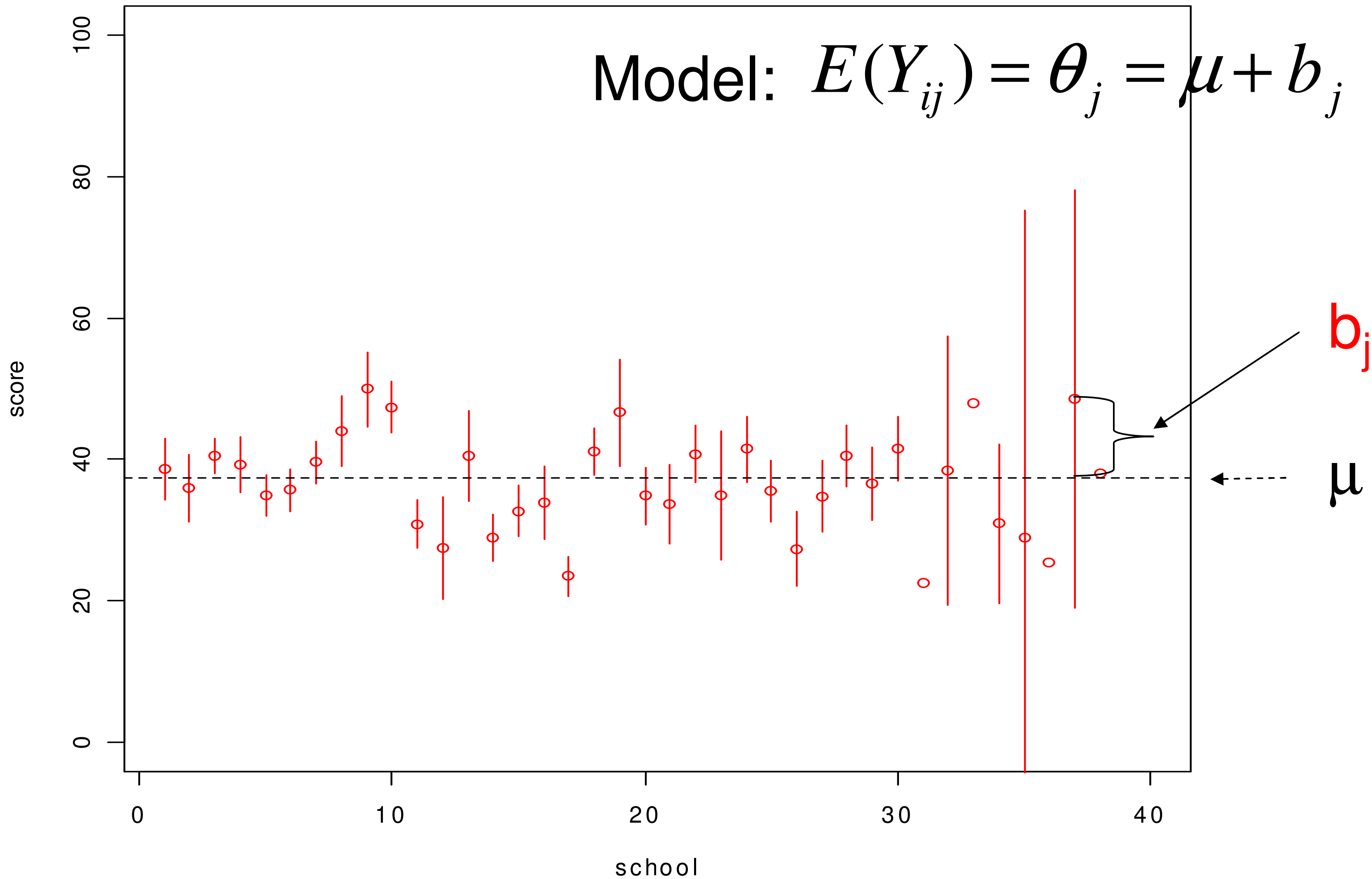
- Approach C reduces to approach A (no pooling) when the shrinkage factor is equal to 1, that is, when the variance between groups is very large
- Approach C reduces to approach B, (complete pooling) when the shrinkage factor is equal to 0, that is, when the variance between group is close to be zero

A Case study: Testing in Schools

- Why borrow across schools?
- Median # of students per school: 48, Range: 1-198
- Suppose small school (N=3) has: 90, 90,10 (avg=63)
- Difficult to say, small N $\Rightarrow$ highly variable estimates
- For larger schools we have good estimates, for smaller schools we may be able to borrow information from other schools to obtain more accurate estimates

Testing in Schools

Mean Scores & C.I.s for Individual Schools



Fixed and Random Effects

- Standard regression models: $\varepsilon_{ij} \sim N(0,\sigma^2)$

$Y_{ij} = \mu + \varepsilon_{ij}$ $E(Y_{ij})=\mu$ (overall average)

$Y_{ij} = \mu + b^*_j + \varepsilon_{ij}$ $E(Y_{ij})=\theta_j$ (observed school avgs)

