

Advanced Regression Methods for Independent Data

STAT/BIOST 570, 2020

Practical Aspects of Regression Models with Weak Assumptions

Mauricio Sadinle
Department of Biostatistics
University of Washington
msadinle@uw.edu

Back to the FEV Data

- FEV: forced expiratory volume. FEV1: amount of air you can force from your lungs in one second.
- Data from 654 children and youths ages 3–19 in East Boston, 1980. (Childhood Respiratory Disease Study).
- For more information visit: <http://www.statsci.org/data/general/fev.html>

```
> url <- "http://www.statsci.org/data/general/fev.txt"
> data <- read.table(file = url, header = T, sep="\t", stringsAsFactors = F)
> data$Sex <- factor(data$Sex, levels=c("Male","Female"), labels=c(0,1))
> data$Smoker <- factor(data$Smoker, levels=c("Non","Current"), labels=c(0,1))
> data$HeightC <- data$Height - 60
> data$AgeC <- data$Age - 10
> data$HeightCSq <- data$HeightC^2
```

An Initial Normal Linear Model

Based on an exploratory analysis, we first looked at a normal linear model with regression function:

$$\begin{aligned} E(\text{FEV1} \mid \mathbf{X}) = & \beta_0 + \beta_1 \text{Height} + \beta_2 \text{Height}^2 + \beta_3 \text{Age} + \\ & (\beta_4 + \beta_5 \text{Height} + \beta_6 \text{Height}^2 + \beta_7 \text{Age}) I(\text{Sex}=\text{Female}) + \\ & (\beta_8 + \beta_9 \text{Height} + \beta_{10} \text{Height}^2 + \beta_{11} \text{Age}) I(\text{Smoker}=\text{Yes}) \end{aligned}$$

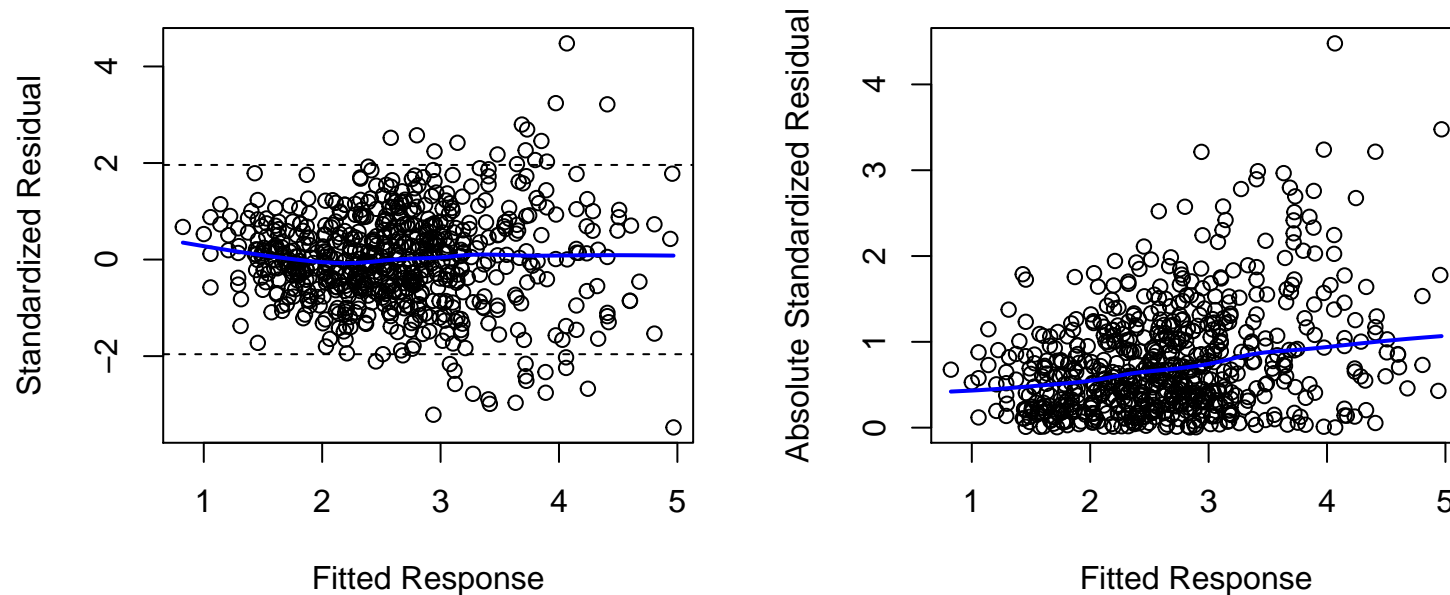
```
> model3 <- lm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   data=data)
```

An Initial Normal Linear Model

Based on the fit of this normal linear model, we looked at the residuals to check whether its assumptions were reasonable

```
> limits3 <- range(data$FEV, fitted(model3))
> par(mfrow = c(1,2))
> library(MASS)
> resid_mod3 <- stdres(model3)
> plot(fitted(model3), resid_mod3,
+      ylab = "Standardized Residual", xlab = "Fitted Response")
> lines(lowess(fitted(model3), resid_mod3), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
> plot(fitted(model3), abs(resid_mod3),
+      ylab = "Absolute Standardized Residual", xlab = "Fitted Response")
> lines(lowess(fitted(model3), abs(resid_mod3)), lwd = 2, col = "blue")
```

An Initial Normal Linear Model



We concluded that lack of homoscedasticity was an issue, so inferences based on a normal linear model wouldn't be reliable

