Building and Fitting Random Effects Models

William S. Cleveland
Lorraine Denby
Chuanhai Liu

Bell Labs
Lucent Technologies
Murray Hill, NJ

URL: http://cm.bell-labs.com/stat/

Data

Customer satisfaction survey of supplier companies in information technology industry

Sample of customers
- from 19 companies
- in each of 9 quarters

Customers rate 9 attributes on 1 (poor) to 10 (excellent) scale

3505 people, no repeats (approximately)

Attributes

prod-qual product quality

over-qual overall quality of all company processes

delivery delivery of the product

cost both the cost of the product to the customer and the process of establishing the cost with the customer

features functionality and technological excellence of product

pre-sup providing product information before delivery to support customer processes

response overall responsiveness of the company to the customer

value value of the product relative to the cost of the product

service service provided by the sales team

Missing Data

No one rates all 9 attributes

Missing by design

- only purchasing agents rate `delivery` and
- only product and process designers rate `pre-sup`

Missing, but not by design

Attributes Rated	1	2	3	4	5	6	7	8
Pollings	5	9	26	64	150	350	1106	1795

$3505 \times 9 = 31545$ = number of observations if complete

25309 = actual number of observations

Counts by Company and Attribute

	DELIVERY	PRE-SUP	PROD-QUAL	SERVICE	COST	RESPONSE	VALUE	FEATURES	OVER-QUAL
NUT	15	15	25	29	28	32	33	33	32
GAS	17	25	26	34	40	43	42	44	46
HAM	13	35	38	43	42	47	46	50	50
PUB	32	32	56	59	61	67	68	70	71
KEY	35	41	49	70	72	76	77	77	76
CAB	25	74	65	85	88	104	107	113	111
EAR	28	74	75	92	94	102	106	106	107
BEE	38	76	87	114	119	121	128	131	132
MUG	58	71	106	116	120	132	136	138	137
JET	26	128	102	138	147	153	158	162	162
RUG	54	100	124	148	163	169	174	167	174
LOG	63	97	121	156	165	167	170	172	171
TOY	47	110	132	145	152	170	177	178	181
OAK	84	110	169	174	197	205	218	219	216
ACE	77	174	186	254	240	270	267	273	275
INN	83	187	180	255	260	283	287	287	285
SKI	112	189	250	282	300	321	334	328	335
DUO	121	292	309	408	389	420	420	431	430
FAN	132	303	320	400	403	431	455	465	463

Goal

Compare suppliers for each attribute

Cannot look at company in isolation because there is a large attribute effect

Raw Data

Histograms of responses concerning one attribute for a company

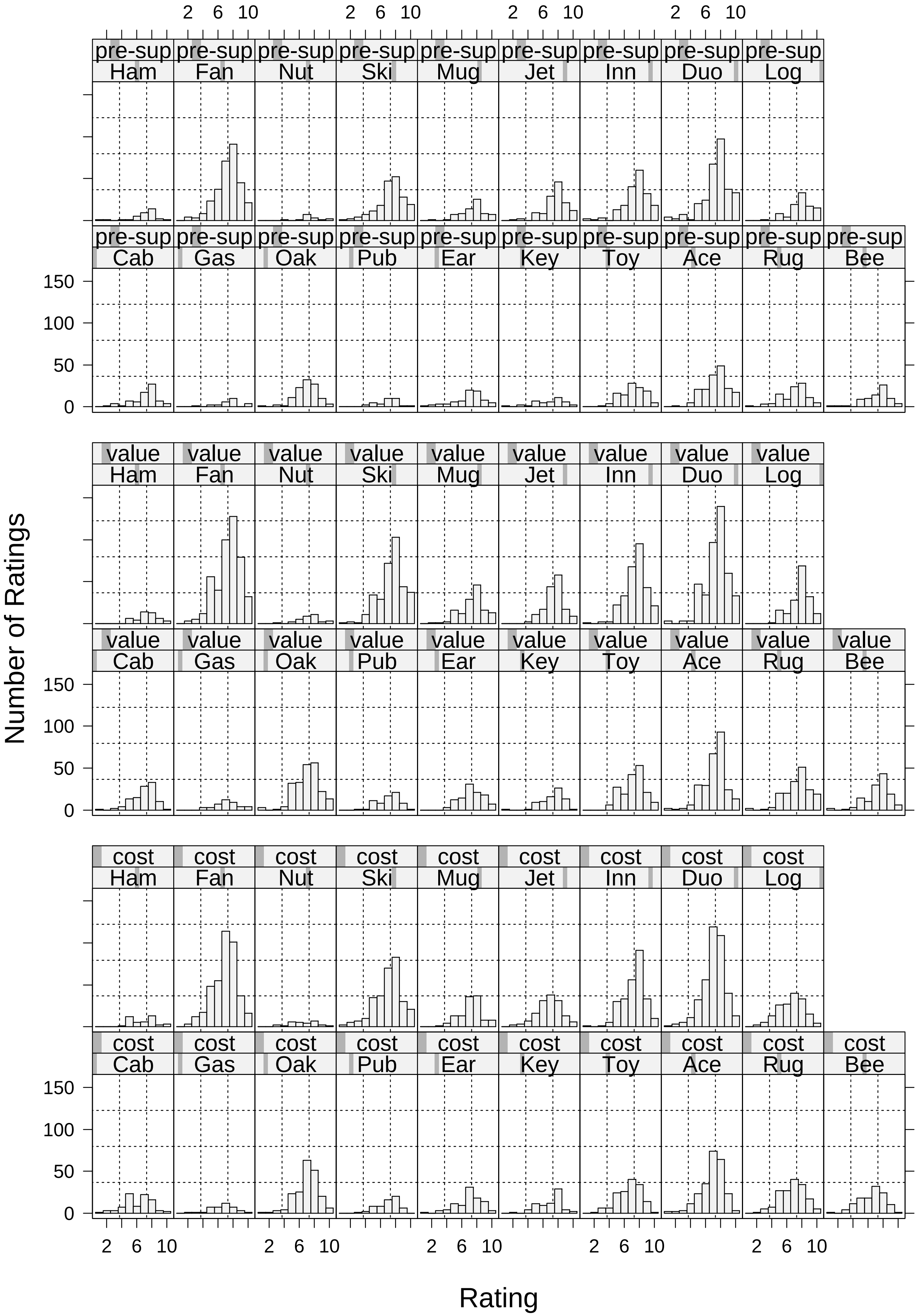
What do we see?

- Mode is 7 or 8
- 1–10 scale leaves more of scale below mode than above
- skewness to left
- no undue buildup at the high end of scale (10)

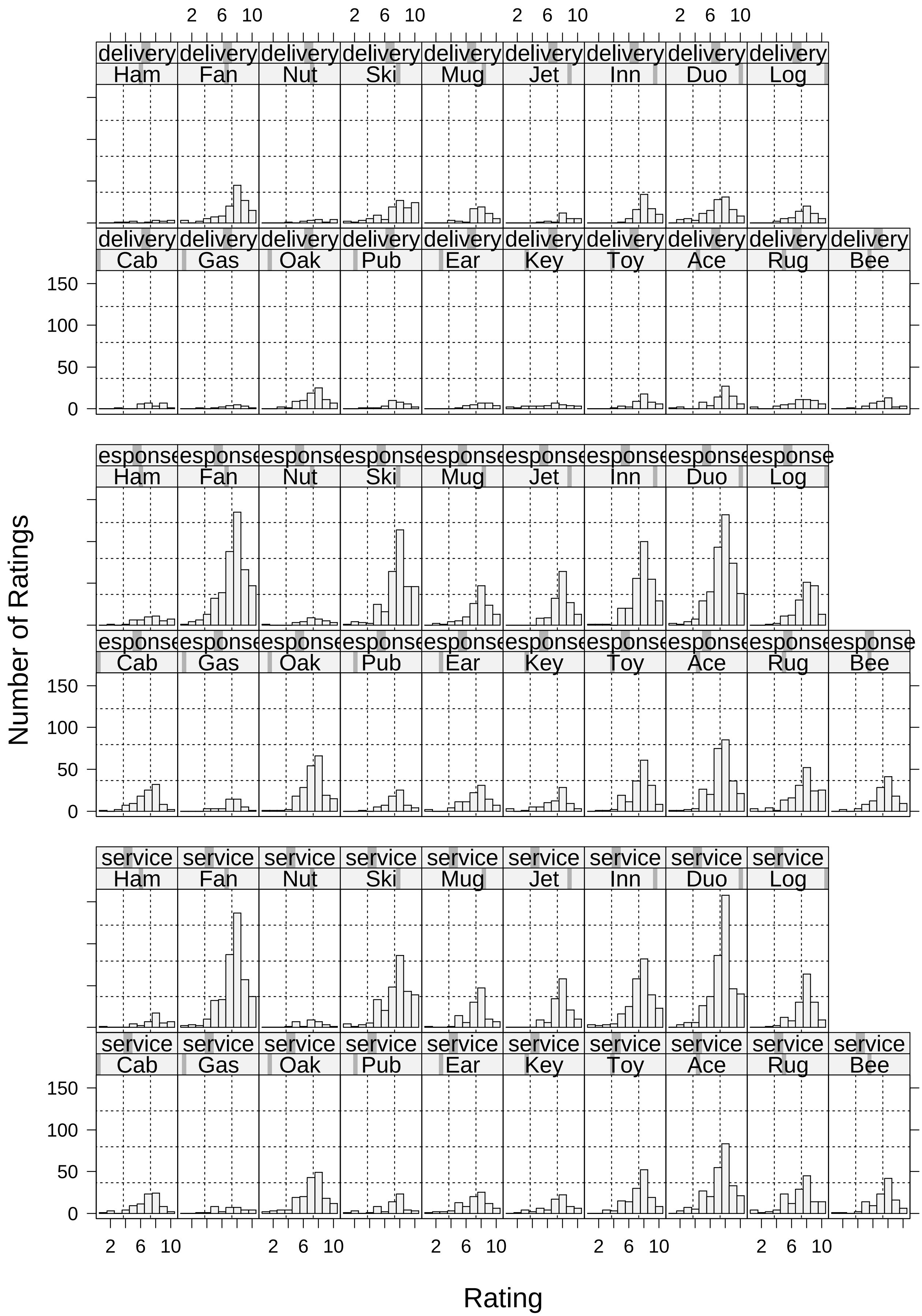
Conclusion

- Regularity of ratings and lack of buildup at 10 implies that continuous model is good first attempt
- Try transformation to make distribution symmetric

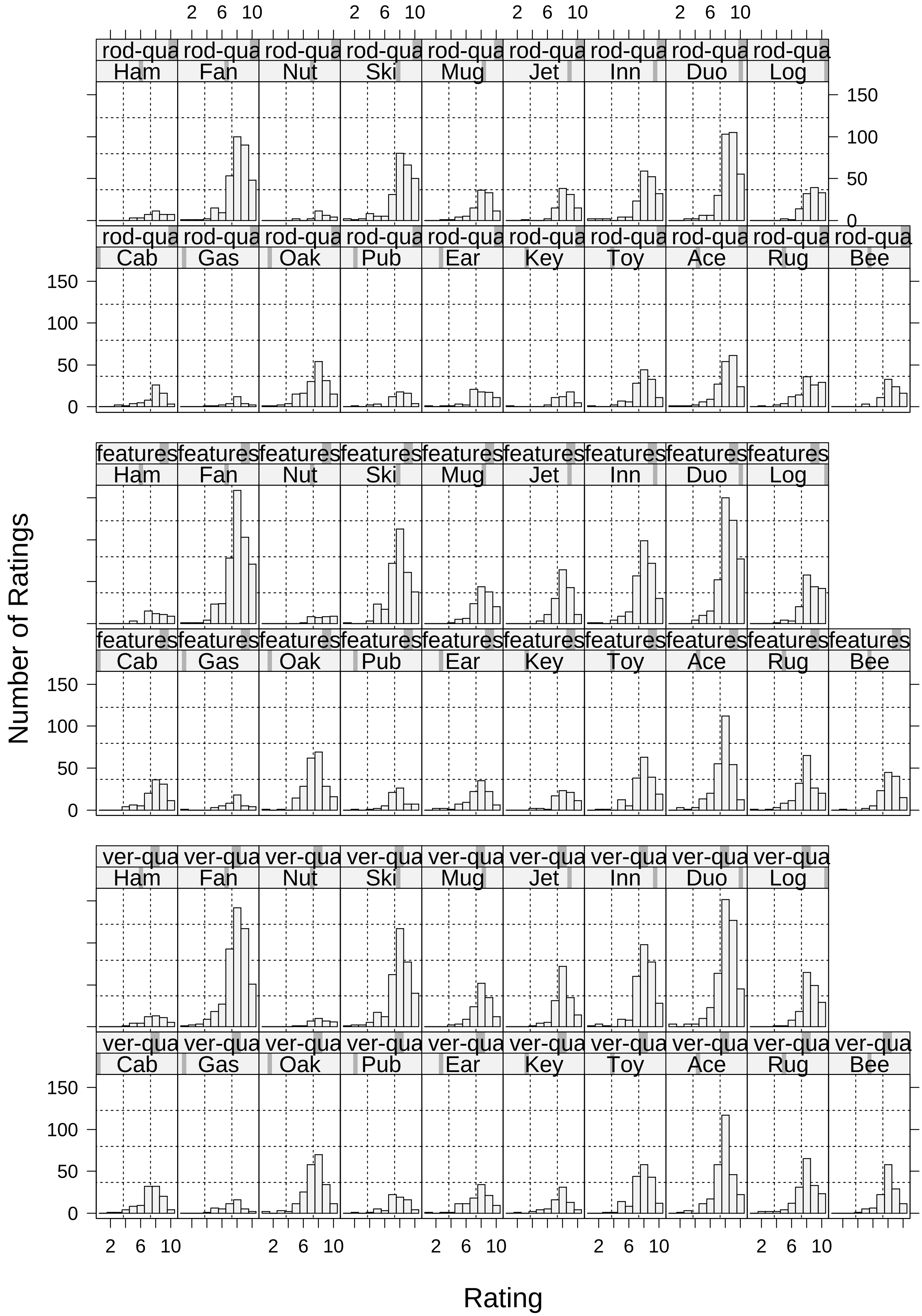
PLOT: Histogram of Raw Data (1)



PLOT: Histogram of Raw Data (2)



PLOT: Histogram of Raw Data (3)



Transformation to Symmetry

Considerations: Want data to range from 1 to 10

Family of transformations: $a \cdot f(11 - x) + b$ for various f where a and b were chosen to force range 1 to 10