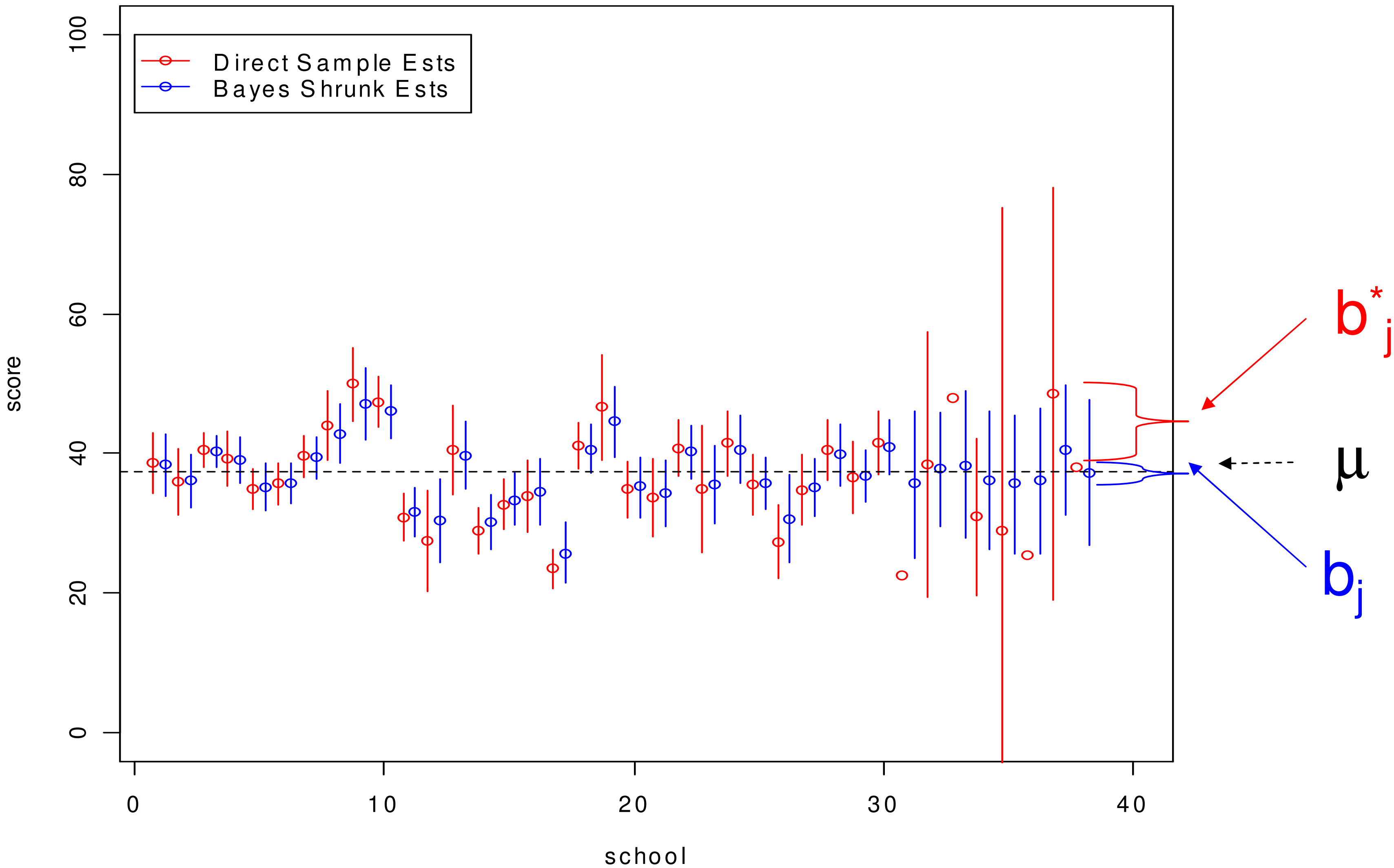
Fixed Effects

- A random effects model:

$Y_{ij} \mid b_j = \mu + b_j + \varepsilon_{ij}$, where: $b_j \sim N(0,\tau^2)$

Random Effects:

Testing in Schools: Shrinkage Plot



Some Bayes Concepts

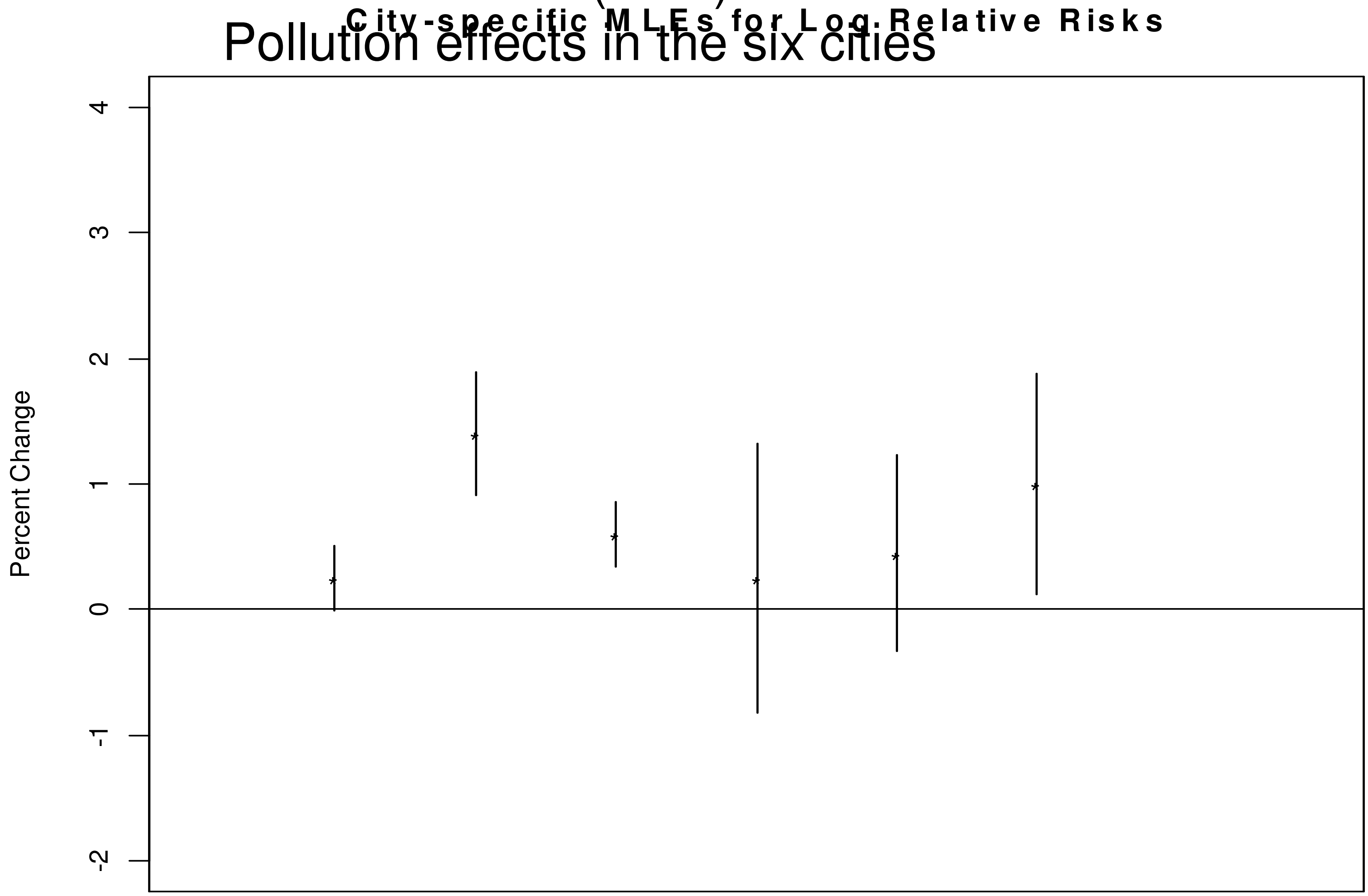
- Frequentist: Parameters are "the truth"
- Bayesian: Parameters have a distribution
- "Borrow Strength" from other observations
- "Shrink Estimates" towards overall averages
- Compromise between model & data
- Incorporate prior/other information in estimates
- Account for other sources of uncertainty