A Gamma GLM

We then considered a gamma GLM, for which

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha V(\mu_i) = \alpha \mu_i^2 = \alpha E(Y_i | \mathbf{x}_i)^2,$$

which we thought could be appropriate for our data

We used the identity link, which is fine here since the outcome is away from zero

```
> FEV_Gamma <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   family = Gamma(link="identity"),  
+   data=data)
```

In this case, $\hat{\alpha}$ was

```
> summary(FEV_Gamma)$dispersion
```

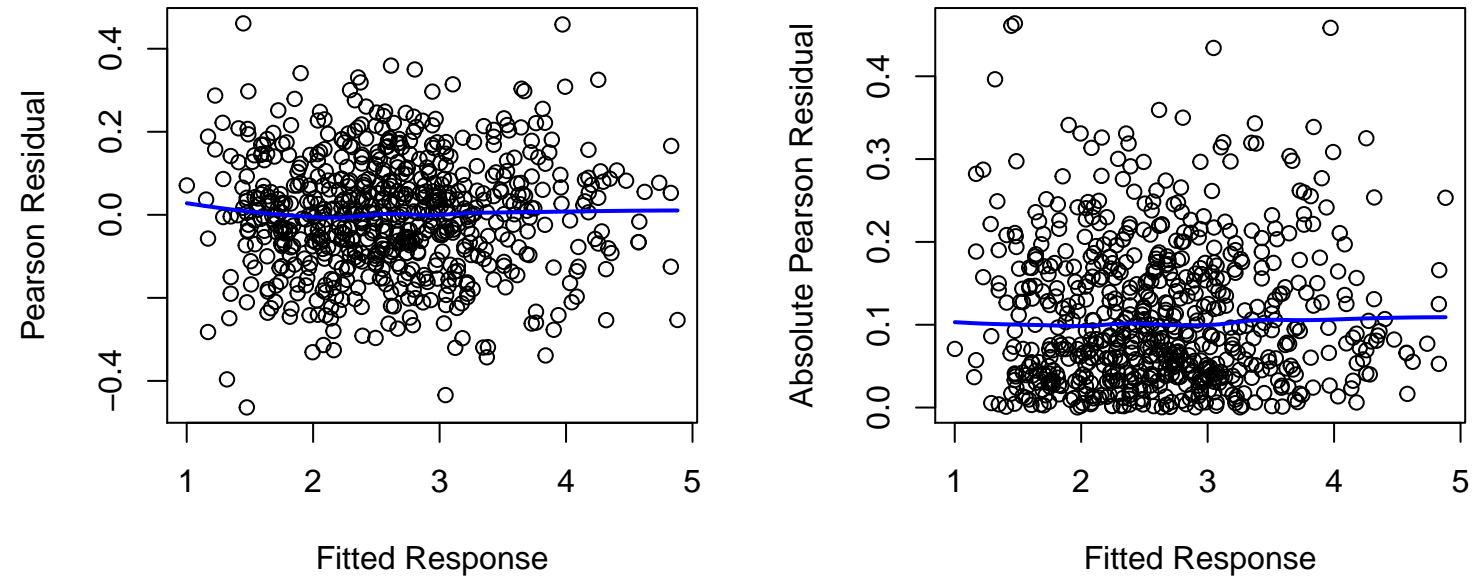
```
[1] 0.02015759
```

A Gamma GLM

We checked the Pearson residuals

```
> resid_Gamma <- residuals(FEV_Gamma, type = "pearson")
> fitted_Gamma <- fitted(FEV_Gamma)
> limits <- range(data$FEV, fitted_Gamma)
> par(mfrow = c(1,2))
> plot(fitted_Gamma, resid_Gamma,
+      ylab = "Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_Gamma, resid_Gamma), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
> plot(fitted_Gamma, abs(resid_Gamma),
+      ylab = "Absolute Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_Gamma, abs(resid_Gamma)), lwd = 2, col = "blue")
```

A Gamma GLM



We did not find patterns in the Pearson residuals that indicated violations to the model's assumptions

Quasi-Likelihood Approaches

In R, some quasi-likelihood approaches are implemented with the `glm` function

The regression function is taken as in a GLM:

$$E(Y_i | \mathbf{x}_i) = \mu_i = \mu(\mathbf{x}_i, \boldsymbol{\beta}) = g^{-1}(\mathbf{x}_i \boldsymbol{\beta})$$

for a link function $g(\cdot)$

The variance function is taken as

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha V(\mu_i)$$

for some known function $V(\cdot)$

Quasi-Likelihood Approaches

To specify a quasi-likelihood model, the family argument of `glm` can be taken as one of the following default options

```
quasi(link = "identity", variance = "constant")  
quasibinomial(link = "logit")  
quasipoisson(link = "log")
```

Available link functions: `logit`, `probit`, `cloglog`, `identity`, `inverse`, `log`, `1/mu^2` and `sqrt`. Also, the function `power` can be used to create a power link function

Available variance functions for `quasi`: `constant`, `mu(1-mu)`, `mu`, `mu^2` and `mu^3`

Variance function for `quasibinomial` is proportional to `mu(1-mu)`

Variance function for `quasipoisson` is proportional to `mu`

Quasi-Likelihood Approaches

Quasi-examples (pun totally intended!):