Tool: Normal probability plot of ratings of a company for an attribute

This is first aid

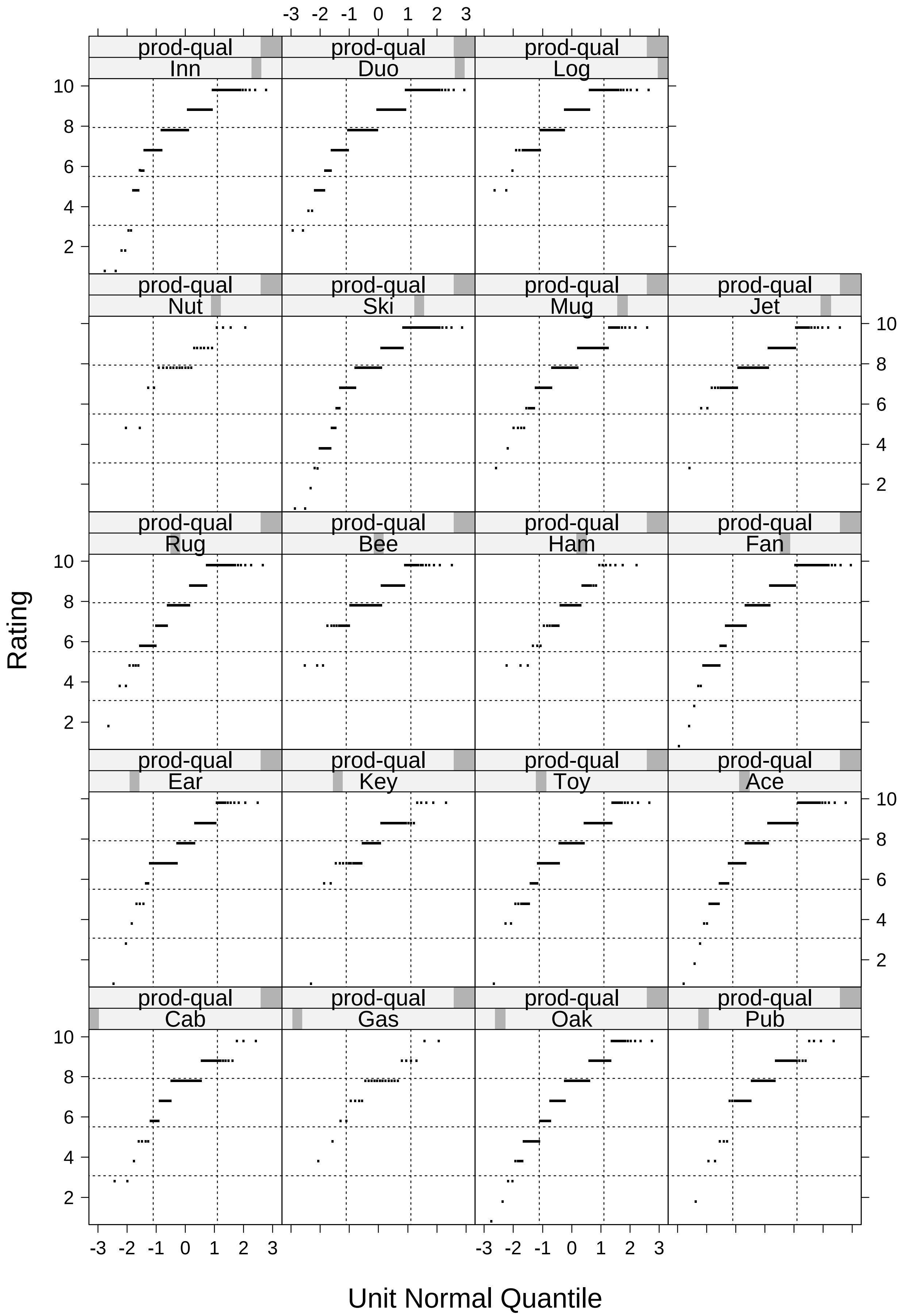
Normal QQ Plot Definition

Suppose F is a theoretical distribution

Plot $x_{(i)}$, the i^{th} order statistic, against $F^{-1}((i-.5)/n)$

Each panel contains ratings for one company for `prod-qual`

PLOT: Normal QQ



Normal QQ Plot Comments

Discreteness makes this plot difficult to read

Need to smooth out plot

Smooth QQ Plot Definition

Consider ratings for one company and one attribute

For $k = 1$ to 10 there are $n(k)$ values equal to k

Replace $n(k)$ values of k by

$$k - .5 + \frac{(i - .5)}{n(k)}$$

Visual Enhancements

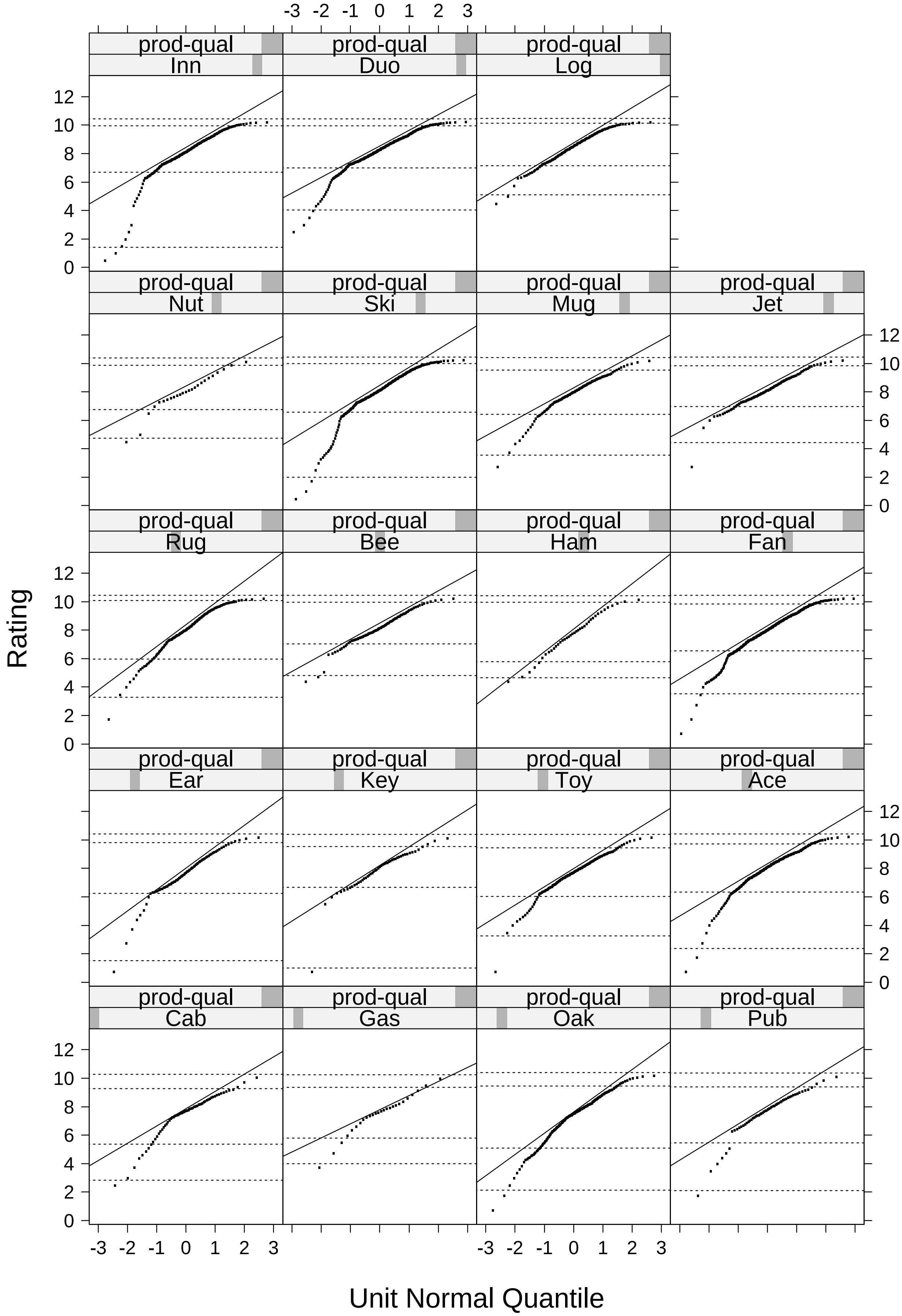
To judge how well the data agree with the reference distribution:

- Line through the lower quartile and the upper quartile points

To help determine where the majority of the data lie:

- Horizontal lines at the 0.01, 0.1, 0.9, and 0.99 quantiles

PLOT: Smooth QQ

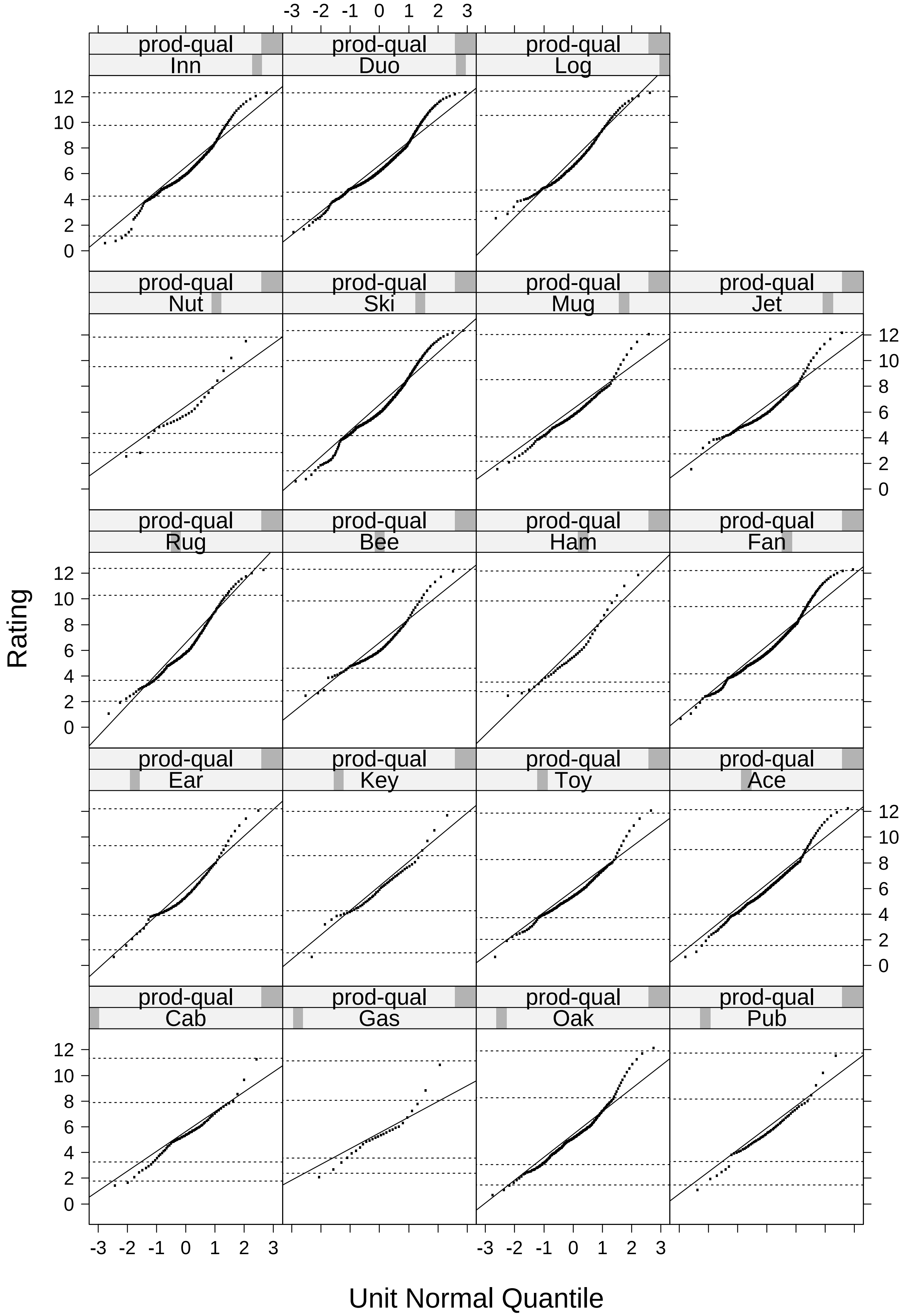


Smooth QQ Plot Comments

Need to push lower tail in and pull upper tail out

Try log!