Relative Risks for Six Largest Cities

City	RR Estimate (% per 10 micrograms/ml y_j	Statistical Standard Error σ_j	Statistical Variance σ_j^2
Los Angeles	0.25	0.13	.0169
New York	1.4	0.25	.0625
Chicago	0.60	0.13	.0169
Dallas/Ft Worth	0.25	0.55	.3025
Houston	0.45	0.40	.1600
San Diego	1.0	0.45	.2025

Approximate values read from graph in Daniels, et al. 2000. AJE

Point estimates (MLE) and 95% CI of the air Pollution effects in the six cities



Two-stage normal normal model

RR estimate in city j

True RR in city j

$$y_j = \theta_j + \varepsilon_j$$

Within city statistical Uncertainty (known)

$$\varepsilon_j \sim N(0, \sigma_j^2)$$

$$\theta_j \sim N(\theta, \tau^2)$$

Heterogeneity across cities in the true RR

Two sources of variance

$$y_j = \theta_j + \varepsilon_j$$

$$\theta_j = \mu + b_j$$

$$y_j = \mu + b_j + \varepsilon_j$$

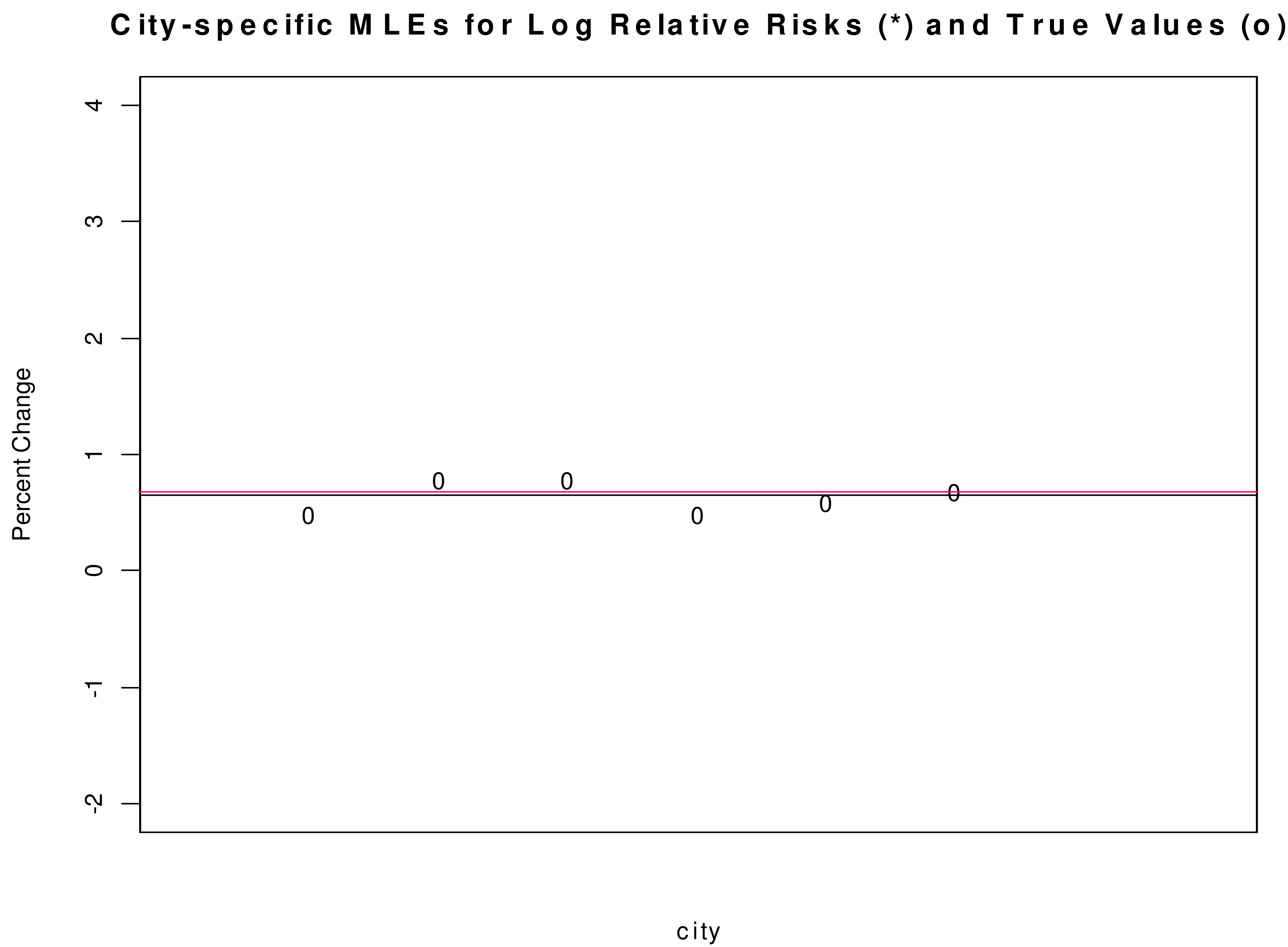
Total variance

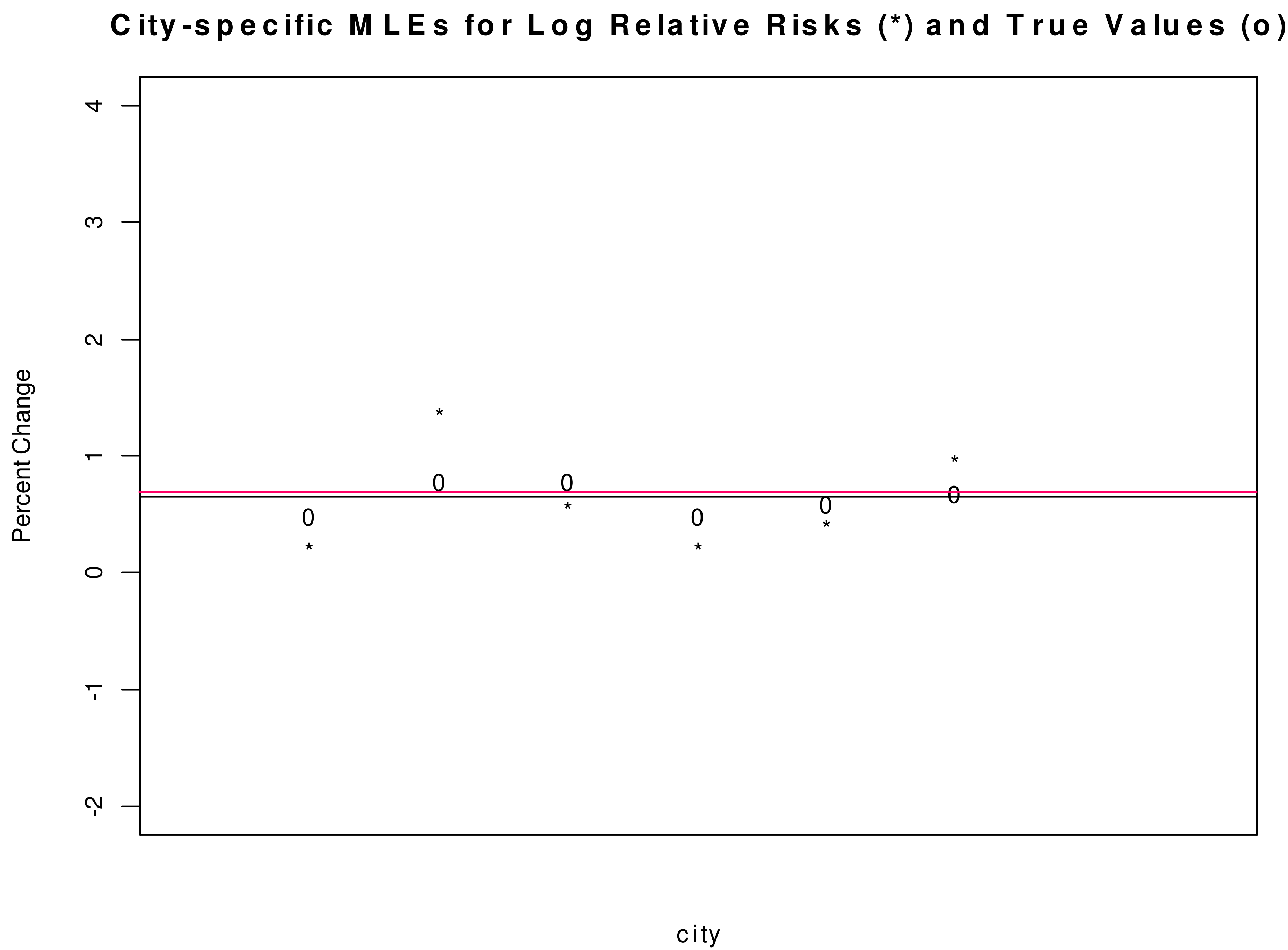
Variance between

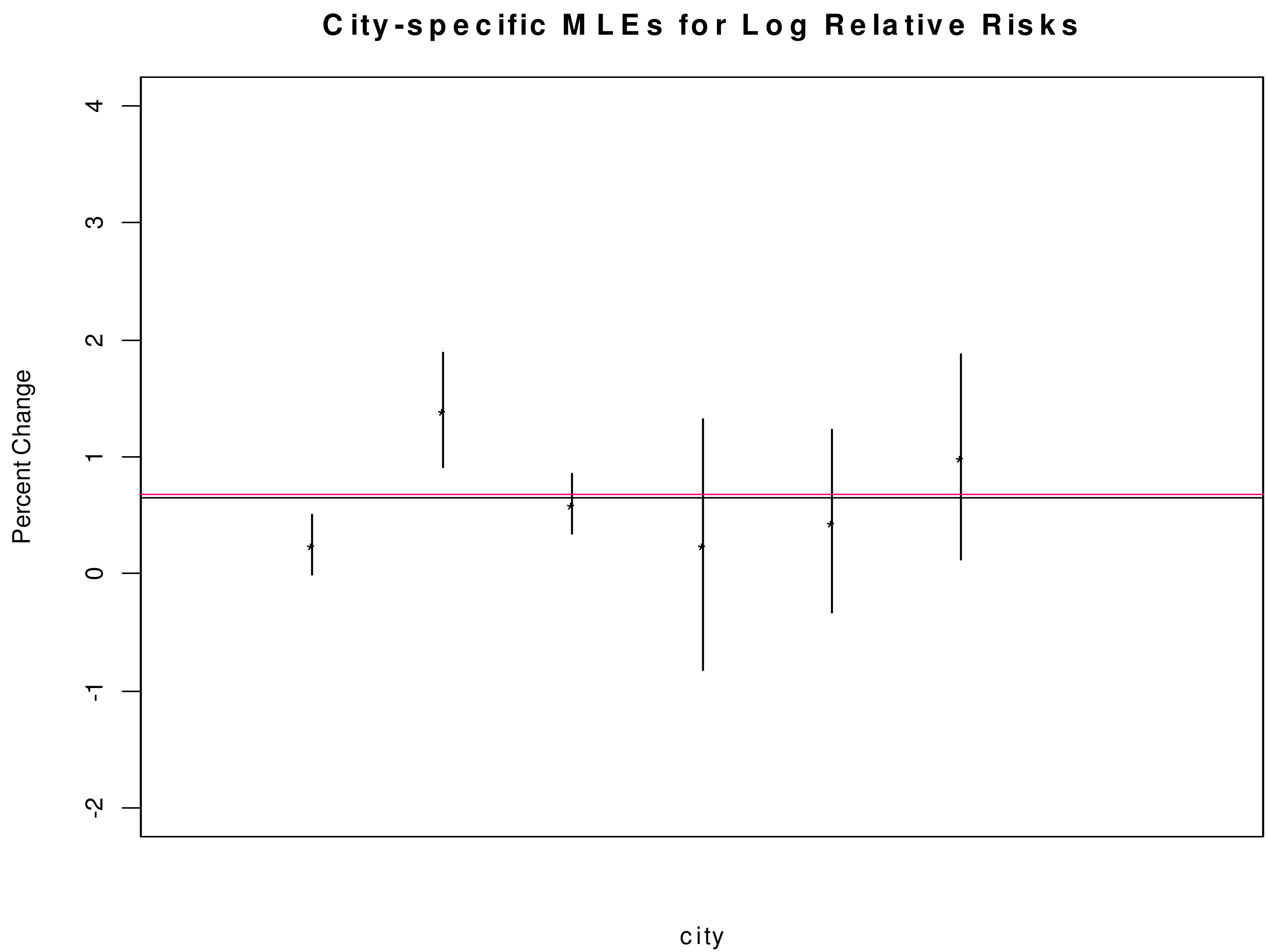
Variance within

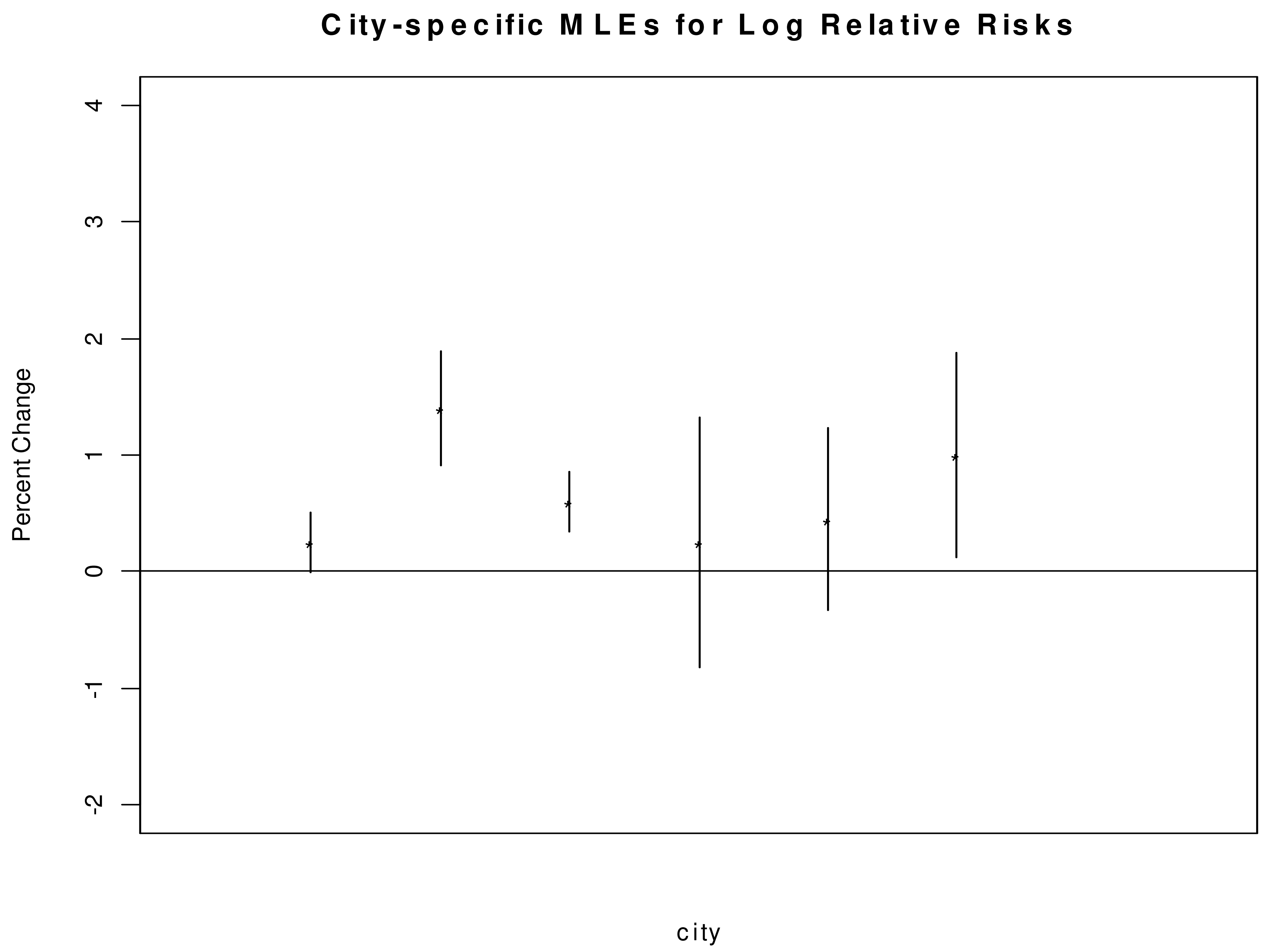
$$V(y_j) = V(b_j) + V(\varepsilon_j) = \tau^2 + \sigma_j^2$$

$$\boxed{\lambda_j = \frac{\tau^2}{\tau^2 + \sigma_j^2} \quad \text{shrinkage factor}}$$


City-specific MLEs for Log Relative Risks (*) and True Values (o)
Percent Change
-2
-1
0
1
2
3
4
0
0
0
0
0
0
city


City-specific MLEs for Log Relative Risks (*) and True Values (o)
Percent Change
4
3
2
1
0
-1
-2
city


City-specific MLEs for Log Relative Risks
Percent Change
4
3
2
1
0
-1
-2
city


City-specific MLEs for Log Relative Risks
Percent Change
city
4
3
2
1
0
-1
-2

Estimating Overall Mean

- Idea: give more weight to more precise values

- Specifically, weight estimates inversely proportional to their variances

Estimating the overall mean (Der Simonian and Laird, Controlled Clinical Trial 1986)

"crude" estimate of the heterogeneity parameter

$$\hat{\tau}^2 = \frac{1}{J-1}\sum_j (y_j - \bar{y})^2 - \frac{1}{J}\sum_j \sigma_j^2$$

$$h_j = \frac{1}{\sigma_j^2 + \tau^2}; w_j = h_j / \sum_j h_j$$

$$\hat{\mu} = \frac{\sum_j w_j y_j}{\sum_j w_j}; \boxed{V(\hat{\mu}) = \frac{1}{\sum_j w_j}}$$

Calculations for Empirical Bayes Estimates *(redo this using the "meta" function in stata..)*