- `glm(y~x1+x2, family = quasi(link="identity", variance="mu^2"))`

corresponds to assuming

$$g(\mu_i) = \mu_i, \quad V(\mu_i) = \mu_i^2,$$

that is,

$$E(Y_i | \mathbf{x}_i) = \mathbf{x}_i \boldsymbol{\beta}, \quad \text{var}(Y_i | \mathbf{x}_i) = \alpha E(Y_i | \mathbf{x}_i)^2$$

- `glm(y~x1+x2, family = quasipoisson(link = "log"))`

corresponds to assuming

$$g(\mu_i) = \log(\mu_i), \quad V(\mu_i) = \mu_i,$$

that is,

$$E(Y_i | \mathbf{x}_i) = \exp(\mathbf{x}_i \boldsymbol{\beta}), \quad \text{var}(Y_i | \mathbf{x}_i) = \alpha E(Y_i | \mathbf{x}_i)$$

A Quasi-Likelihood Example

Continuing with the FEV data, let us first consider a quasi-likelihood approach with

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha V(\mu_i) = \alpha \mu_i = \alpha E(Y_i | \mathbf{x}_i)$$

```
> FEV_Quasi1 <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   family = quasi(link="identity", variance="mu"),  
+   data=data)
```

In this case, $\hat{\alpha}$ is

```
> summary(FEV_Quasi1)$dispersion
```

```
[1] 0.05291511
```

A Quasi-Likelihood Example

The quasi-likelihood approach relies on correct specification of the variance function,
 $\text{var}(Y_i \mid \mathbf{X}_i) = \alpha V(\mu_i)$

Note that if the mean-variance relationship holds, then the Pearson residuals

$$e_i^* = \frac{Y_i - \mu_i(\hat{\beta})}{\sqrt{\hat{\alpha} V(\mu_i(\hat{\beta}))}}$$

should fluctuate around 0 over the whole range of the covariates

As a function of $\mu_i(\hat{\beta})$, the Pearson residuals should also fluctuate around 0 with homogeneous variance

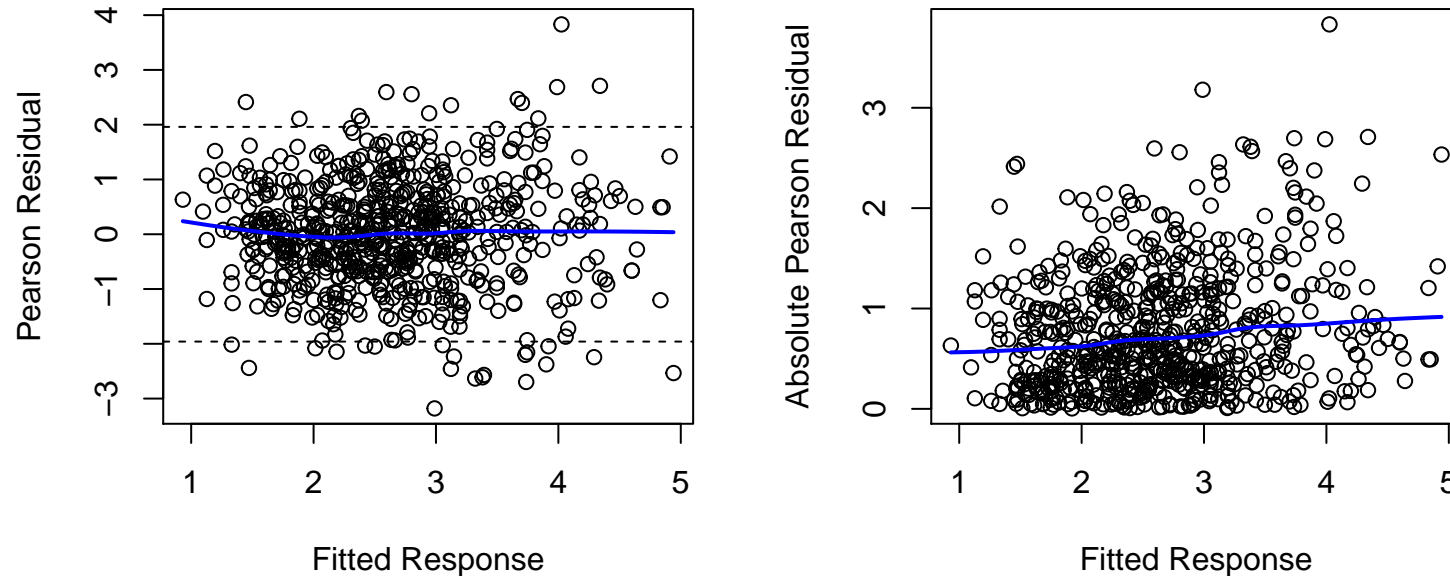
Note that in R, the `residuals` function doesn't use $\hat{\alpha}$ in computing Pearson residuals

A Quasi-Likelihood Example

Computation and plotting of the Pearson residuals

```
> resid_FEV_Quasi1 <- residuals(FEV_Quasi1, type = "pearson")/sqrt(summary(FEV_Quasi1)$dispersion)
> fitted_FEV_Quasi1 <- fitted(FEV_Quasi1)
> limits <- range(data$FEV, fitted_FEV_Quasi1)
> par(mfrow = c(1,2))
> plot(fitted_FEV_Quasi1, resid_FEV_Quasi1,
+      ylab = "Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi1, resid_FEV_Quasi1), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
> plot(fitted_FEV_Quasi1, abs(resid_FEV_Quasi1),
+      ylab = "Absolute Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi1, abs(resid_FEV_Quasi1)), lwd = 2, col = "blue")
```