PLOT: QQ for Log



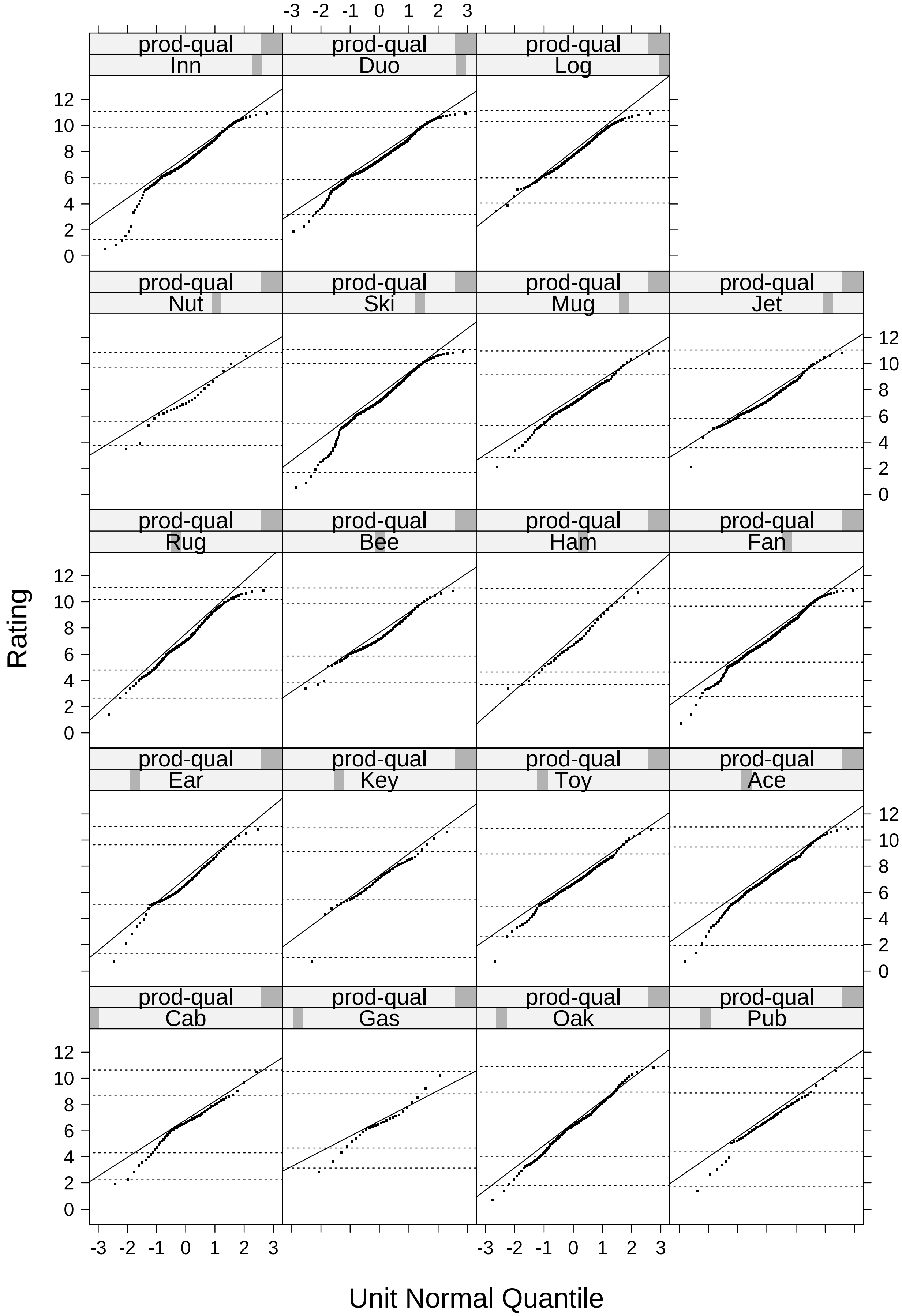
Log Transformation Comments

Log transformation is too drastic

Data now skewed to right

Try square root transformation

PLOT: QQ for Square Root



Square Root Transformation Comments

Data now symmetric

Longer tails than normal

Final transformation

$$-4.16\sqrt{(11-\text{rating})}+14.16$$

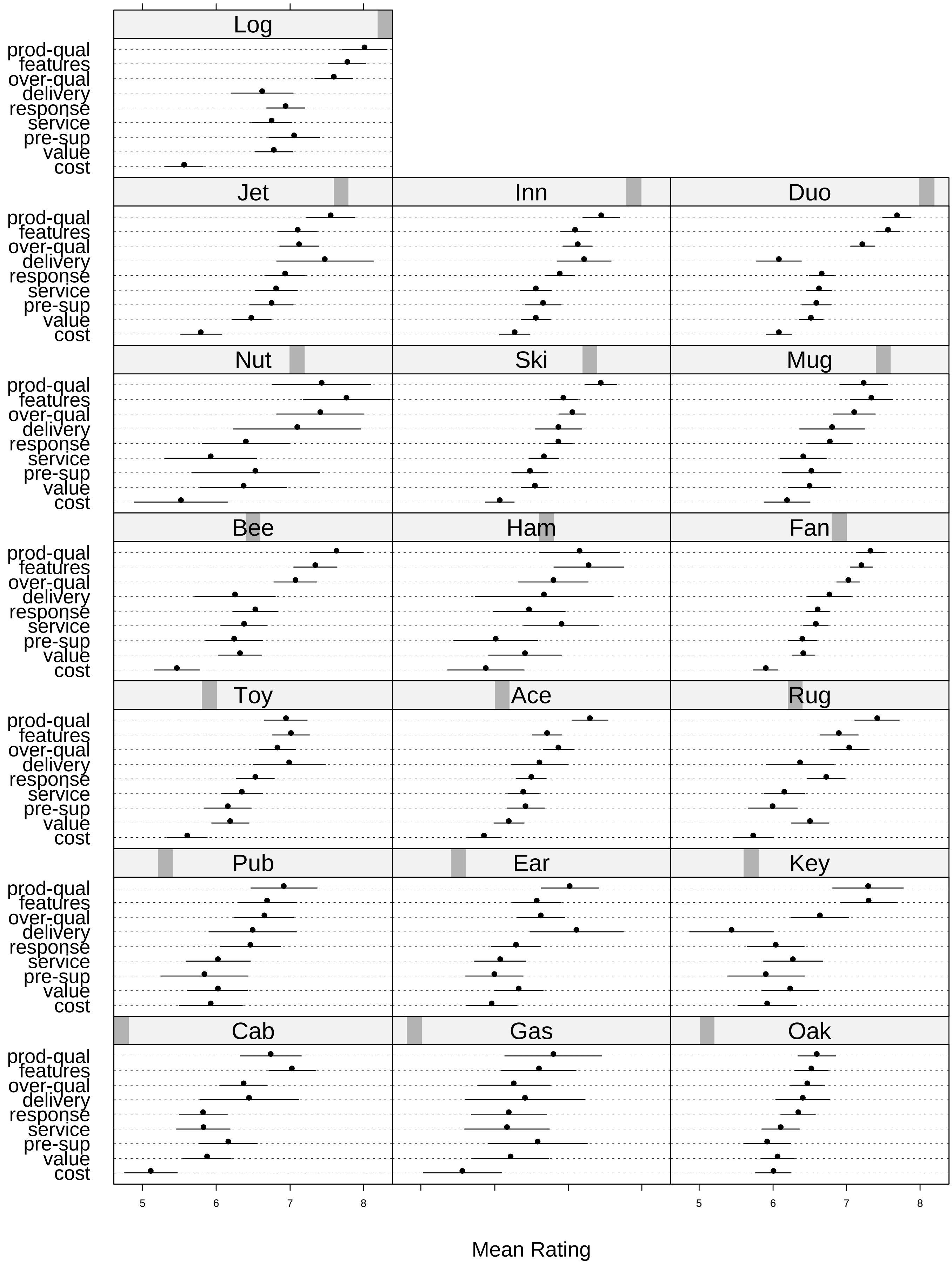
All modeling on transformed rating

Company/Attribute Means

Trellis chart shows means of ratings about an attribute for each company

Segment is $\pm 2\times$ sample standard deviation for company/attribute means

PLOT: Company/Attribute Means

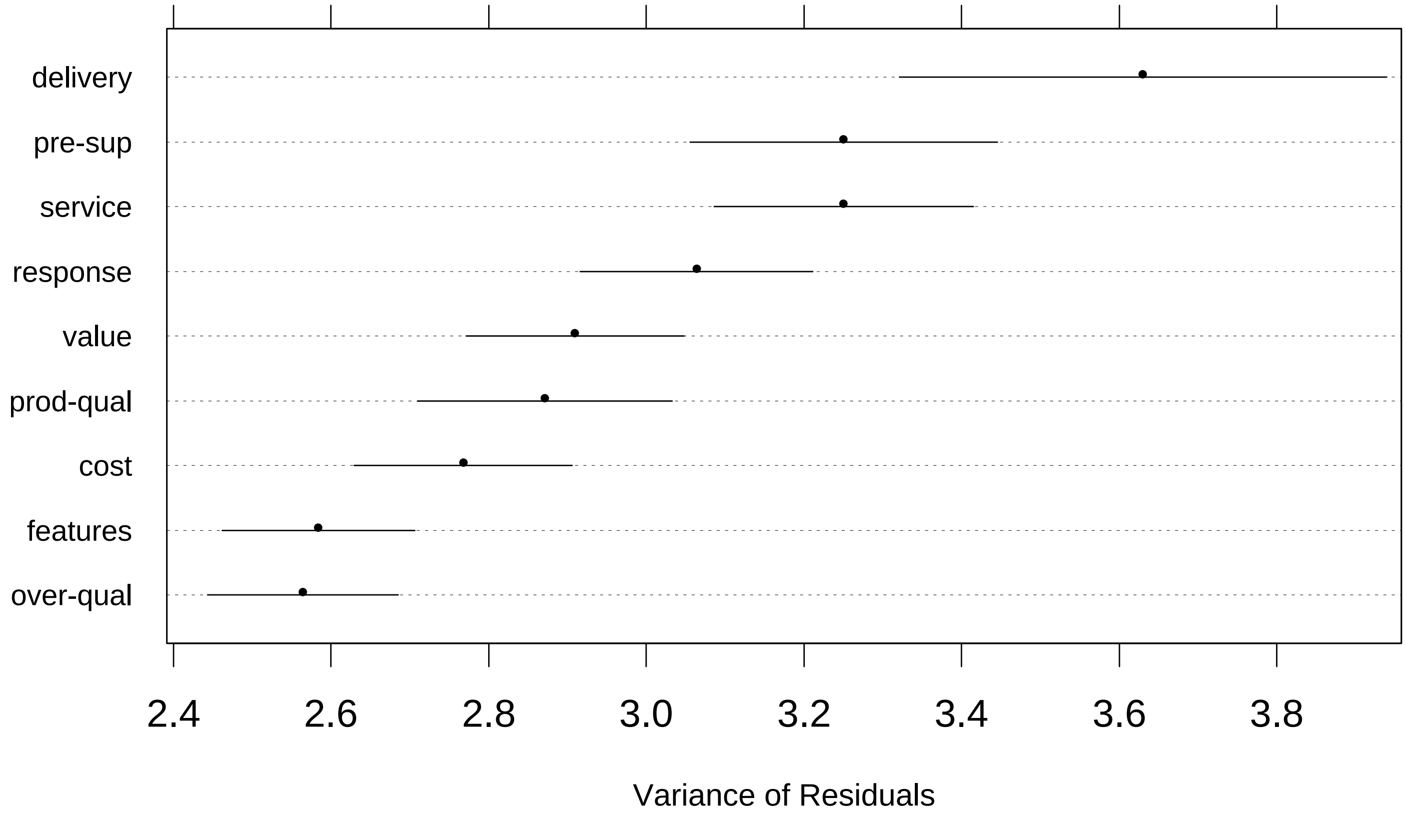


Attribute Variances

Residuals from the company/attribute means are used to calculate the variance associated with each attribute

Dot plot of attribute sample variances shows that there is a variance associated with attribute which we will treat as a fixed effect in our model

PLOT: Attribute Variances



Time

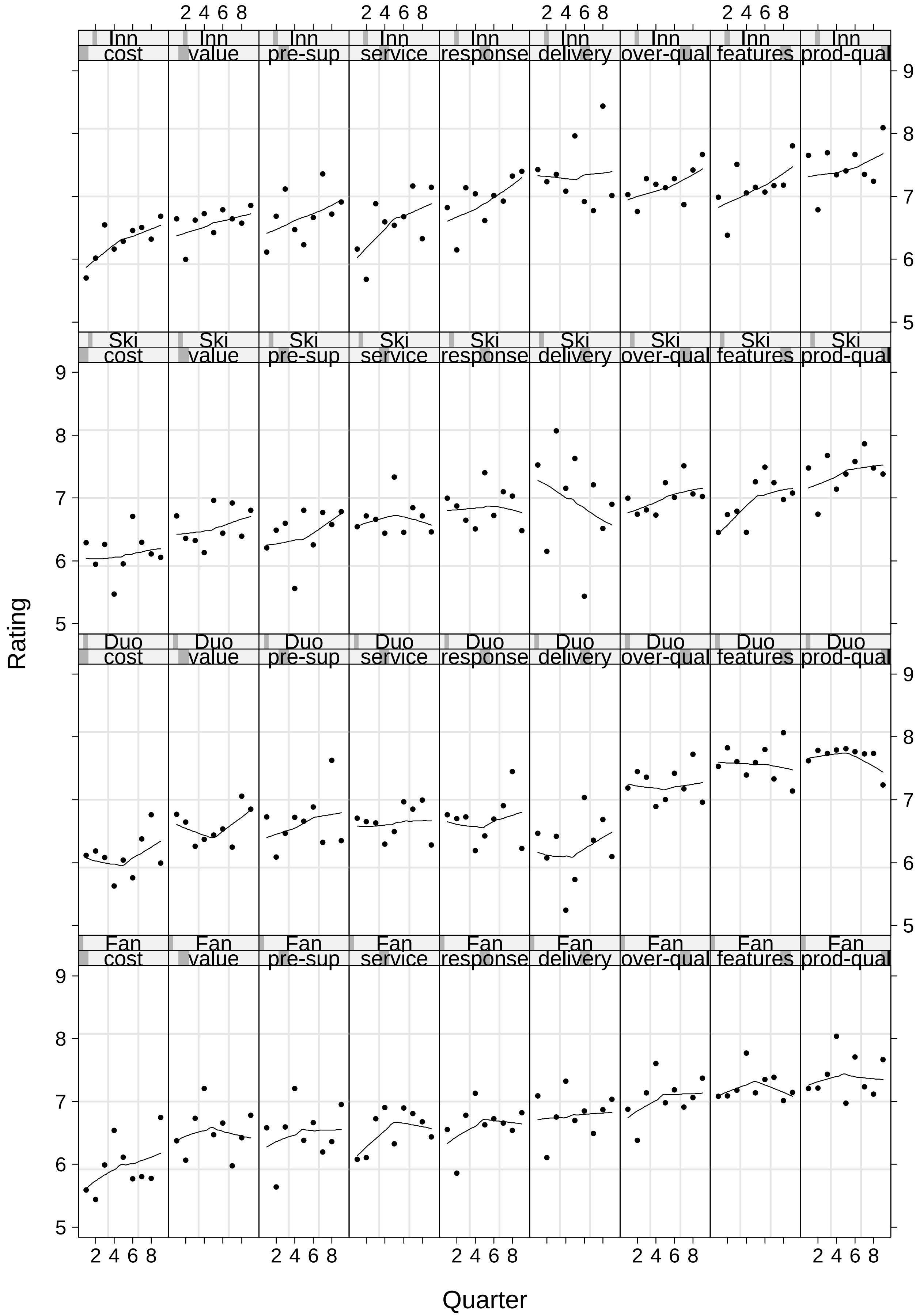
Each panel displays the means for a fixed value of a and s against q together with a loess curve fitted to the all of the ratings

Each plotting dot is one mean for the triple (a, s, q)

Quarterly values of the company/attribute means exhibit quarterly variation, but mainly random

Smooth curve shows possible mild upward or downward trend in some cases

PLOT: Quarterly Company/Attribute Means



What We Have Learned

Intermediate conclusions:

- Need to transform for symmetry
- Strong main effects for company/attributes
- Variance related to attribute
- Small time effect

Check for other sources of variation

- random location effects
- random scale effects

Respondents Effects

Random effects for survey data

People locate themselves at different places on the subjective scale

People use up different amounts of the subjective scale

Some Notation

$$r_{uj} = j^{\text{th}} \text{ transformed rating for respondent } u$$

$$\alpha_{asq} = \text{effect for supplier } s\text{, attribute } a \text{ and}$$
$$\text{quarter } q \text{ — treated as fixed effect}$$

$$\beta_u = \text{random location effect for respondent } u$$

$$\gamma_u = \text{random scale effect for respondent } u$$

$$\varepsilon_{uj} = \text{error terms, } N\left(0, \sigma^2_{a(u,j)}(\varepsilon)\right)$$

$$\sigma^2_{a(u,j)} = \text{error variance of attribute } a$$

Candidate Model

From initial exploration of the data and exogenous information

$$\begin{aligned}
r_{uj} &= \alpha_{asq} + \beta_u + \gamma_u \varepsilon_{uj} \\
E\beta_u &= 0 \\
E\varepsilon_{uj} &= 0 \\
E\varepsilon_{uj}^2 &= \sigma_{a(u,j)}^2 \\
E\gamma_u^2 &= 1 \\
Er_{ac} &= \alpha_{asq}
\end{aligned}$$

α_{asq} = effect of attribute a for supplier company s in quarter q

In terms of linear model:

$$\alpha_{asq} = \sum_{k=1}^{k_\alpha} \alpha_{(k)} v_{(k)uj}$$

where $k_\alpha = 9 \times 9 \times 19 = 1539$ and $v_{(k)uj}$ is a dummy variable that takes the value 1 when r_{uj} is a rating of attribute a for supplier s in quarter q and is zero otherwise