City	RR	Stat Var	Total Var (TV)	1/TV	wj
LA	0.25	.0169	.0994	10.1	.27
NYC	1.4	.0625	.145	6.9	.18
Chi	0.60	.0169	.0994	10.1	.27
Dal	0.25	.3025	.385	2.6	.07
Hou	0.45	.160	,243	4.1	.11
SD	1.0	.2025	.285	3.5	.09
Over-all	0.65			37.3	1.00

overall = .27 0.25 + .18*1.4 + .27*0.60 + .07*0.25 + .11*0.45 + 0.9*1.0 = 0.65*

Software in R

```
yj <-c(0.25,1.4,0.60,0.25,0.45,1.0)

sigmaj <- c(0.13,0.25,0.13,0.55,0.40,0.45)

tausq <- var(yj) - mean(sigmaj^2)

TV <- sigmaj^2 + tausq

tmp<- 1/TV

ww <- tmp/sum(tmp)

v.muhat <- sum(ww)^{-1}

muhat <- v.muhat*sum(yj*ww)
```

Two Extremes

- Natural variance >> Statistical variance
  - Weights wj approximately constant
  - Use ordinary mean of estimates regardless of their relative precision

- Statistical variance >> Natural variance
  - Weight each estimator inversely proportional to its statistical variance

Empirical Bayes Estimation

$$\hat{\theta}_j = \lambda_j \bar{y}_j + (1 - \lambda_j)\boxed{\hat{\mu}}$$

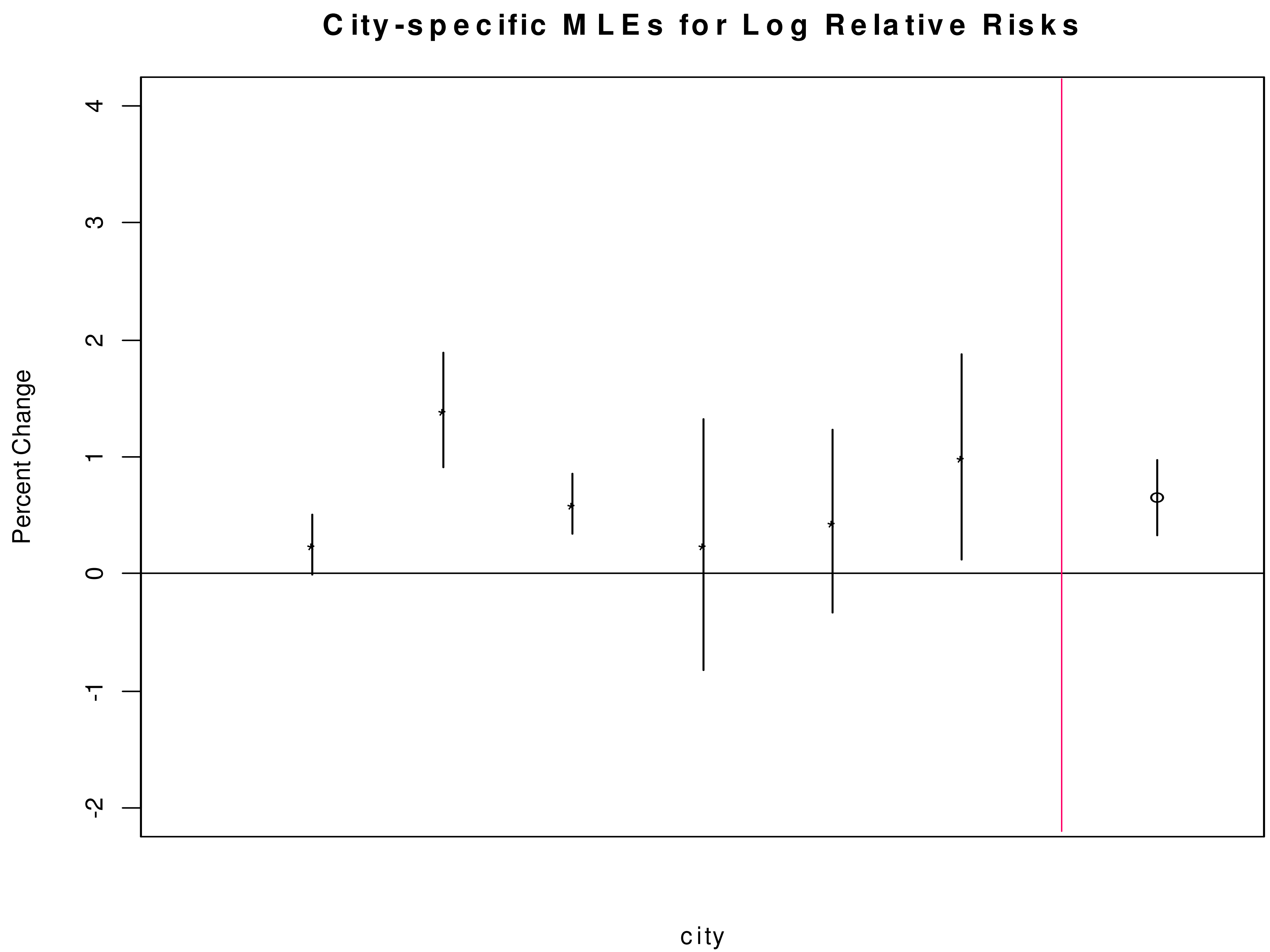
$$\lambda_j = \frac{\tau^2}{\tau^2 + \sigma_j^2}$$