A Quasi-Likelihood Example



The Pearson residuals reveal that the assumption $\text{var}(Y_i | \mathbf{x}_i) = \alpha\mu_i$ does not quite capture the dependence of the variability of the response as a function of the mean, as we can still see an increase in the variance of the residuals as a function of the fitted means

A Quasi-Likelihood Example

Let us now consider a quasi-likelihood approach with

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha V(\mu_i) = \alpha \mu_i^2 = \alpha E(Y_i | \mathbf{x}_i)^2$$

```
> FEV_Quasi2 <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   family = quasi(link="identity", variance="mu^2"),  
+   data=data)
```

In this case, $\hat{\alpha}$ is

```
> summary(FEV_Quasi2)$dispersion
```

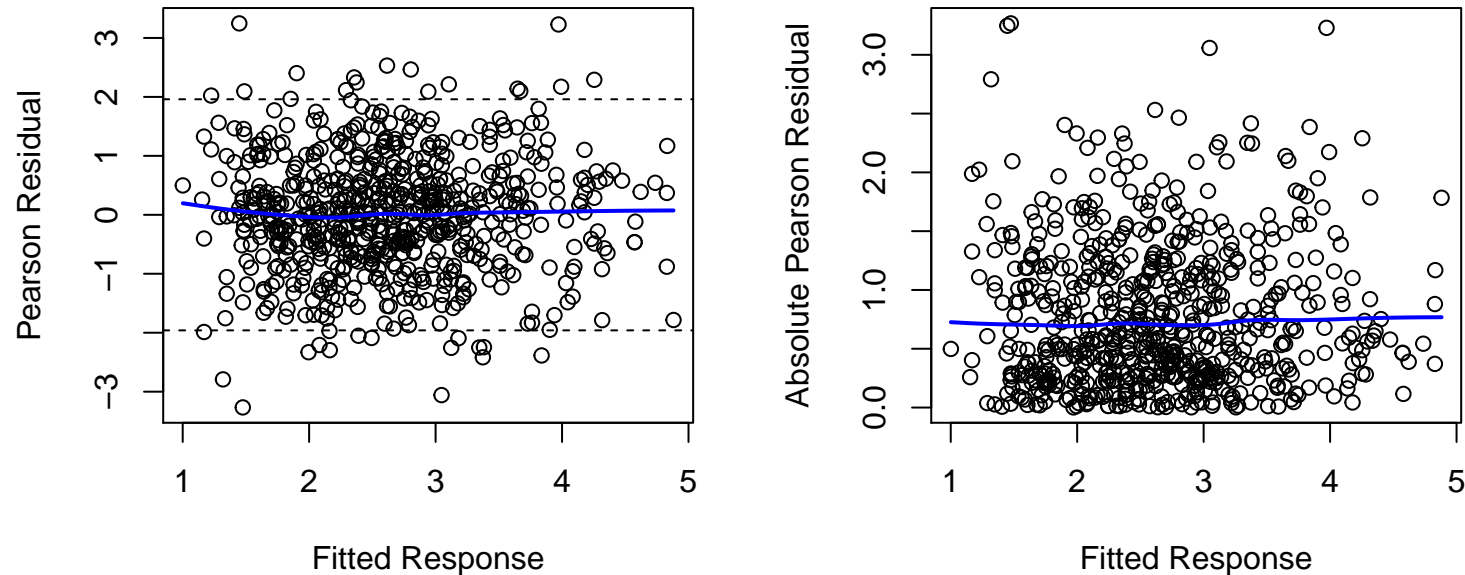
```
[1] 0.02015759
```


A Quasi-Likelihood Example

Computation and plotting of the Pearson residuals

```
> resid_FEV_Quasi2 <- residuals(FEV_Quasi2, type = "pearson")/sqrt(summary(FEV_Quasi2)$dispersion)
> fitted_FEV_Quasi2 <- fitted(FEV_Quasi2)
> limits <- range(data$FEV, fitted_FEV_Quasi2)
> par(mfrow = c(1,2))
> plot(fitted_FEV_Quasi2, resid_FEV_Quasi2,
+      ylab = "Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi2, resid_FEV_Quasi2), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
> plot(fitted_FEV_Quasi2, abs(resid_FEV_Quasi2),
+      ylab = "Absolute Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi2, abs(resid_FEV_Quasi2)), lwd = 2, col = "blue")
```

A Quasi-Likelihood Example



The Pearson residuals indicate that the assumption $\text{var}(Y_i \mid \mathbf{x}_i) = \alpha \mu_i^2$ seems reasonable

Question: how does this compare with the gamma GLM?

A Quasi-Likelihood Example

The gamma GLM and the quasi-likelihood approach with $\text{var}(Y_i \mid \mathbf{x}_i) = \alpha \mu_i^2$ are indistinguishable from their estimating functions!

```
> identical(summary(FEV_Quasi2)$dispersion, summary(FEV_Gamma)$dispersion)
```

```
[1] TRUE
```

```
> identical(fitted(FEV_Quasi2), fitted(FEV_Gamma))
```

```
[1] TRUE
```

```
> identical(vcov(FEV_Quasi2), vcov(FEV_Gamma))
```

```
[1] TRUE
```

A Quasi-Likelihood Example

One more quasi-likelihood approach: let us now consider

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha V(\mu_i) = \alpha \mu_i^3 = \alpha E(Y_i | \mathbf{x}_i)^3$$

```
> FEV_Quasi3 <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   family = quasi(link="identity", variance="mu^3"),  
+   data=data)
```