β_u = random location effect for respondent u

γ_u = random scale effect for respondent u

ε_{uj} = error term

Distributional Assumptions

ε_{uj}: is normal with variance related to attribute

β_u: will be determined from data

γ_u: will be determined from data

All variables are independent

Steps to Model Building

1) Estimate and adjust for fixed effects

- Location
- Scale

2) Use unit regression equations to estimate random location and scale effects

3) Use deconvolution methods to determine random effects distributions

4) Use Bayesian methods to fit the model

Adjust for Fixed Location Effects

Model:

$$r_{uj} = \alpha_{asq} + \beta_u + \gamma_u \varepsilon_{uj}$$

Estimate α_{asq} by least squares and assume estimates are true

Adjust for fixed location effects:

$$\begin{aligned} r^*_{uj} &= r_{uj} - \alpha_{asq} \\ &= \beta_u + \gamma_u \varepsilon_{uj} \end{aligned}$$

Adjust for Fixed Scale Effects

Estimate fixed scale effects and assume true,

$$\sigma_a^2 = \sigma_{a(u,j)}^2(\varepsilon)$$

Adjust for fixed scale:

$$\begin{aligned} r^*_{uj}/\sigma_a &= \beta_u/\sigma_a + \gamma_u\varepsilon_{uj}/\sigma_a \\ y_{uj} &= \beta_u x_{uj} + \gamma_u\zeta_{uj} \end{aligned}$$

where

$$x_{uj} = 1/\sigma_a$$

and

$$\zeta_{uj} = \varepsilon_{uj}/\sigma_a$$

are independent normal with mean 0 and variance 1

Unit Regression Method

Use to estimate β_u and γ_u

A regression for each u:

$$y_{uj} = \beta_u x_{uj} + \gamma_u \zeta_{uj}$$

For each u, a regression with 8 or fewer observations

Estimate β_u and γ_u by least-squares and compute residuals of various types

Estimate Random Location Effects

The least squares estimate of β_u:

$$\begin{aligned}
\hat{\beta}_u &= \frac{\sum_{j=1}^{n_u} y_{uj} x_{uj}}{\sum_{j=1}^{n_u} x_{uj}^2} \\
&= \frac{\sum_{j=1}^{n_u} (\beta_u x_{uj} + \gamma_u \zeta_{uj}) x_{uj}}{\sum_{j=1}^{n_u} x_{uj}^2} \\
&= \beta_u + \frac{\gamma_u \sum_{j=1}^{n_u} x_{uj} \zeta_{uj}}{\sum_{j=1}^{n_u} x_{uj}^2} \\
&= \beta_u + \gamma_u \xi_u
\end{aligned}$$

where

$$\xi_u \sim N(0, \sigma^2(\xi_u))$$

and

$$\sigma^2(\xi_u) = \frac{1}{\sum_{\ell=1}^{n_u} x_{u\ell}^2} = \frac{1}{\sum_{a(u,j) \in A(u)} \sigma_{a(u,j)}^{-2}}$$

since ζ_u has mean 0 and variance 1

Distribution of $\hat{\beta}_u$ complex:

- additive-multiplicative convolution of distributions of β_u, γ_u, and ξ_u

Residuals

$$
\begin{aligned}
\hat{\tau}_{uj} &= y_{uj} - \hat{\beta}_u x_{uj} \\
&= (\beta_u x_{uj} + \gamma_u \zeta_{uj}) - (\beta_u x_{uj} + \gamma_u \xi_{uj}) \\
&= \gamma_u \hat{\zeta}_{uj}
\end{aligned}
$$

where

$$
\hat{\zeta}_{uj} = \zeta_{uj} - \frac{\sum_{\ell=1}^{n_u} x_{u\ell} \zeta_{u\ell}}{\sum_{\ell=1}^{n_u} x_{u\ell}^2}
$$

The variance of $\hat{\zeta}_{uj}$:

$$
\bar{p}_{ujj} = 1 - \frac{x_{uj}^2}{\sum_{\ell=1}^{n_u} x_{u\ell}^2}
$$

where $\bar{P}_u$ is the projection matrix onto the space orthogonal to the space spanned by the columns of X_u

Standardized residuals

$$\hat{\psi}_{uj} = \frac{\hat{\tau}_{uj}}{\sqrt{\bar{p}_{ujj}}} = \gamma_u \frac{\hat{\zeta}_{uj}}{\sqrt{\bar{p}_{ujj}}} \sim \gamma_u N(0,1)$$

Distribution of $\hat{\psi}_{uj}$ is independent of β_u, depends on γ_u, and does not vary with n_u:

- multiplicative convolution of the γ_u and the standard normal

For example, suppose

$$\gamma_u^2 \sim IG(h,(h-1)^{-1}) \tag{1}$$

where

- IG is the inverse gamma distribution
- h is the shape
- the scale $(h-1)^{-1}$ is chosen so that $E(\gamma_u^2) = 1$
- $d = 2h$ is integer greater than 2

$$\hat{\psi}_{uj} \sim T\left(d, 0, \sqrt{1-2/d}\right)$$

where $T(d,0,\lambda)$ is a t distribution with d degreees of freedom, location 0, and scale λ

Residual Variance

$$
\begin{aligned}
s_u^2 &= \frac{\sum_{j=1}^{n_u} \hat{\tau}_{uj}^2}{n_u - 1} \\
&= \gamma_u^2 \frac{\sum_{j=1}^{n_u} \hat{\zeta}_{uj}^2}{n_u - 1} \\
&\sim \gamma_u^2 MSQ(n_u - 1)
\end{aligned}
$$

where $MSQ(d)$ is a *mean-square distribution* with d degrees of freedom — the distribution of a chi-square random variable divided by its degrees of freedom

Distribution of s_u^2 is independent of β_u, depends on γ_u, and varies with n_u

For example, if $\gamma_u^2 \sim IG$ then

$$s_u^2 \sim (1 - 2/d) F(n_u - 1, d)$$

where $F(f_1, f_2)$ is an F-distribution with f_1 and f_2 degrees of freedom

If no random scales then $\sigma^2(\gamma^2) = 0$, and

$$\sigma^2(s_u^2) = \frac{2}{n_u - 1}$$

The effect of $\sigma^2(\gamma^2) > 0$ is to inflate $\sigma^2(s_u^2)$ because

$$\sigma^2(s_u^2) = \frac{2}{n_u - 1} + \sigma^2(\gamma^2)\left(1 + \frac{2}{n_u - 1}\right)$$

Studentized Residuals

$$\hat{\phi}_{uj} = \frac{\hat{\Psi}_{uj}}{s_u} \sim \sqrt{n_u - 1} DSB(0.5, 0.5n_u - 1)$$

where s_u^2 is the residual variance for regression u

A random variable on [-1,1] has a double square beta distribution, $DSB(p,q)$, if its distribution is symmetric and its square is distributed $BETA(p,q)$

This distribution is independent of β_u and γ_u, but changes with n_u

Estimation of Fixed Location Effect

If no smoothness constraint:
α_{asq} would be sample means of r_{uj}, one for each combination of a, s, and q
$(9 \times 9 \times 19 = 1782$ of them)

Smoothness makes sense here:

- Use loess to estimate the α_{asq}
- For each combination of a and s, the ratings are smoothed as a function of q
- The loess smoothing parameter will be 1, and the local fitting will be linear

For fixed a and s, the estimate, $\hat{\alpha}_{asq}$, is the value of the loess curve at q

Adjusted r_{uj}:

$$r^*_{uj} = r_{uj} - \hat{\alpha}_{asq}$$

Estimation of Fixed Scale Effect

Remember:

$$
\begin{aligned}
r^*_{uj}/\sigma_a &= \beta_u/\sigma_a + \gamma_u \varepsilon_{uj}/\sigma_a \\
y_{uj} &= \beta_u x_{uj} + \gamma_u \zeta_{uj}
\end{aligned}
$$

So, for a particular a, $Var(\gamma_u \hat{\zeta}_{uj}) = E \sum_{(u,j)\in I(z)} \gamma_u^2 \hat{\zeta}_{uj}^2$

Also:

$$\hat{\zeta}_{uj} = \zeta_{uj} - \frac{\sum_{\ell=1}^{n_u} x_{u\ell}\zeta_{u\ell}}{\sum_{\ell=1}^{n_u} x_{u\ell}^2}$$

So, variance of $\hat{\zeta}_{uj}$ is

$$Var(\hat{\zeta}_{uj}) = \bar{p}_{ujj} = 1 - \frac{x_{uj}^2}{\sum_{\ell=1}^{n_u} x_{u\ell}^2}$$

and

$$x_{uj} = \sigma_{a(u,j)}^{-1}$$

For this data there are 9 attributes so 9 σ_a to be estimated

For each attribute a:

$$\sum_{(u,j)\in I(a)} \left(\frac{r^*_{uj} - \hat{\beta}_u}{\sigma_{a(u,j)}(\varepsilon)} \right)^2 = \sum_{(u,j)\in I(a)} \left(1 - \frac{\sigma_{a(u,j)}^{-2}(\varepsilon)}{\sum_{a(u,\ell)\in A(u)} \sigma_{a(u,\ell)}^{-2}} \right)$$

Estimates of σ_a

over-qual	0.79
response	1.17
prod-qual	1.19
pre-sup	1.42
value	1.46
features	1.61
delivery	1.88
cost	1.90
service	1.93

Where are We?

Data has been adjusted for fixed location and scale effects

Need to check assumptions and determine distributions of random effects

- Normality of the errors
- Distribution of γ_u
- Distribution of β_u

Checking Normality of Residuals ε_{uj}

Studentized residuals

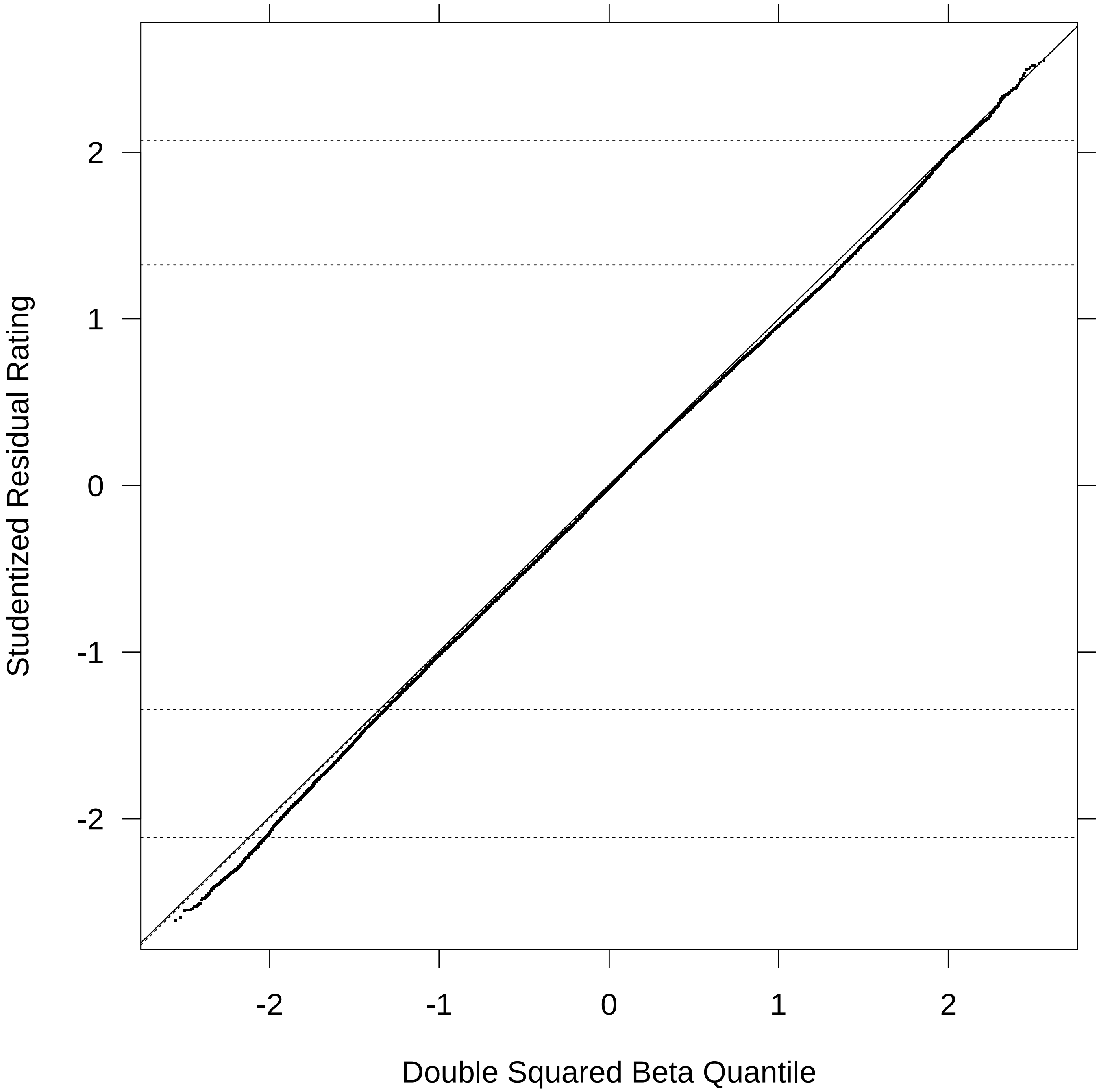
$$\hat{\phi}_{uj} = \frac{\hat{\psi}_{uj}}{s_u} \sim \sqrt{n_u - 1} DSB(0.5, 0.5n_u - 1)$$

A random variable on [-1,1] has a double square beta distribution, $DSB(p,q)$, if its distribution is symmetric and its square is distributed $BETA(p,q)$

This distribution is independent of β_u and γ_u, but changes with n_u

Good agreement!

PLOT: Check Normality of Residuals



Quantile Plots Definition

Empirical quantiles of a set of data — d_i for $i = 1$ to n — are graphed against the quantiles of a reference distribution $p = F(x)$

Used to check the appropriateness of an assumption that the distribution of the data is well approximated by the reference distribution

Two cases:

- i.d. quantile plot — assume the d_i are identically distributed
- mixture quantile plot — the data are not i.d. but assume that the means and variances of the d_i are known

I.D. Quantile Plot Definition

$d_{(i)}$ — data ordered from smallest to largest

$F^{-1}(i(n+1)^{-1})$ — the quantile of probability $i(n+1)^{-1}$ of the reference distribution

$d_{(i)}$ is plotted against $F^{-1}(i(n+1)^{-1})$

Mixture Quantile Plots Definition

d_i — non-identically distributed d_i with known means and variance

F_i — the reference distribution for d_i

d_i are centered and have variance $v_i = \sigma^2(d_i)$

Define an empirical distribution function with point mass p_i at d_i where

$$p_i = n(n+1)^{-1} v_i^{-1} \sum_{\ell=1}^{n} v_\ell^{-1}$$

and $d_{(i)}$ is the quantile of order $i(n+1)^{-1}$ of this empirical distribution

Mixture Quantile Plots Execution

Reference distribution is taken to be a mixture:

(1) take a sample with replacement of size n of the indices 1 to n with probability p_i for index i; Denote the indices by $i_1, \ldots, i_n$

(2) generate a sample d_i^* of size n from the distributions F_{i_j} for $j = 1$ to n

(3) generate samples from this mixture distribution by simulation, repeatedly carrying out (1) and (2), sorting the output of (2) to form order statistics $d_{(i)}^*$ for $i = 1$ to n, and then averaging each order statistic across samples to form $\overline{d}_{(i)}^*$

$d_{(i)}$ is plotted against $\overline{d}_{(i)}^*$

Are There Random Scale Effects?