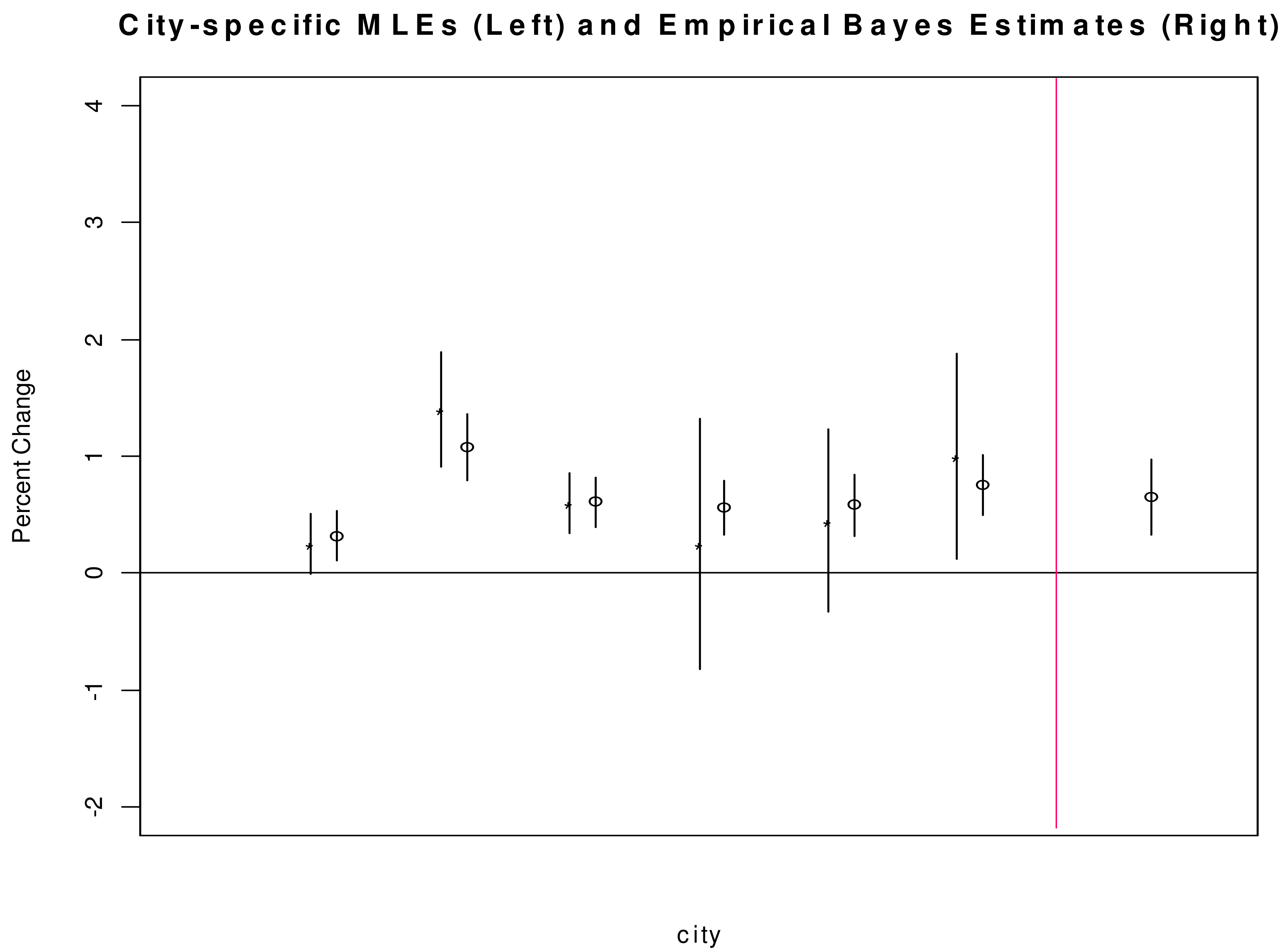
Calculations for Empirical Bayes Estimates