In this case, $\hat{\alpha}$ is

```
> summary(FEV_Quasi3)$dispersion
```

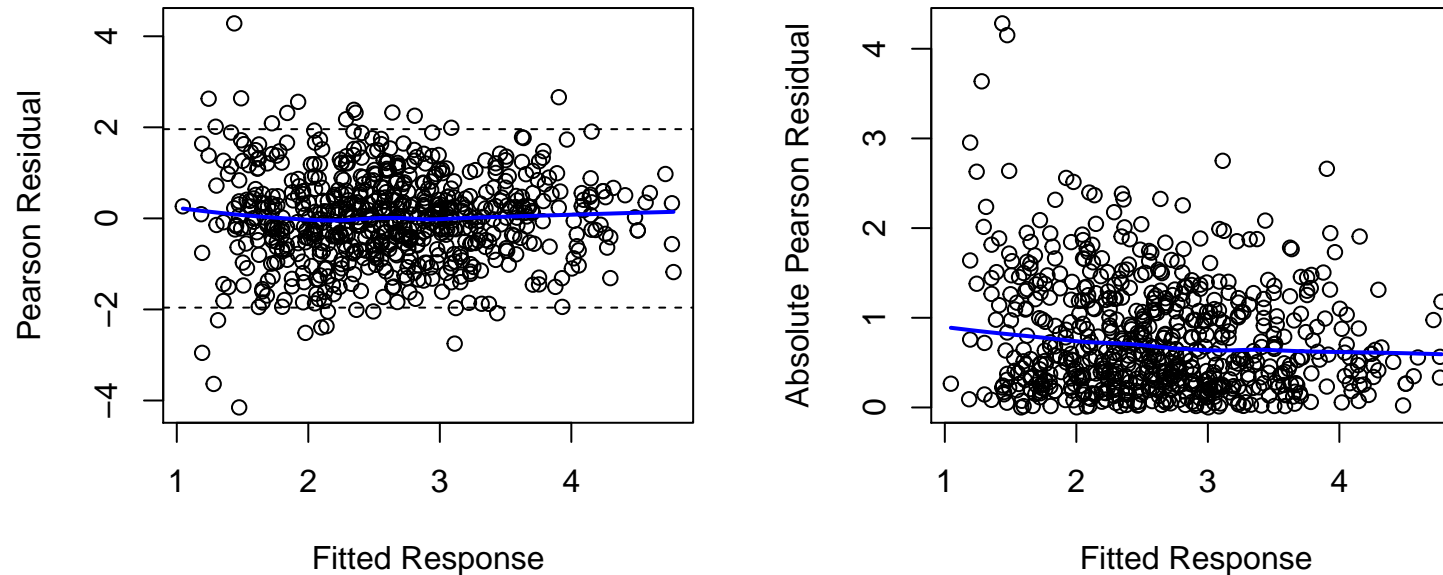
```
[1] 0.00845347
```

A Quasi-Likelihood Example

Computation and plotting of the Pearson residuals

```
> resid_FEV_Quasi3 <- residuals(FEV_Quasi3, type = "pearson")/sqrt(summary(FEV_Quasi3)$dispersion)
> fitted_FEV_Quasi3 <- fitted(FEV_Quasi3)
> limits <- range(data$FEV, fitted_FEV_Quasi3)
> par(mfrow = c(1,2))
> plot(fitted_FEV_Quasi3, resid_FEV_Quasi3,
+      ylab = "Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi3, resid_FEV_Quasi3), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
> plot(fitted_FEV_Quasi3, abs(resid_FEV_Quasi3),
+      ylab = "Absolute Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_FEV_Quasi3, abs(resid_FEV_Quasi3)), lwd = 2, col = "blue")
```

A Quasi-Likelihood Example



These Pearson residuals now indicate that assuming $\text{var}(Y_i \mid \mathbf{x}_i) = \alpha \mu_i^3$ is not appropriate, as it makes the variance of the responses grow too fast, which can be seen from the variability of the Pearson residuals decreasing as a function of the fitted means