Can be answered by checking if

Standardized residuals:

$$\begin{aligned} \hat{\psi}_{uj} &= \frac{\hat{\tau}_{uj}}{\sqrt{\bar{p}_{ujj}}} \\ &\sim N(0,1) \end{aligned}$$

Residual variance:

$$s_u^2 \sim MSQ(n_u - 1)$$

Normality of $\hat{\psi}_{uj}$

Recall that: Coefficient of kurtosis (0 for normal):

$$\frac{E(\hat{\psi}_{uj}^4)}{E^2(\hat{\psi}_{uj}^2)} - 3 = 3\sigma^2(\gamma^2)$$

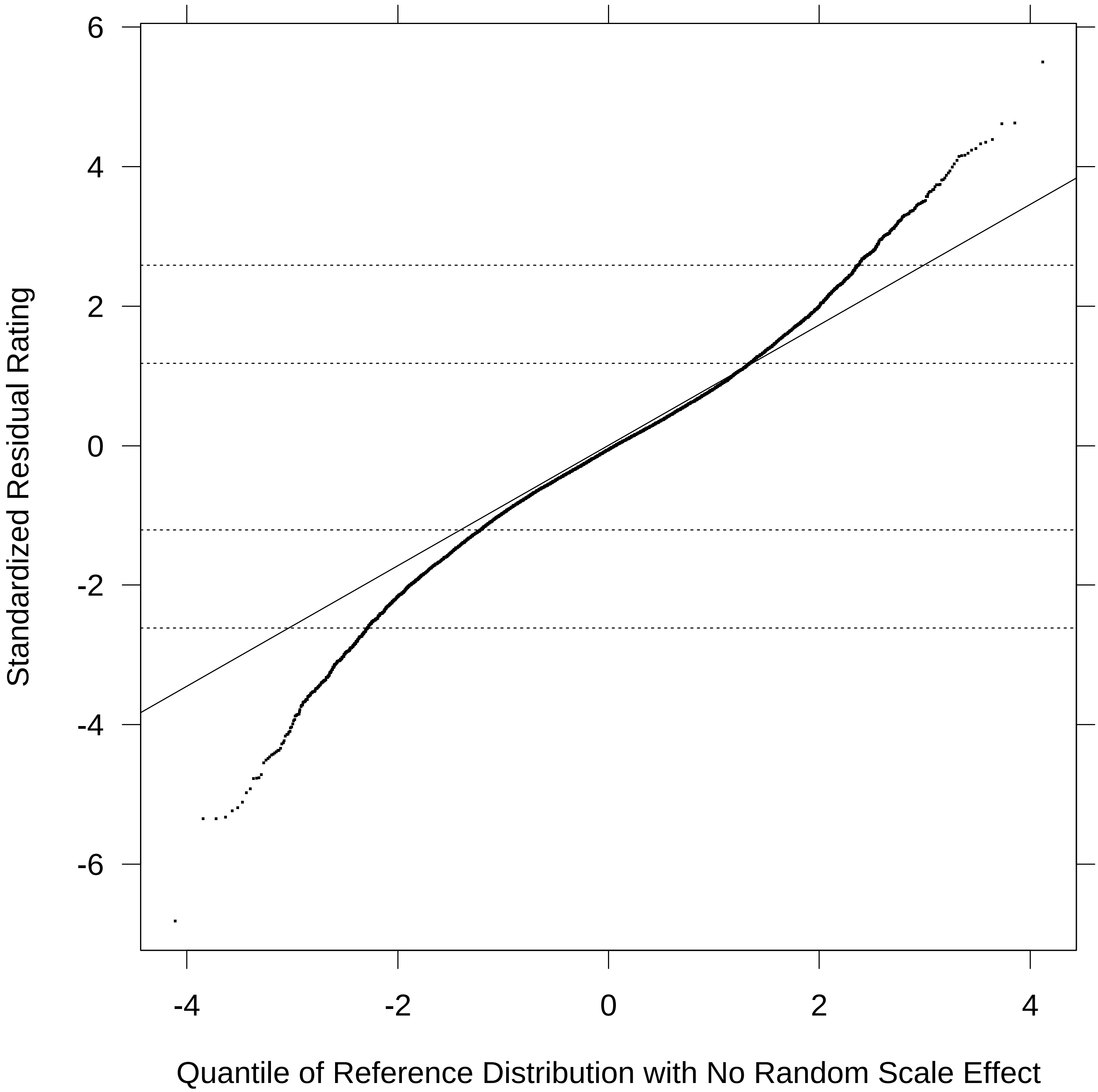
As $\sigma^2(\gamma^2)$ increases, the coefficient increases

Heavy tails are the expected behavior when random scale effects are present

See that $\hat{\psi}_{uj}$ has tails heavier than normal

Thus, random scale effects!

PLOT: Normal QQ of $\hat{\psi}_{uj}$



Are $s_u^2 \sim$ MSQ($n_u - 1$)?

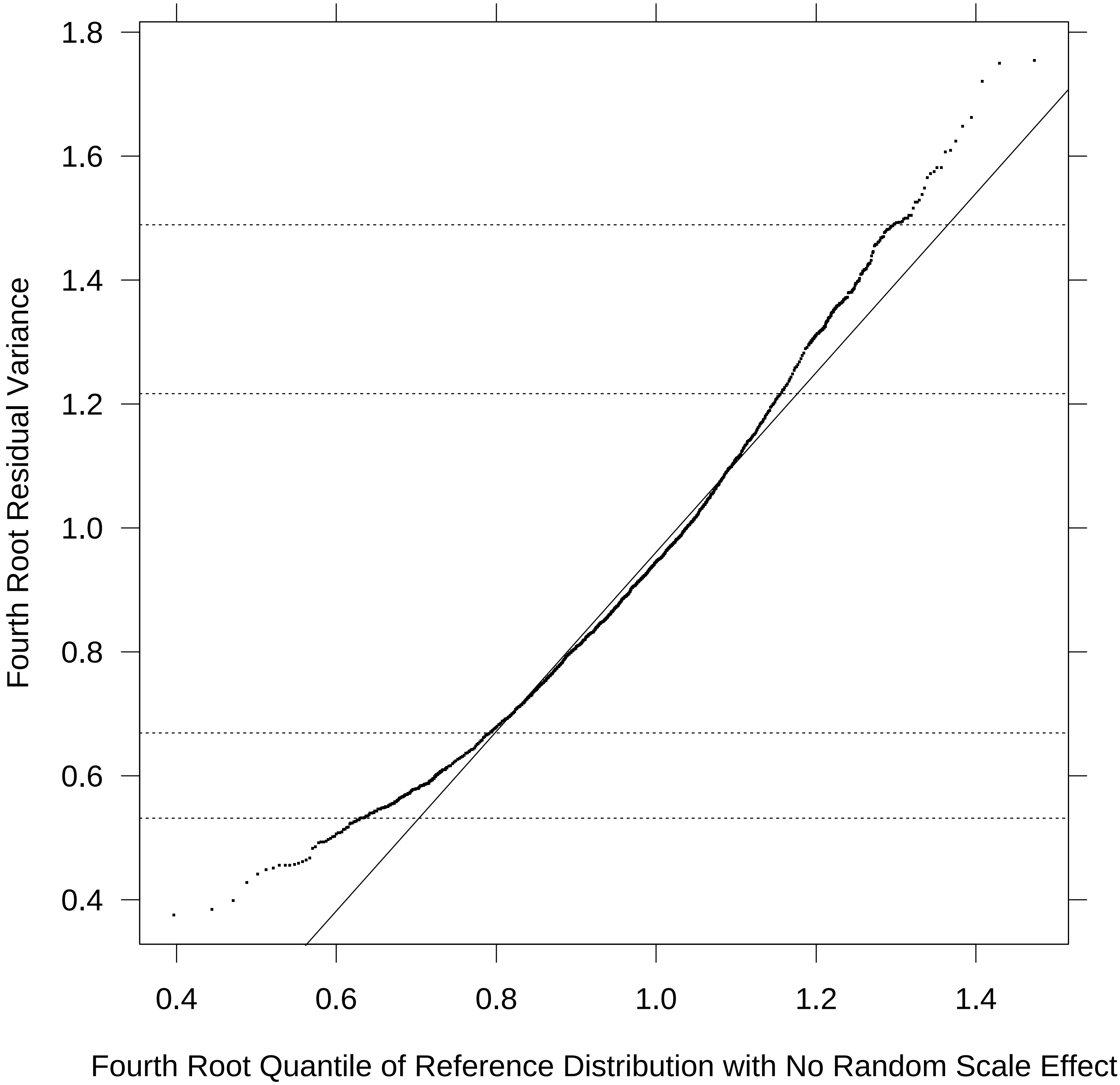
Mixture quantile plot of fourth root s_u^2 against fourth root quantiles of the $MSQ(n_u - 1)$ mixture distribution

The variance of s_u^2 taken to be $2(n_u - 1)^{-1}$, the variance when there are no random scale effects

- Random scale effects increase the variance of s_u^2
- Empirical distribution is highly skewed relative to the reference distribution

Yes, random scale effects!

PLOT: QQ of s_u^2



Two Assumptions in Place

1) $\varepsilon_{uj} \sim N(0, \sigma^2_{a(u,j)})$

2) Existence of random scale effects

Next need to determine the distributions, $F(\beta_u)$ and $F(\gamma_u)$

Model Random Scale Distribution

We know

$$s_u^2 \sim \gamma_u^2 MSQ(n_u - 1)$$

$$\hat{\psi}_{uj} \sim \gamma_u N(0, 1)$$

Use graphical deconvolution of s_u^2 and $\hat{\psi}_{uj}$ separately to identify the unknown distribution of γ_u

Deconvolution of s_u^2

Posit distribution for γ_u

Generate mixture quantiles of reference distributions $\gamma_u^2 MSQ(n_u - 1)$ — either by simulation or direct methods

Need estimates of $\hat{\sigma}^2(s_u^2)$: If no random scale effects, then $\sigma^2(\gamma^2) = 0$, and $\sigma^2(s_u^2) = 2(n_u - 1)^{-1}$

Plot the fourth roots of the order statistics of s_u^2 against the fourth roots of the generated reference quantiles

Families to consider for γ_u^2

1. Gamma: $G(h, \lambda)$ where h is the shape parameter and λ the scale

2. Weibull: $W(h, \lambda)$ where h is the transformation parameter and λ the scale

3. Inverse Gamma: $IG(h, \lambda)$ where h is the shape and λ the scale

4. Log Normal: $LN(\mu, \lambda^2)$ where μ and λ^2 are the mean and variance of the natural log

Use moment matching to specify the unkown parameters, producing a specific candidate distribution to carry out the graphical deconvolution

Moment Matching

Match based on the mean and variance of γ_u^2: $E(\gamma_u^2)$ and $\sigma^2(\gamma^2)$:

Mean has the value 1 because of our model normalization of scale

Use the estimate, $\hat{\sigma}^2(\gamma^2)$ for variance

Estimate of $\hat{\sigma}^2(\gamma^2)$

Since we assumed that $E\gamma_u^2$ is 1, the variance of γ_u^2 is

$$\sigma^2(\gamma^2) = E\gamma_u^4 - 1$$

Because $\hat{\zeta}_{uj}$ is normal and $E\left(\frac{\hat{\zeta}_{uj}^2}{\bar{p}_{ujj}}\right) = 1$,

$$E\left(\frac{\hat{\zeta}_{uj}^4}{\bar{p}_{ujj}^2}\right) = 3$$

Thus, $E(\hat{\psi}_{uj}^4) = E(\gamma_u^4)E\left(\frac{\hat{\zeta}_{uj}^4}{\bar{p}_{ujj}^2}\right) = 3(\sigma^2(\gamma^2) + 1)$

Coefficient of kurtosis (0 for normal):

$$\frac{E(\hat{\psi}_{uj}^4)}{E^2(\hat{\psi}_{uj}^2)} - 3 = 3\sigma^2(\gamma^2)$$

As $\sigma^2(\gamma^2)$ increases, the coefficient increases, so the tails get heavier

So, estimate of variance $\sigma^2(\gamma^2)$

$$\hat{\sigma}^2(\gamma^2) = \frac{\sum_{u=1}^{m}\sum_{j=1}^{n_u}\hat{\psi}_{uj}^4}{3n} - 1$$

Illustrate with Log Normal Family

$$\begin{aligned}\kappa_1 &= e^{\mu} \\ \kappa_2 &= e^{\lambda^2/2}\end{aligned}$$

Then

$$E(\gamma_u^2) = \kappa_1 \kappa_2$$

and

$$\sigma^2(\gamma_u^2) = \kappa_1^2 \kappa_2^2 (\kappa_2^2 - 1) = \kappa_2^2 - 1$$