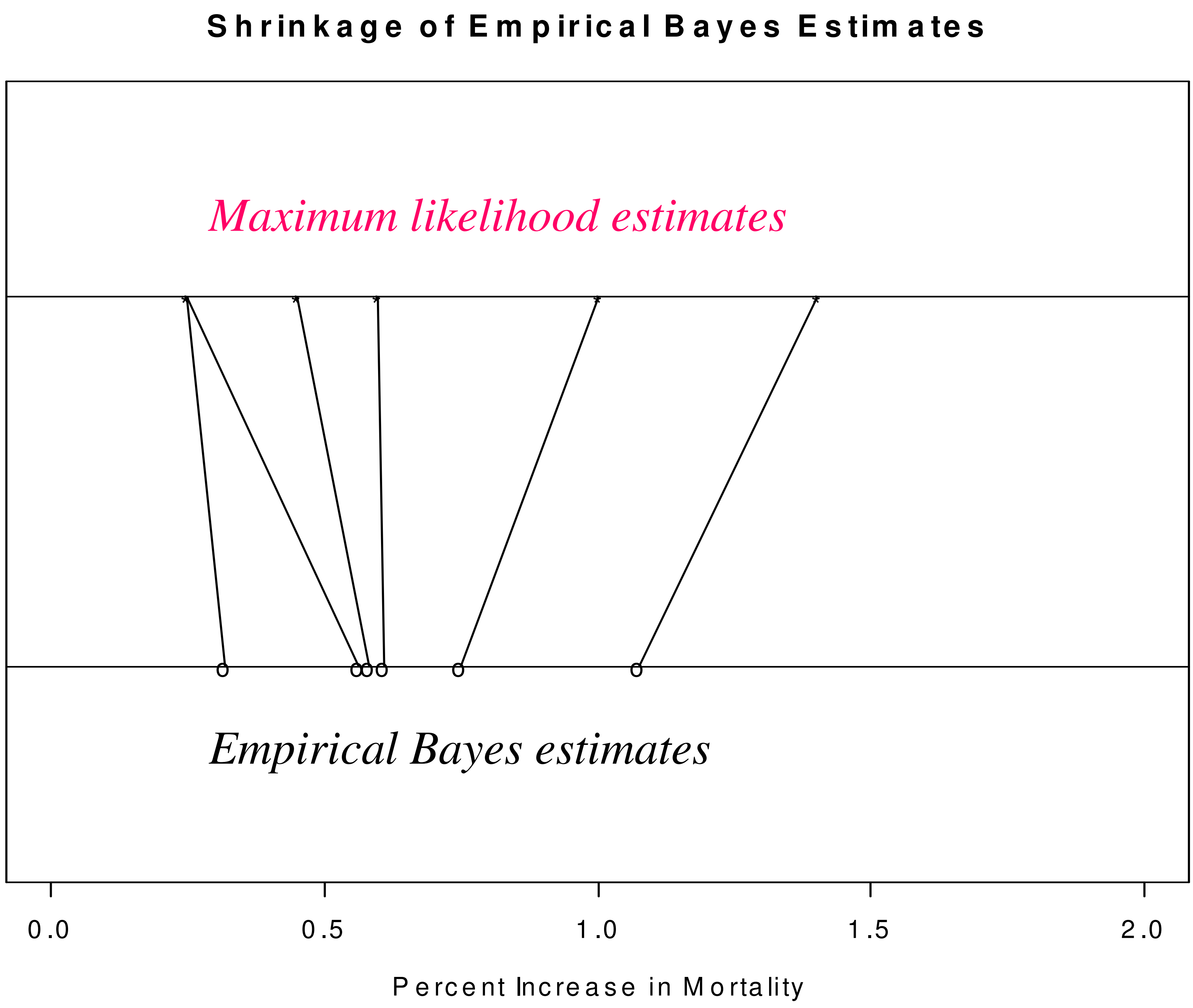
0.83x0.25 + (1-0.83)x0.65

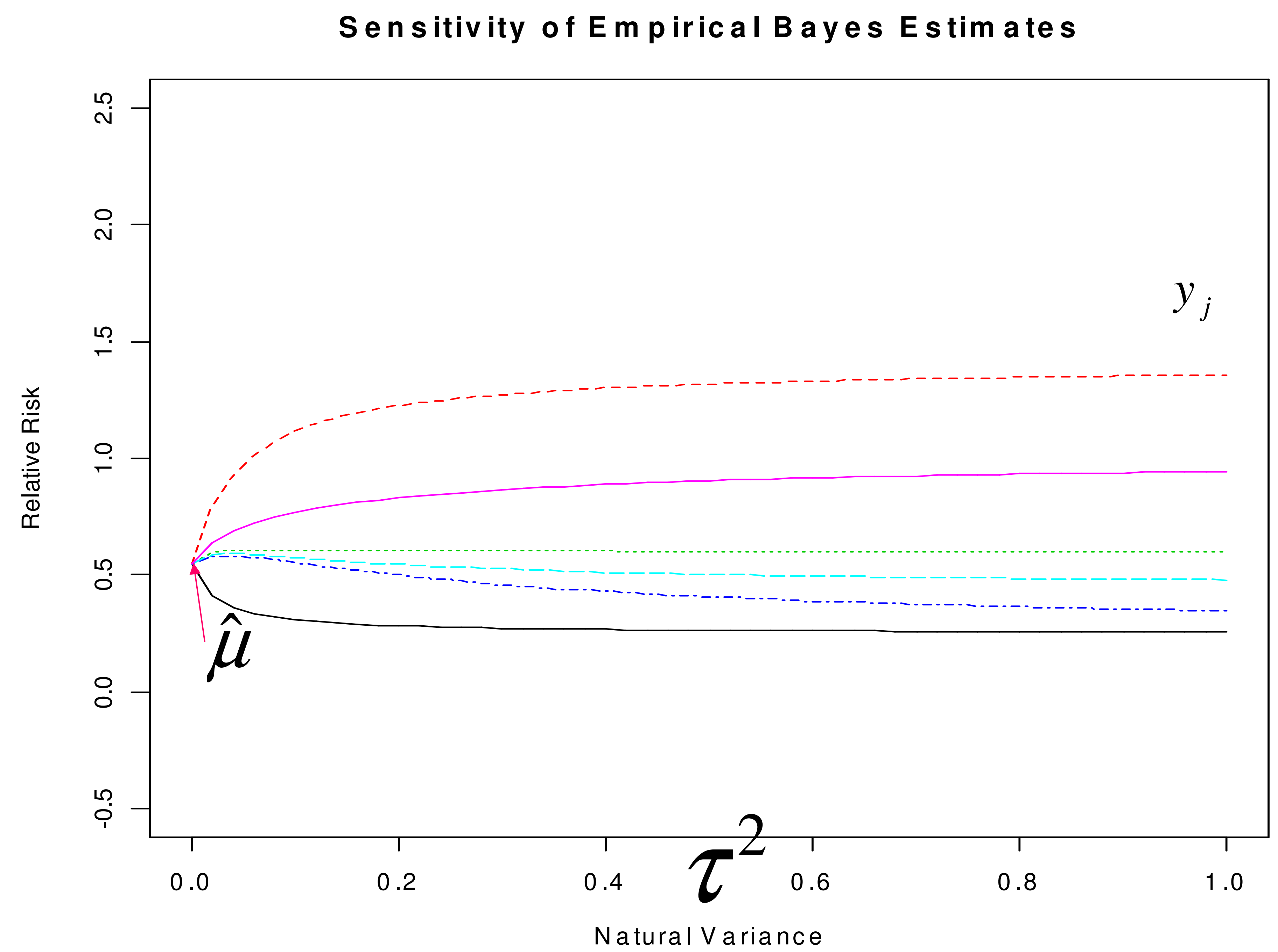
City	RR	Stat Var	Total Var	1/TV	w_j	λ_j	$\hat{\theta}_j$	$se(\hat{\theta}_j)$
LA	0.25	.0169	.0994	10.1	.27	.83	0.32	0.12
NYC	1.4	.0625	.145	6.9	.18	.57	1.1	0.19
Chi	0.60	.0169	.0994	10.1	.27	.83	0.61	0.12
Dal	0.25	.3025	.385	2.6	.07	.21	0.56	0.25
Hou	0.45	.160	,243	4.1	.11	.34	0.58	0.23
SD	1.0	.2025	.285	3.5	.09	.29	0.75	0.24
Over-all	0.65	1/37.3= 0.027		37.3	1.00		0.65	0.16

$$se(\hat{\theta}_j) = \left(\frac{1}{\sigma_j^2} + \frac{1}{\tau^2} \right)^{-1}$$


City-specific MLEs for Log Relative Risks
Percent Change
4
3
2
1
0
-1
-2
city


City-specific MLEs (Left) and Empirical Bayes Estimates (Right)
Percent Change
city
4
3
2
1
0
-1
-2


Shrinkage of Empirical Bayes Estimates
Maximum likelihood estimates
Empirical Bayes estimates
0.0
0.5
1.0
1.5
2.0
Percent Increase in Mortality


Sensitivity of Empirical Bayes Estimates
Relative Risk
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
y_j
$\hat{\mu}$
τ^2
0.0
0.2
0.4
0.6
0.8
1.0
Natural Variance

Key Ideas

- Better to use data for all cities to estimate the relative risk for a particular city
  - Reduce variance by adding some bias
  - Smooth compromise between city specific estimates and overall mean
- Empirical-Bayes estimates depend on measure of natural variation
  - Assess sensitivity to estimate of NV (heterogeneity parameter τ^2)

Caveats

- Used simplistic methods to illustrate the key ideas:
  - Treated natural variance and overall estimate as known when calculating uncertainty in EB estimates
  - Assumed normal distribution or true relative risks
- Can do better using Markov Chain Monte Carlo methods – more to come

In Stata (see 1.4 and 1.6, also Lab 1)

- **xtreg** with the **mle** option
- **xtmixed:** *preferred for continuous outcomes*
- **gllamm**