A Quasi-Likelihood Example

The usual table with test statistics for the individual parameters can be obtained from `summary` or from `coeftest` in the `lmtest` package

```
> round(coeftest(FEV_Quasi2),3)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.436	0.030	82.239	<2e-16	***
HeightC	0.109	0.006	17.851	<2e-16	***
HeightCSq	0.004	0.000	7.499	<2e-16	***
AgeC	0.052	0.014	3.833	<2e-16	***
Sex1	-0.017	0.038	-0.443	0.657	
Smoker1	-0.067	0.130	-0.521	0.603	
HeightC:Sex1	-0.019	0.010	-1.999	0.046	*
HeightC:Smoker1	0.029	0.049	0.596	0.551	
HeightCSq:Sex1	-0.002	0.001	-2.882	0.004	**
HeightCSq:Smoker1	-0.002	0.005	-0.421	0.674	
AgeC:Sex1	0.011	0.018	0.636	0.525	
AgeC:Smoker1	-0.031	0.027	-1.138	0.255	

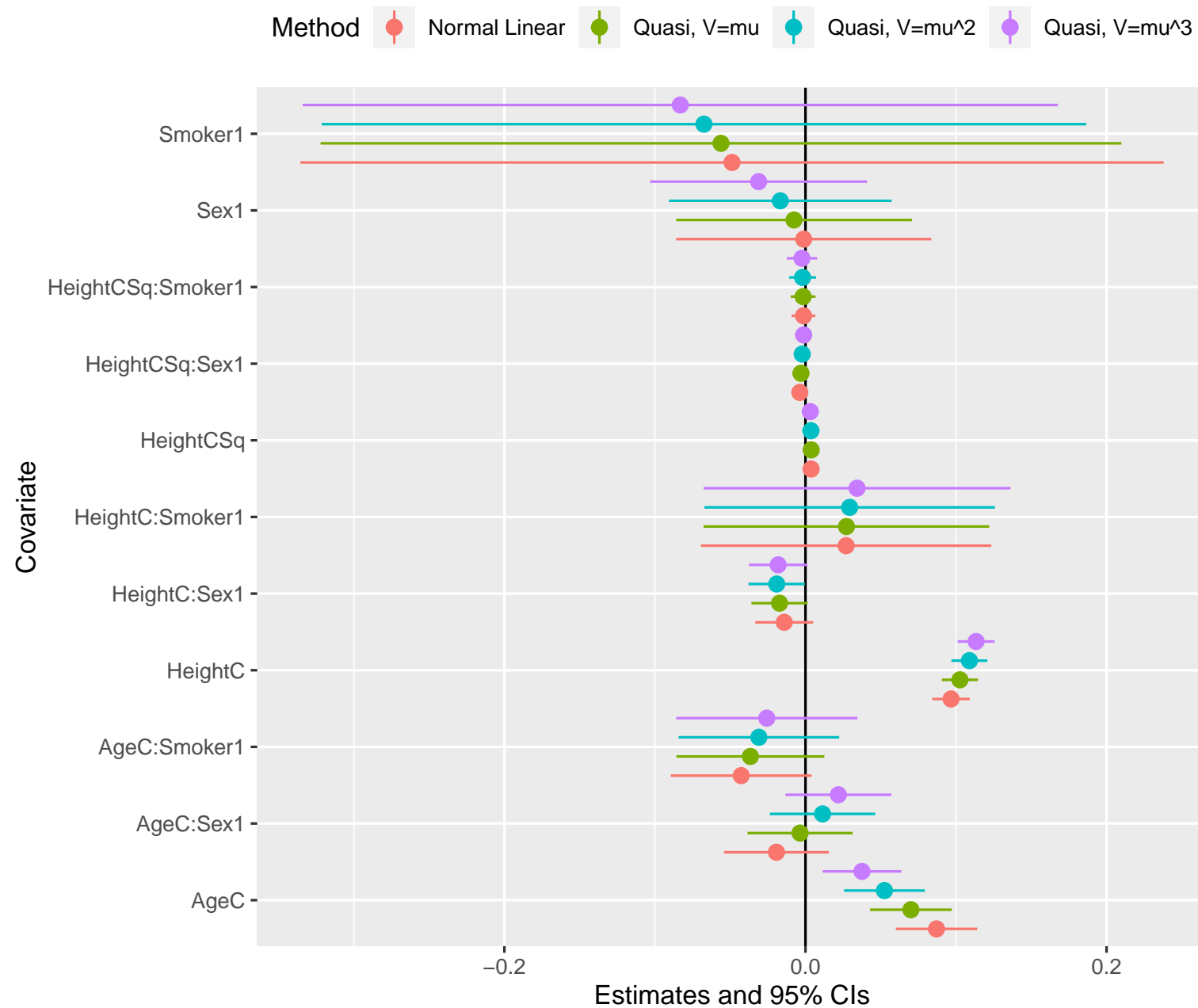
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A Quasi-Likelihood Example

Let's compare the point estimates and confidence intervals obtained from each quasi-likelihood approach

```
> normal_CIs <- data.frame(coefci(model3, vcov = vcov(model3)), method="Normal Linear")
> quasi1_CIs <- data.frame(coefci(FEV_Quasi1, vcov = vcov(FEV_Quasi1)), method="Quasi, V=mu")
> quasi2_CIs <- data.frame(coefci(FEV_Quasi2, vcov = vcov(FEV_Quasi2)), method="Quasi, V=mu^2")
> quasi3_CIs <- data.frame(coefci(FEV_Quasi3, vcov = vcov(FEV_Quasi3)), method="Quasi, V=mu^3")
> all_CIs <- rbind(normal_CIs[-1,],quasi1_CIs[-1,],quasi2_CIs[-1,],quasi3_CIs[-1,])
> all_CIs <- data.frame(all_CIs, param=rownames(normal_CIs)[-1],
+   pointest = c(model3$coeff[-1],FEV_Quasi1$coeff[-1],
+               FEV_Quasi2$coeff[-1],FEV_Quasi3$coeff[-1])
+   )
> names(all_CIs)[1:2] <- c("lower","upper")
> myplot <- ggplot(data = all_CIs, aes(x = param, y = pointest,
+   ymin = lower, ymax = upper, group = method, color = method))
> myplot + geom_hline(yintercept=0, color="black") +
+   geom_pointrange(position = position_dodge(width = 1)) +
+   theme(legend.position = "top") + coord_flip() +
+   labs(x = "Covariate", y = "Estimates and 95% CIs", color = "Method")
```


A Quasi-Likelihood Example



A Quasi-Likelihood Example

We had the following question: is there any association between FEV and smoking after controlling for other variables?

Based on the model for the mean

$$\begin{aligned} E(\text{FEV1} \mid \mathbf{X}) = & \beta_0 + \beta_1 \text{Height} + \beta_2 \text{Height}^2 + \beta_3 \text{Age} + \\ & (\beta_4 + \beta_5 \text{Height} + \beta_6 \text{Height}^2 + \beta_7 \text{Age}) I(\text{Sex}=\text{Female}) + \\ & (\beta_8 + \beta_9 \text{Height} + \beta_{10} \text{Height}^2 + \beta_{11} \text{Age}) I(\text{Smoker}=\text{Yes}) \end{aligned}$$

we can test the null hypothesis

$$H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$$

to try to answer the question

Let us do this comparing the fit of the quasi-likelihood with quadratic variance FEV_Quasi2 with a reduced model that does not include the Smoker variable, say FEV_Quasi2_red:

```
> FEV_Quasi2_red <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*Sex,  
+   family = quasi(link="identity", variance="mu^2"),  
+   data=data)
```

A Quasi-Likelihood Example

```
> waldtest(FEV_Quasi2_red, FEV_Quasi2, test = "Chisq") # Wald test
```

Wald test

Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex

Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)

	Res.Df	Df	Chisq	Pr(>Chisq)
1	646			
2	642	4	3.0416	0.5509

```
> anova(FEV_Quasi2_red, FEV_Quasi2, test = "Rao") # Score test
```

Analysis of Deviance Table

Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex

Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)

	Resid. Df	Resid. Dev	Df	Deviance	Rao	Pr(>Chi)
1	646	13.462				
2	642	13.404	4	0.058237	0.056386	0.5923

Therefore the larger model is *not* a better fit, and so we lack evidence that FEV is associated with smoking status after controlling for other variables in this model

Note that the likelihood-ratio test does not make sense under quasi-likelihood, as there is no likelihood

Sandwich-Based Intervals and Tests

We can base our inferences on sandwich estimators, so as not to rely on correct specification of the mean-variance relationship

The R package `sandwich` contains functions to extract the pieces required to compute sandwich estimators from the fit of the function `glm`, `lm` and others, see Zeileis (2004, 2006, *Journal of Statistical Software*)

The package `sandwich` still relies on the regression function being correctly specified

It computes the most general version of $\hat{\mathbf{B}}_n$ for uncorrelated data as

$$\hat{\mathbf{B}}_n = \frac{1}{n} \hat{\mathbf{D}}^T \hat{\mathbf{V}}^{-1} \text{diag}\{[Y_i - \mu_i(\hat{\beta}_n)]^2\} \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}},$$

which is known as the Eicker-Huber-White (HC0) estimator, and the most general version of $\hat{\mathbf{A}}_n$ under correct specification of the mean model

$$\hat{\mathbf{A}}_n = \frac{1}{n} \hat{\mathbf{D}}^T \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}}$$

See slides9.pdf, p. 39

Sandwich-Based Intervals and Tests

```
> library(sandwich) # Default is the Eicker-Huber-White estimator
> coeftest(FEV_Quasi2, vcov = sandwich) # Same as vcov = vcovHC(FEV_Quasi2, type = "HCO")
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.43615775	0.03226790	75.4979	< 2.2e-16	***
HeightC	0.10888453	0.00556352	19.5712	< 2.2e-16	***
HeightCSq	0.00362528	0.00064325	5.6359	1.742e-08	***
AgeC	0.05244215	0.01275235	4.1124	3.917e-05	***
Sex1	-0.01673334	0.03929730	-0.4258	0.67024	
Smoker1	-0.06745471	0.14136928	-0.4772	0.63325	
HeightC:Sex1	-0.01910880	0.00987039	-1.9360	0.05287	.
HeightC:Smoker1	0.02935413	0.04568911	0.6425	0.52056	
HeightCSq:Sex1	-0.00217012	0.00088874	-2.4418	0.01461	*
HeightCSq:Smoker1	-0.00192167	0.00382853	-0.5019	0.61571	
AgeC:Sex1	0.01136177	0.01707719	0.6653	0.50585	
AgeC:Smoker1	-0.03095755	0.02233657	-1.3860	0.16576	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The argument type can also be specified as HC1, which replaces $e_i^2 := [Y_i - \mu_i(\hat{\beta}_n)]^2$ with $\frac{n}{n-k-1}e_i^2$ in $\hat{\mathbf{B}}_n$; other options are available, see Zeileis (2004)

Sandwich-Based Intervals and Tests

Compare these two:

Relies on correct mean-variance specification

```
> waldtest(FEV_Quasi2_red, FEV_Quasi2, test = "Chisq")
```

Wald test

```
Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex
Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)
  Res.Df Df  Chisq Pr(>Chisq)
1      646
2      642  4 3.0416    0.5509
```

Doesn't assume correct variance specification

```
> waldtest(FEV_Quasi2_red, FEV_Quasi2, test = "Chisq", vcov = sandwich)
```