Using $\hat{\sigma}^2(\gamma^2)$ in place of $\sigma^2(\gamma^2)$, and using $E(\gamma_u^2) = 1$, the candidate log normal has

$$\lambda^2 = \log(\kappa_2^2) = \log(1 + \hat{\sigma}^2(\gamma^2))$$

and

$$\mu = \log(\kappa_1) = -0.5 \log(1 + \hat{\sigma}^2(\gamma^2))$$

Distribution of s_u^2

Recall

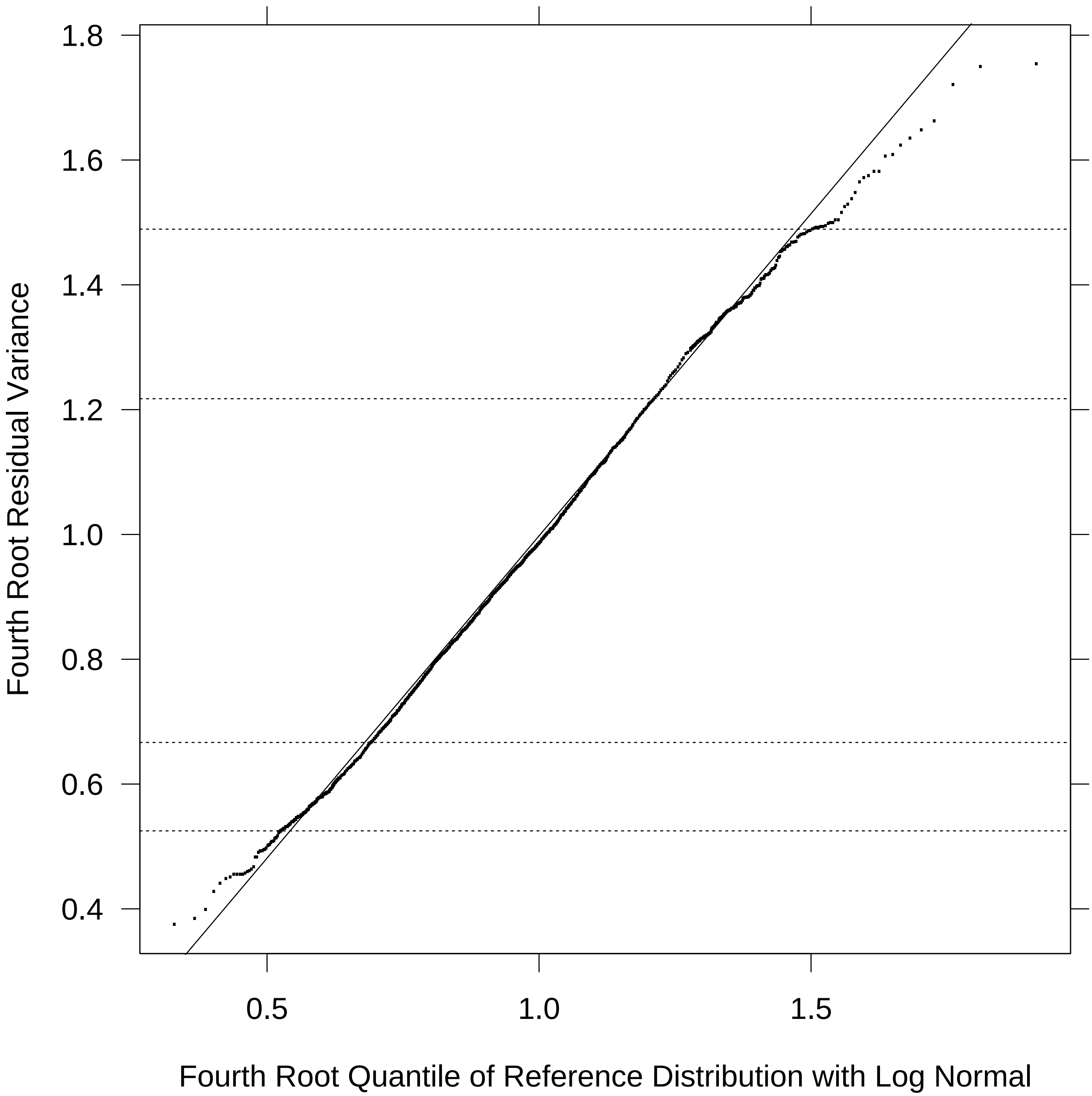
$$s_u^2 \sim \gamma_u^2 MSQ(n_u - 1)$$

For the rater data, $\hat{\sigma}^2(\gamma^2) = 0.50$, so the candidate log normal is

$$\gamma_u^2 \sim LN(-0.20, 0.41)$$

Want to check if $s_u^2 \sim LN(-0.20, 0.41) \times \chi^2/(n_u - 1)$ where the χ^2 has degrees of freedom equal to $n_u - 1$

PLOT: QQ of s_u^2 for LN Reference Distribution



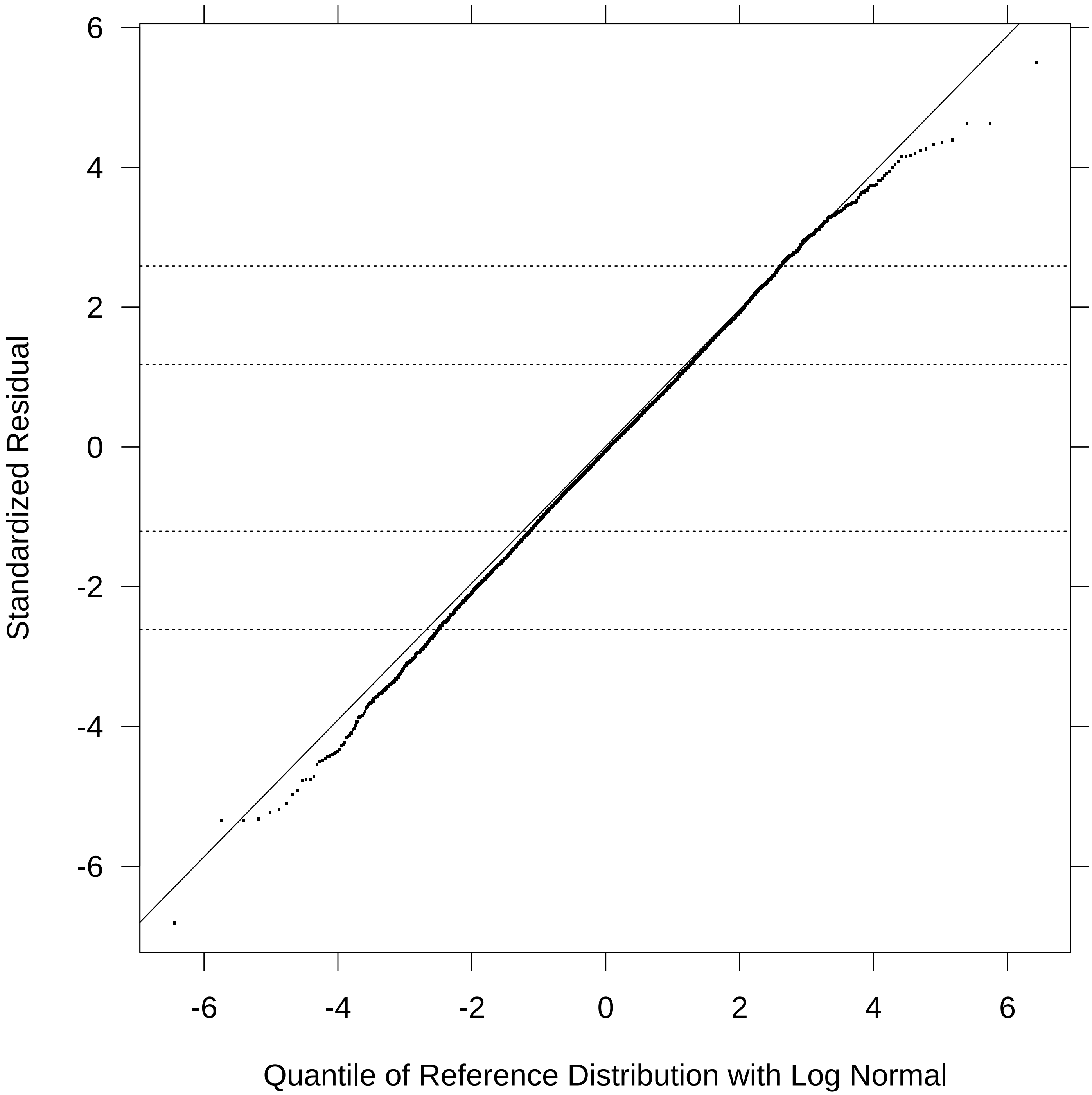
Distribution of $\hat{\psi}_{uj}$

Recall

$$\hat{\psi}_{uj} \sim \gamma_u N(0,1)$$

Want to check if $\hat{\psi}_{uj} \sim LN(-0.20, 0.41) \times N(0,1)$

PLOT: QQ of $\hat{\psi}_{uj}$ for LN Reference Distribution



Comments on Log Normality of γ_u^2

Good agreement of the empirical distributions with the log normal reference distribution

Check Standard Distributional Assumption

Usually random scale assumed $\sqrt{IG(h,\lambda)}$ where h is the shape and λ is the scale

First moment matching results in candidate distribution $IG(h,(h-1)^{-1})$

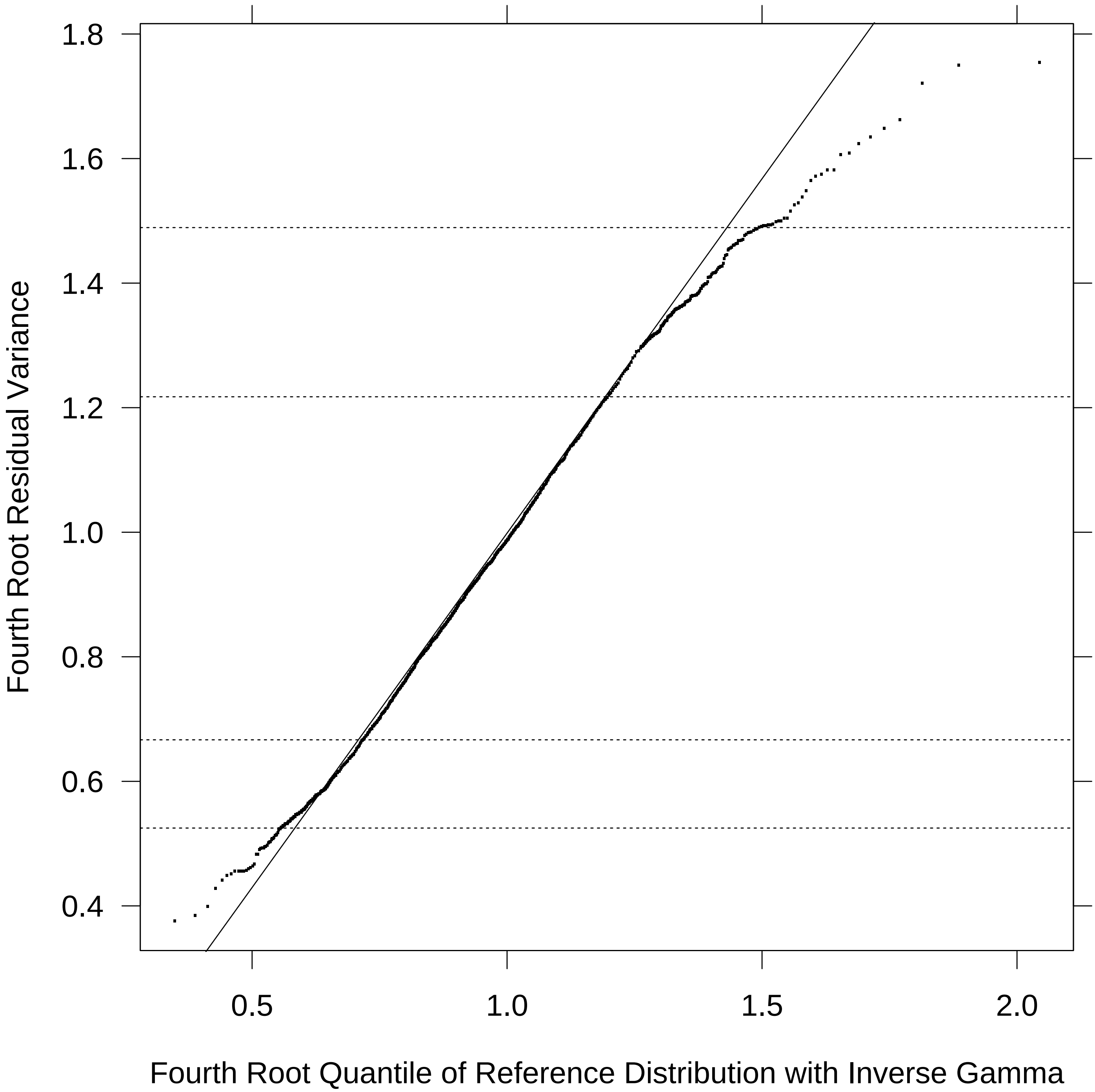
Second moment matching results in $h = 4.0$

Let's look at the deconvolution quantile plots

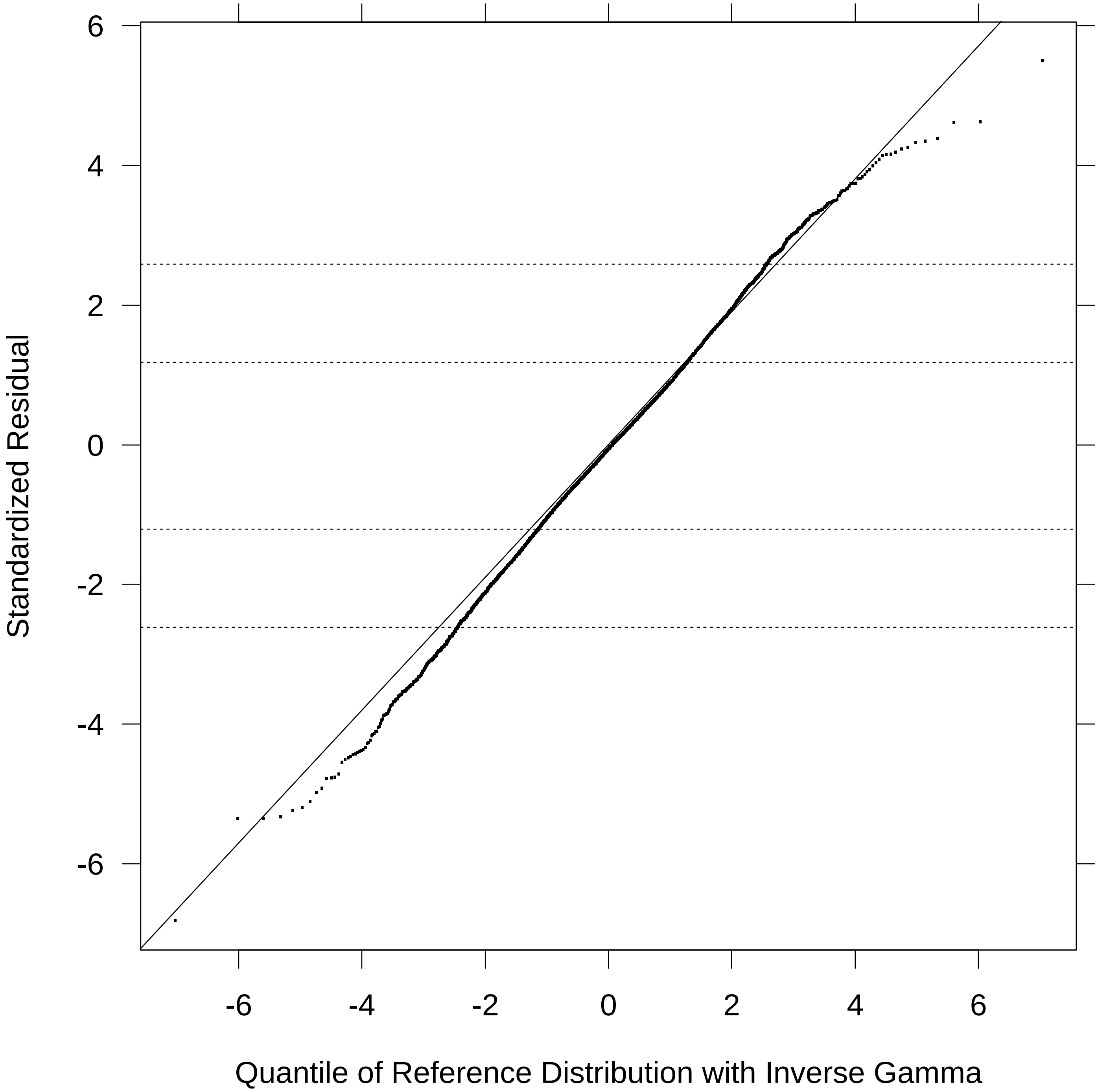
Check quantile plots for s_u^2 and $\hat{\psi}_{uj}$

- Mixture quantile plot of fourth root of residual variance for the $\sqrt{IG(4,1/3)}$ random scale distribution
- Quantile plot of standardized residuals for $\sqrt{IG(4,1/3)}$ random scale distribution

PLOT: QQ for Residual Variance for IG Reference



PLOT: QQ for Standardized Residuals for IG Reference



Comments on Inverse Gamma Assumption

The empirical distributions are not as well approximated as for the log normal

Conclude: Log normal is a better choice!

Two Distributional Specifications in Place

$$\zeta_{uj} \sim N(0,1)$$
$$\gamma_u^2 \sim LN(-0.20, 0.41)$$

Still need the distribution of β_u

Modeling the Random Location Distribution

The unit regression coefficients, $\hat{\beta}_u$, will be used to identify the distribution of the random location parameters, β_u:

$$\hat{\beta}_u \sim \beta_u + \gamma_u \xi_u$$

We know the distribution of $\gamma_u \xi_u$

γ_u is log normal and

$$\xi_u \sim N(0, \sigma^2(\xi_u))$$

where

$$\sigma^2(\xi_u) = \frac{1}{\sum_{a(u,j) \in A(u)} \sigma_{a(u,j)}^{-2}}$$

Need to prescribe distribution for β_u and then use graphical deconvolution to find good approximating distribution for β_u

Finding Distribution for β_u

Use deconvolution

Entertain a tentative distribution for β_u

Convolve with the known product distribution

Plot the quantiles of the $\hat{\beta}_u$ against the quantiles of the convolution reference distribution

First Assume β_u Normal

$$\hat{\beta}_u \sim \beta_u + \gamma_u \xi_u$$

If

$$\beta_u \sim N(0, \sigma^2(\beta)),$$

then

$$\hat{\beta}_u \sim N(0, \sigma^2(\beta)) + \sqrt{LN(-0.20, 0.41)} N\left(0, \sigma^2(\xi_u)\right)$$

Need estimate of $\hat{\sigma}^2(\beta)$

Estimate of $\hat{\sigma}^2(\beta)$

$$\begin{aligned}\hat{\beta}_u &= (X_u'X_u)^{-1}X_u'y_u \\ &= \beta_u + \gamma_u(X_u'X_u)^{-1}X_u'\zeta_u \\ &= \beta_u + \gamma_u\xi_u\end{aligned}$$

where

$$\xi_u = \frac{\sum_{j=1}^{n_u} x_{uj}\zeta_{uj}}{\sum_{j=1}^{n_u} x_{uj}^2} \sim N(0, \sigma^2(\xi_u))$$

where

$$\sigma^2(\xi_u) = \frac{1}{\sum_{\ell=1}^{n_u} x_{u\ell}^2} = \frac{1}{\sum_{a(u,j)\in A(u)} \sigma_{a(u,j)}^{-2}}$$

Then

$$\sigma^2(\hat{\beta}) = \sigma^2(\beta) + \sigma^2(\xi_u)$$

For this data

$$\hat{\sigma}^2(\beta) = \frac{1}{3491}\sum_{u=1}^{3491}\left(\hat{\beta}_u^2 - \sigma^2(\xi_u)\right) = 1.50$$

Are $\hat{\beta}_u$ Normal?

Quantiles of the data, $\hat{\beta}_u$, are graphed against the mixture quantiles of the distribution,

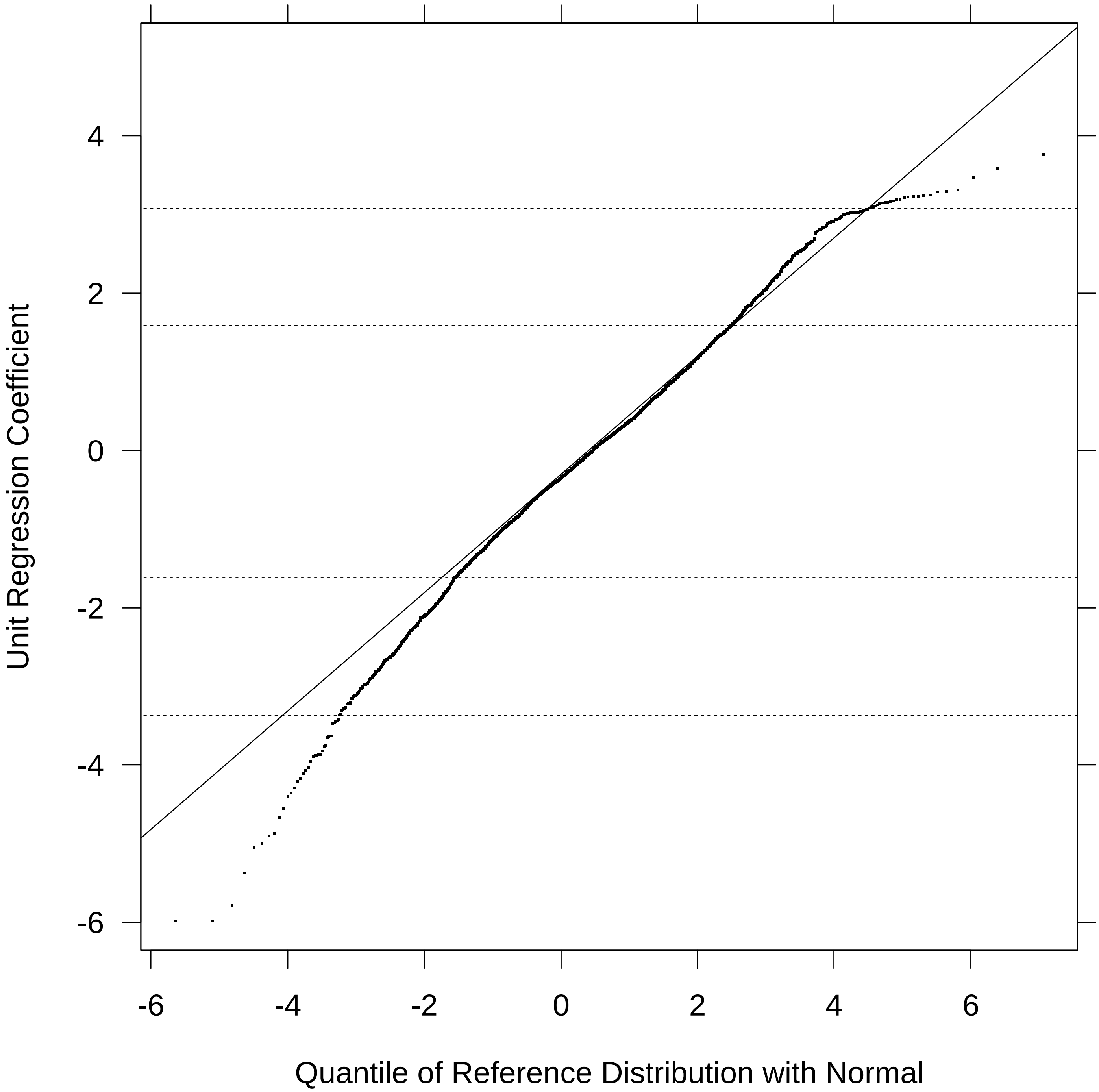
$$N(0,1.5)+\sqrt{LN(-0.20,0.41)}N\left(0,\sigma^2(\xi_u)\right)$$

Not good agreement!

- Particularly at right where data is bounded by 10
- But this departure involves only small fraction of the data
- Main departure: remaining 99% of the distribution show heavier tails than the normal

Try $T(d,0,\lambda^2)$ family

PLOT: QQ of $\hat{\beta}_u$ for Normal Reference



Are β_u from a T Distribution?

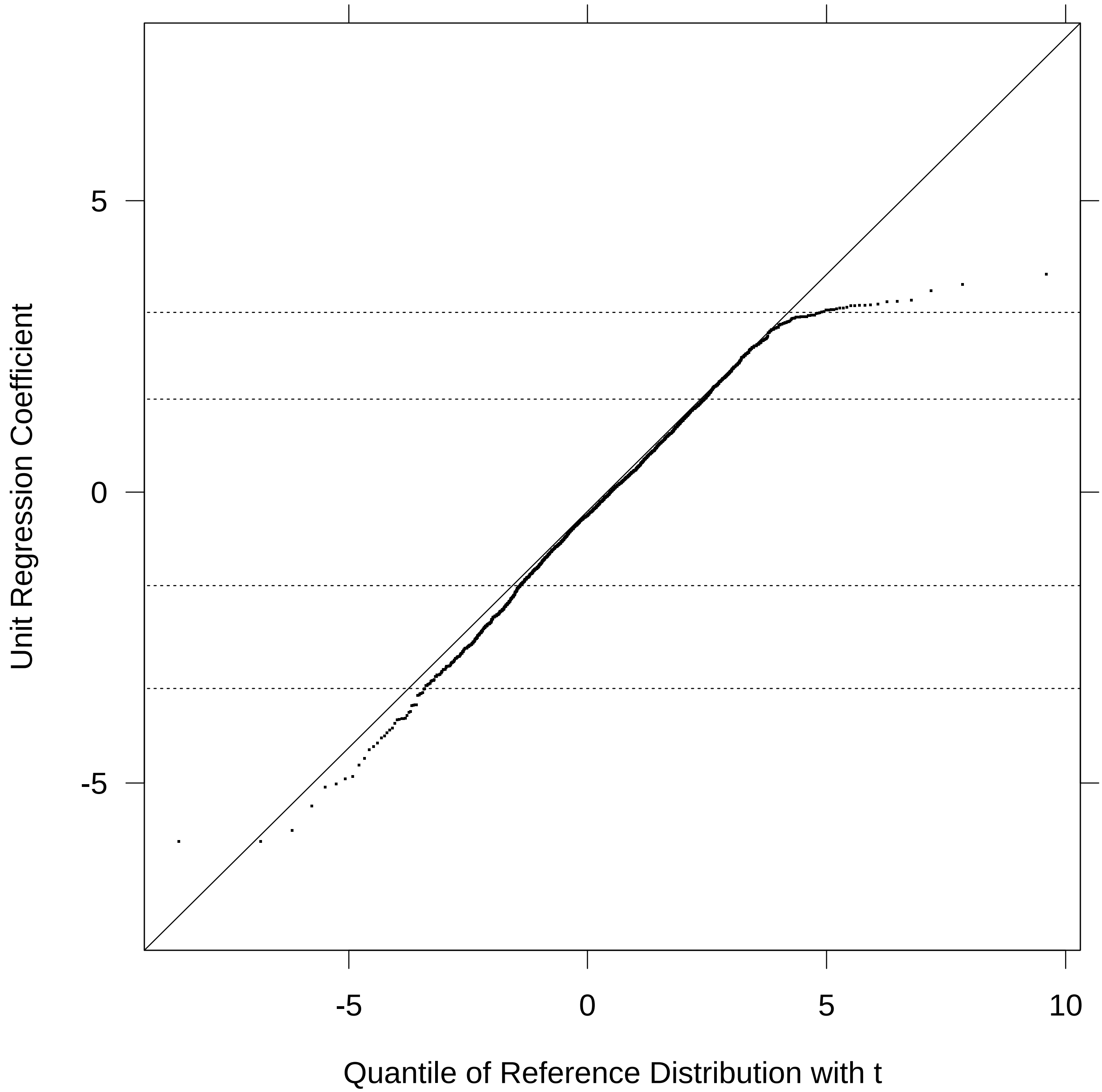
Try $T(d, 0, \lambda^2)$ family

- λ selected by moment matching
- So, $T(d, 0, 1.50(1 - 2d^{-1}))$
- d will be chosen by the quantile plot method — taking the value of d to be the one that produces the best visual match on the mixture quantile plot
- $d = 5$ provides a good match

Last distributional assumption in place:

$$F(\beta_u) \sim T(5, 0, 1.30)$$

PLOT: QQ of $\hat{\beta}_u$ for T Reference



Final Model Specification

$$
\begin{aligned}
r_{uj} &= \alpha_{asq} + \beta_u + \gamma_u \varepsilon_{uj} \\
E\beta_u &= 0 \\
E\varepsilon_{uj} &= 0 \\
E\gamma_u^2 &= 1 \\
Er_{ac} &= \alpha_{asq}
\end{aligned}
$$

α_{asq} smooth across quarters for a given *as*

$\varepsilon_{uj} \sim$ normal

$\gamma_u^2 \sim$ log normal

$\beta_u \sim$ t with 5 degrees of freedom

Now check for independence

Dependence of γ_u^2 on β_u

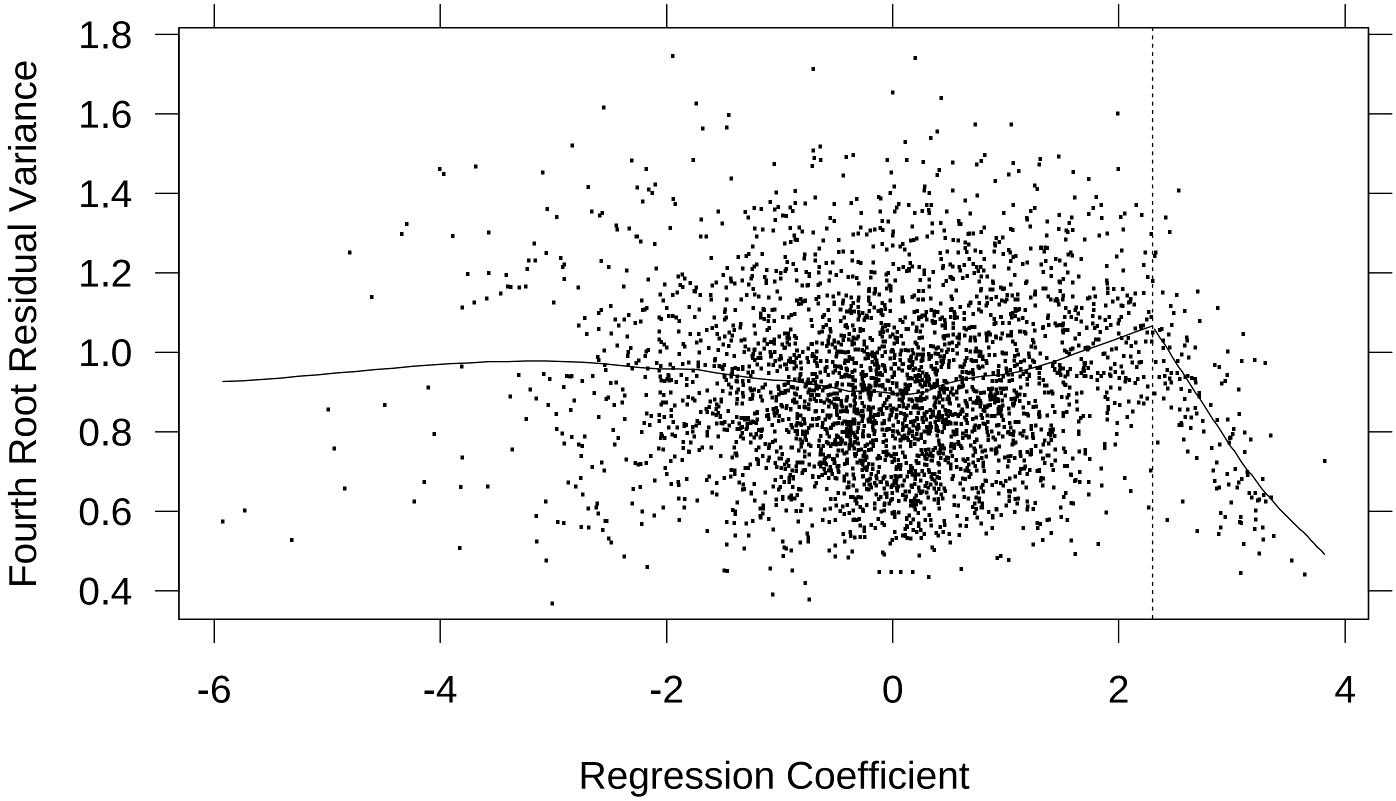
Model specifies that the amount of variation in $\gamma_u \varepsilon_{uj}$ does not depend on β_u