Wald test

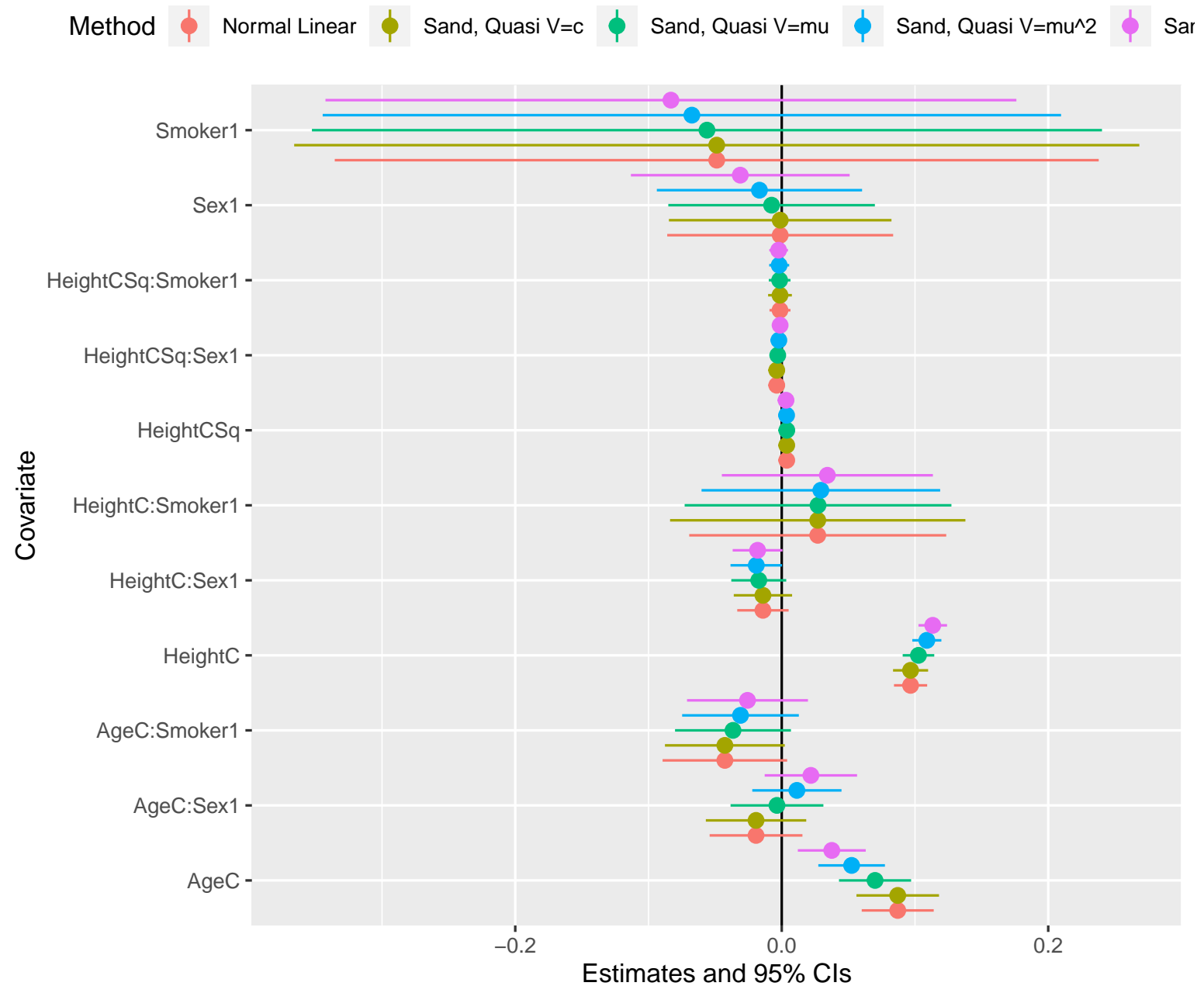
```
Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex
Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)
  Res.Df Df  Chisq Pr(>Chisq)
1      646
2      642  4 3.9253    0.4162
```

Sandwich-Based Intervals and Tests

Let's compare the point estimates and confidence intervals obtained from each quasi-likelihood approach

```
> normal_CIs <- data.frame(coefci(model3, vcov = vcov(model3)), method="Normal Linear")
> sand_quasi0_CIs <- data.frame(coefci(model3, vcov = sandwich), method="Sand, Quasi V=c")
> sand_quasi1_CIs <- data.frame(coefci(FEV_Quasi1, vcov = sandwich), method="Sand, Quasi V=mu")
> sand_quasi2_CIs <- data.frame(coefci(FEV_Quasi2, vcov = sandwich), method="Sand, Quasi V=mu^2")
> sand_quasi3_CIs <- data.frame(coefci(FEV_Quasi3, vcov = sandwich), method="Sand, Quasi V=mu^3")
> all_CIs <- rbind(normal_CIs[-1,], sand_quasi0_CIs[-1,], sand_quasi1_CIs[-1,],
+               sand_quasi2_CIs[-1,], sand_quasi3_CIs[-1,])
> all_CIs <- data.frame(all_CIs, param=rownames(normal_CIs)[-1],
+               pointtest = c(model3$coeff[-1], model3$coeff[-1], FEV_Quasi1$coeff[-1],
+               FEV_Quasi2$coeff[-1], FEV_Quasi3$coeff[-1])
+               )
> names(all_CIs)[1:2] <- c("lower", "upper")
> myplot <- ggplot(data = all_CIs, aes(x = param, y = pointtest,
+               ymin = lower, ymax = upper, group = method, color = method))
> myplot + geom_hline(yintercept=0, color="black") +
+       geom_pointrange(position = position_dodge(width = 1)) +
+       theme(legend.position = "top") + coord_flip() +
+       labs(x = "Covariate", y = "Estimates and 95% CIs", color = "Method")
```

Sandwich-Based Intervals and Tests



Final Comments

How can we use all of this in practice?

- Think carefully about the regression function specification; make sure it makes scientific sense
- Try to choose a reasonable $V(\cdot)$ function for your variance, to base your estimating function
- Use sandwich estimators, just in case you are wrong!
- If you want to rely on stronger assumptions (say in small samples, for efficiency reasons), make sure you can justify them given your data