Potential relationship: raters who center themselves in one region of the rating scale may use up a different amount of the scale than those in another region

Use diagnostic display to check for a relationship:

- Fourth root s_u^2 plotted against $\hat{\beta}_u$
- One loess curve fits the points to the left of the vertical dashed line; the smoothing parameter is 0.15
- Second loess curve fits the points to the right of the line; the smoothing parameter is 0.75
- Line is drawn at 2.6 and 2.7% of points are to the right of it

PLOT: Diagnostic γ_u^2 vs. β_u



Comments on Relationship of γ_u^2 on β_u

Clearly, the variability decreases for the very highest values of $\hat{\beta}_u$ because for these raters, the ratings are close to 10, so their variation is constrained

Boundary effect is the same as in the quantile plots of $\hat{\beta}_u$

Only small fraction of data affected by boundary effect, so ignore in model building

Check Correlation in Residuals

Another plausible dependence is attribute correlation induced at the rater level

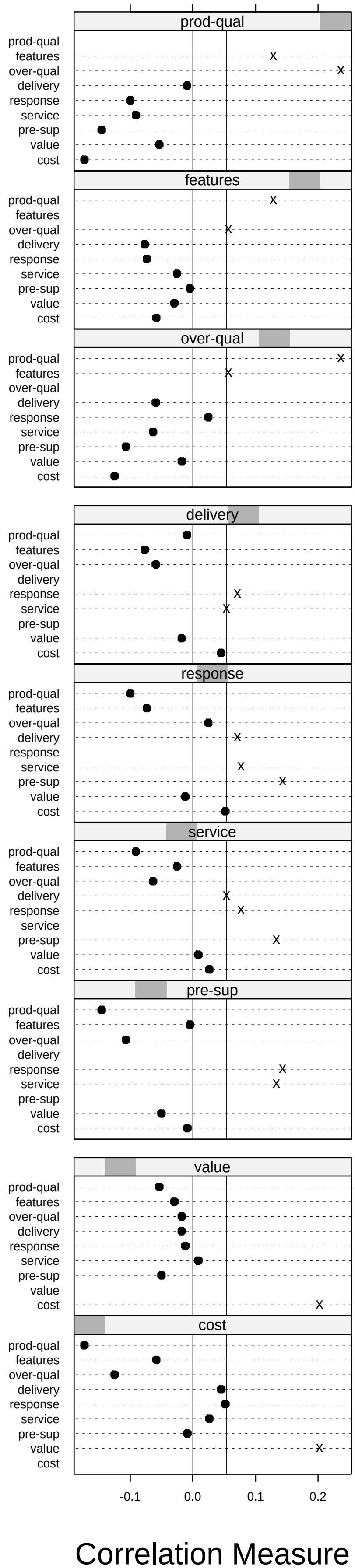
To check:

- Compute correlation coefficients of the studentized residuals for all pairs of attributes across raters
- Subtract expected correlations under an assumption of no correlation
- Difference measure is check for correlation in the errors

Diagnostic plot:

- Shows the correlation measure given each attribute
- Each panel shows the correlation between the panel attribute and all other attributes for which there is a measure (some pairs never appear because of the survey plan)

PLOT: Diagnostic to Check Correlation



Clustering of Attributes

Simple clustering of the attributes

Breaks between the panels divide the attributes into four clusters x's show differences for attribute pairs in the same cluster and the •'s show differences for attribute pairs in different clusters

Clusters formed by:

- Begin with all attributes in separate clusters
- At each stage merged the two clusters that had the largest single value of the difference measure for all pairs of attributes, one in one cluster and one in the other
- Continued until there were no singletons

Differences within the cluster are greater than the differences for one attribute in the cluster and one not in the cluster

Right vertical line on the panels drawn at the minimum within-cluster measure

Interpretation of Clusters

Clusters match our external information:

Top cluster concerns the physical product, its quality and technology

Second concerns the supplier-customer interaction

Third concerns the price of the product and whether it is worth what the customer paid for it

Correlation is convincing but magnitude is small, so it can be ignored

Bayesian Computation

The full Bayesian model

Markov Chain (MC) Monte Carlo Methods

Conditional posterior distributions

Bayesian Model: Notation

$r_{uj} = \alpha_{a(u,j)s(u,j)} + \beta_u + \gamma_u \varepsilon_{uj}$			
α_{as}	β_u	γ_u^2	ε_{uj}
	$\lambda^2 \equiv$ $0.6\mathrm{Var}(\beta_u)$	$\nu^2 \equiv$ $\mathrm{Var}(\ln(\gamma_u^2))$	$\sigma^2_{a(u,j)} \equiv$ $\mathrm{Var}(\varepsilon_{uj})$

$$\alpha \equiv \{\alpha_{as} : a = 1, \ldots, 9; s = 1, \ldots, 19\}$$
$$\beta \equiv \{\beta_u : u = 1, \ldots, m\}$$
$$\gamma^2 \equiv \{\gamma_u^2 : u = 1, \ldots, m\}$$
$$\sigma^2 \equiv \{\sigma_a^2 : a = 1, \ldots, 9\}$$

$q(u, j)$ is ignored.
ν^2 is fixed.

Bayesian Model: Review

(1) Missing data mechanism: ignorable

(2) Observed ratings: for $u = 1, \ldots, m$ and $j = 1, \ldots, n_u$

$$r_{uj} | (\beta, \gamma, \theta) \overset{ind}{\sim} \mathrm{N}\left(\alpha_{a(u,j)s(u,j)} + \beta_u, \gamma_u^2 \sigma^2_{a(u,j)}\right)$$

(3) Location effects: for $u = 1, \ldots, m$

$$\beta_u | (\gamma, \theta) \overset{ind}{\sim} \mathbf{t}\left(d = 5, 0, \lambda^2\right)$$

or, equivalently,

$$\begin{aligned} \omega_u | (\gamma, \theta) &\overset{ind}{\sim} \mathrm{G}(d/2, d/2) \\ \beta_u | (\omega, \gamma, \theta) &\overset{ind}{\sim} \mathrm{N}\left(0, \lambda^2 / \omega_u\right) \end{aligned}$$

where $\omega = \{\omega_1, \ldots, \omega_m\}$

(4) Scale effects: for $u = 1, \ldots, m$

$$\ln(\gamma_u^2) | \theta \overset{ind}{\sim} \mathrm{N}\left(-0.5\nu^2, \nu^2\right) \qquad (\text{fixed } \nu^2 = 0.41)$$

$\mathrm{G}(s, p)$: Gamma, where s is the shape parameter and $1/p$ the scale.

Bayesian Model: Prior Distribution

All the components of the parameters are independent of each other with

$$\alpha_{as} \sim \text{Flat}$$

$$\sigma_a^2 \sim \text{IG}(\phi_{\sigma^2}, \rho_{\sigma^2})$$

$$\lambda^2 \sim \text{IG}(\phi_{\lambda^2}, \rho_{\lambda^2})$$

where ϕ_{σ^2}, ϕ_{λ^2}, ρ_{σ^2}, and ρ_{λ^2} are known (small) positive constants

Inference: Posterior Distribution

Posterior: $f(\Theta|\text{Data}) \equiv f(\alpha, \sigma^2, \beta, \omega, \lambda^2, \gamma^2|\text{Data})$

Marginals: for example, $f(\alpha|\text{Data})$

$$= \int f(\alpha|\sigma^2, \beta, \omega, \lambda^2, \gamma^2|\text{Data}) \cdot f(\sigma^2, \beta, \omega, \lambda^2, \gamma^2|\text{Data}) d\sigma^2 d\beta d\omega d\lambda^2 d\gamma^2$$

$$\approx \frac{1}{N}\sum_{t=1}^{N} f(\alpha|(\sigma^2)^{(t)}, \beta^{(t)}, \omega^{(t)}, (\lambda^2)^{(t)}, (\gamma^2)^{(t)}|\text{Data})$$

Moments: for example, $\text{E}(\alpha_{as}|\text{Data})$

$$\approx \frac{1}{N}\sum_{t=1}^{N} \text{E}(\alpha_{as}|(\sigma^2)^{(t)}, \beta^{(t)}, \omega^{(t)}, (\lambda^2)^{(t)}, (\gamma^2)^{(t)}|\text{Data})$$

$$\approx \frac{1}{N}\sum_{t=1}^{N} \alpha_{as}^{(t)}$$

$\alpha^{(t)}, \sigma^{(t)}, \beta^{(t)}, \omega^{(t)}, \lambda^{(t)}$, and $\gamma^{(t)}$ are draws from the posterior distribution $f(\alpha, \sigma, \beta, \omega, \lambda, \gamma|\text{Data})$

Objective: take draws from $f(x) = f(x_1,\ldots,x_m)$, where $x = (x_1,x_2,\ldots,x_m) \in \mathcal{X}$ and x_i is a vector

Algorithm: (Alternating Conditional Sampling) starting with $x^{(0)} \in \mathcal{X}$, the t-th iteration consists of m steps:

Step 1: draw $x_1^{(t)}$ from $f(x_1|x_2^{(t-1)},\ldots,x_m^{(t-1)})$

$\vdots$

Step i: draw $x_i^{(t)}$ from

$$f(x_i|x_1^{(t)},\ldots,x_{i-1}^{(t)},x_{i+1}^{(t-1)},\ldots,x_m^{(t-1)})$$

$\vdots$

Step m: draw $x_m^{(t)}$ from $f(x_m|x_1^{(t)},\ldots,x_{m-1}^{(t)})$

As $t \to \infty$, $x^{(t)} \sim f(x)$.

Gibbs Sampler with Approximations

Objective: to approximate $f(x_i|x_{-i})$ with $g(x_i|x_{-i})$, from which sampling is simple

Correction: replace the Gibbs step

$$x_i = \texttt{draw}\ f(x_i|x_{-i})$$

with the Metropolis-Hastings step

$$x_i^* = \texttt{draw}\ g(x_i|x_{-i})$$

$$r = \frac{f(x_i^*|x_{-i})}{f(x_i|x_{-i})} \cdot \frac{g(x_i|x_{-i})}{g(x_i^*|x_{-i})}$$

$$x_i = x_i^*\ \texttt{with probability min}(r, 1)$$

---

where, x_{-i} is the current draw of all the components of x, except for x_i

Reference: Gelman, Carlin, Stern, and Rubin (1995)

MC Simulation: draw α and σ^2

Given $\Theta \setminus \alpha$ (*i.e.*, all the components, except for α), all the α_{as} are independent and α_{as} follows

$$\mathrm{N}\left(\frac{\sum_{(u,j)\in I(a,s)} \gamma_u^{-2}(r_{uj}-\beta_u)}{\sum_{(u,j)\in I(a,s)} \gamma_u^{-2}}, \frac{\sigma_a^2}{\sum_{(u,j)\in I(a,s)} \gamma_u^{-2}}\right)$$

Given $\Theta \setminus \sigma^2$, all the σ_a^2 are independent and σ_a^2 follows

$$\mathrm{IG}\left(\phi_{\sigma^2} + \frac{1}{2}\sum_{(u,j)\in I(a)} 1, \rho_{\sigma^2} + \frac{1}{2}\sum_{(u,j)\in I(a)} \gamma_u^{-2} e_{uj}^2\right)$$

where

$$e_{uj} = r_{uj} - \alpha_{a(u,j)s(u,j)} - \beta_u$$

An alternative (faster) scheme is to first draw σ^2 from $f(\sigma^2|\Theta \setminus \{\alpha, \sigma^2\})$ and then α's from $f(\alpha|\Theta \setminus \alpha)$

MC Simulation: draw β, ω, and λ^2

Given $\Theta \setminus \beta$, β's are independent and β_u follows the normal distribution with the mean

$$\frac{\gamma_u^{-2} \sum_{j=1}^{n_u} \sigma_{a(u,j)}^{-2} (r_{uj} - \alpha_{a(u,j)s(u,j)})}{\omega^2 \lambda^{-2} + \gamma_u^{-2} \sum_{j=1}^{n_u} \sigma_{a(u,j)}^{-2}}$$

and variance

$$\frac{1}{\omega^2 \lambda^{-2} + \gamma_u^{-2} \sum_{j=1}^{n_u} \sigma_{a(u,j)}^{-2}}$$

Given $\Theta \setminus \omega$, ω's are independent and ω_u follows

$$\mathrm{G}\left(\frac{d+1}{2}, \frac{d + \beta_u^2/\lambda^2}{2}\right)$$

Given $\Theta \setminus \lambda^2$, λ^2 follows

$$\mathrm{IG}\left(\phi_{\lambda^2} + \frac{m}{2}, \rho_{\lambda^2} + \frac{1}{2} \sum_{u=1}^{m} \omega_u \beta_u^2\right)$$

MC Simulation: draw γ^2

Given $\Theta \setminus \gamma^2$, γ_1^2, ..., and γ_m^2 are independent. For each u, $f(\gamma_u^{-2}|\Theta \setminus \gamma^2)$ has two factors

Factor 1: $\left(\gamma_u^{-2}\right)^{n_u/2} \exp\left\{-\gamma_u^{-2} z_u^2/2\right\}$, where $z_u^2 = \sum_{j=1}^{n_u} e_{uj}^2/\sigma_{a(u,j)}^2$

Factor 2: $\ln(\gamma_u^2)|(\theta) \stackrel{ind}{\sim} \mathrm{N}\left(-0.5\nu^2, \nu^2\right)$, with known ν^2

Approximation: replace Factor 2 with

$$\gamma_u^{-2}|(\theta) \sim \mathrm{G}\left(2+\delta(\nu^2), 1+\delta(\nu^2)\right),$$

where $\delta(\nu^2) = 1/\left(\exp\{\nu^2\} - 1\right)$. This leads to

$$\begin{aligned} & g(\gamma_u^{-2}|\Theta \setminus \gamma^2) \\ = \quad & \mathrm{G}\left(2+\delta(\nu^2)+n_u/2, 1+\delta(\nu^2)+z_u^2/2\right), \end{aligned}$$

Correction: use a Metropolis-Hastings step