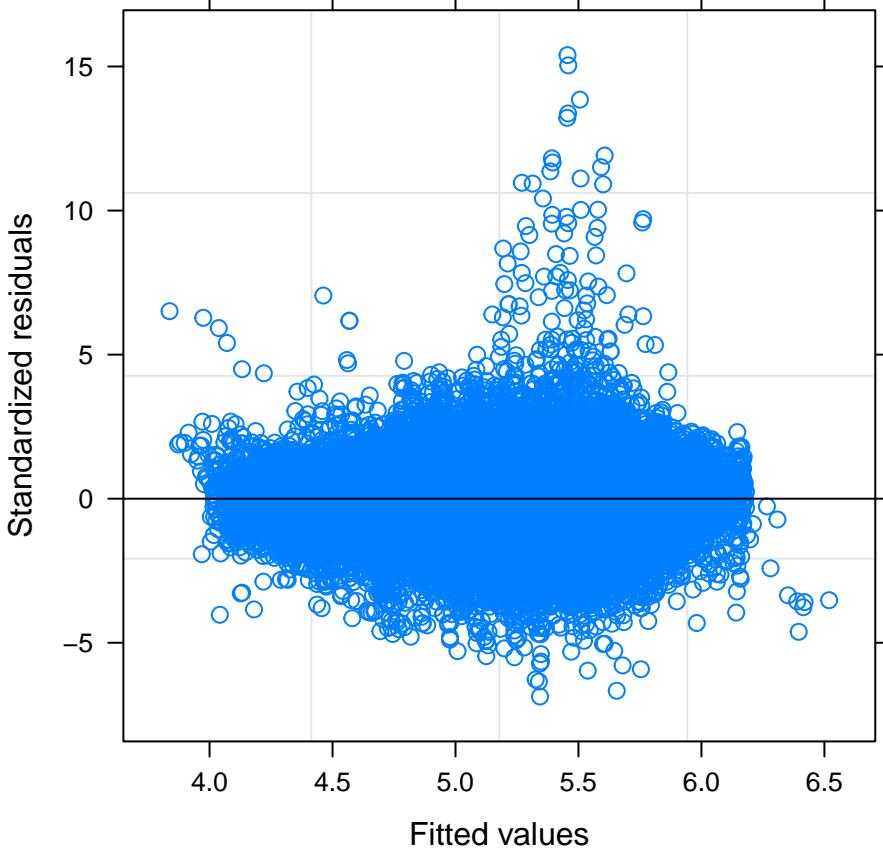


```
plot(fit1.log)
```



```

fit2 <- gls(Avg.Fare ~ Quarter + Avg.Pass, method="REML",
             data = rout)
anova(fit1,fit2) #The Random intercept route is significant

##      Model df     AIC     BIC   logLik   Test  L.Ratio p-value
## fit1     1 5 1100692 1100740 -550340.8
## fit2     2 4 1237751 1237789 -618871.5 1 vs 2 137061.5 <.0001

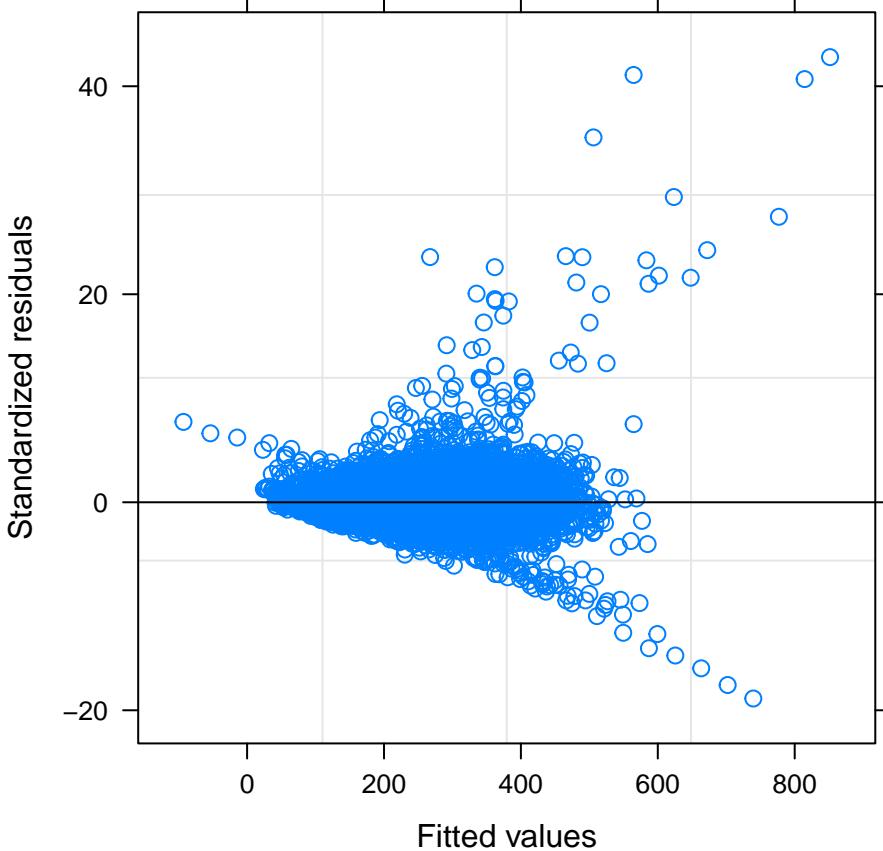
fit.1 <-lme(Avg.Fare ~ Quarter + Avg.Pass,
            random = ~ Quarter|Route, method="REML",
            data = rout) #Consider a random Slope effect
summary(fit.1)

## Linear mixed-effects model fit by REML
## Data: rout
##      AIC      BIC    logLik
## 1077630 1077698 -538808.2
##
## Random effects:
##  Formula: ~Quarter | Route
##  Structure: General positive-definite, Log-Cholesky parametrization

```

```
##           StdDev   Corr
## (Intercept) 81.679782 (Intr)
## Quarter      2.376023 -0.699
## Residual     30.249884
##
## Fixed effects: Avg.Fare ~ Quarter + Avg.Pass
##                   Value Std.Error    DF   t-value p-value
## (Intercept) 218.32204 1.2948835 104423 168.60362      0
## Quarter      -0.52241 0.0387505 104423 -13.48139      0
## Avg.Pass     -0.04247 0.0010394 104423 -40.86151      0
## Correlation:
##             (Intr) Quartr
## Quarter   -0.686
## Avg.Pass  -0.162 -0.013
##
## Standardized Within-Group Residuals:
##           Min        Q1        Med        Q3        Max
## -18.86909007 -0.43831676 -0.02782324  0.38196193  42.79907379
##
## Number of Observations: 108602
## Number of Groups: 4177

plot(fit.1) #Model indicates a worse fit than before
```



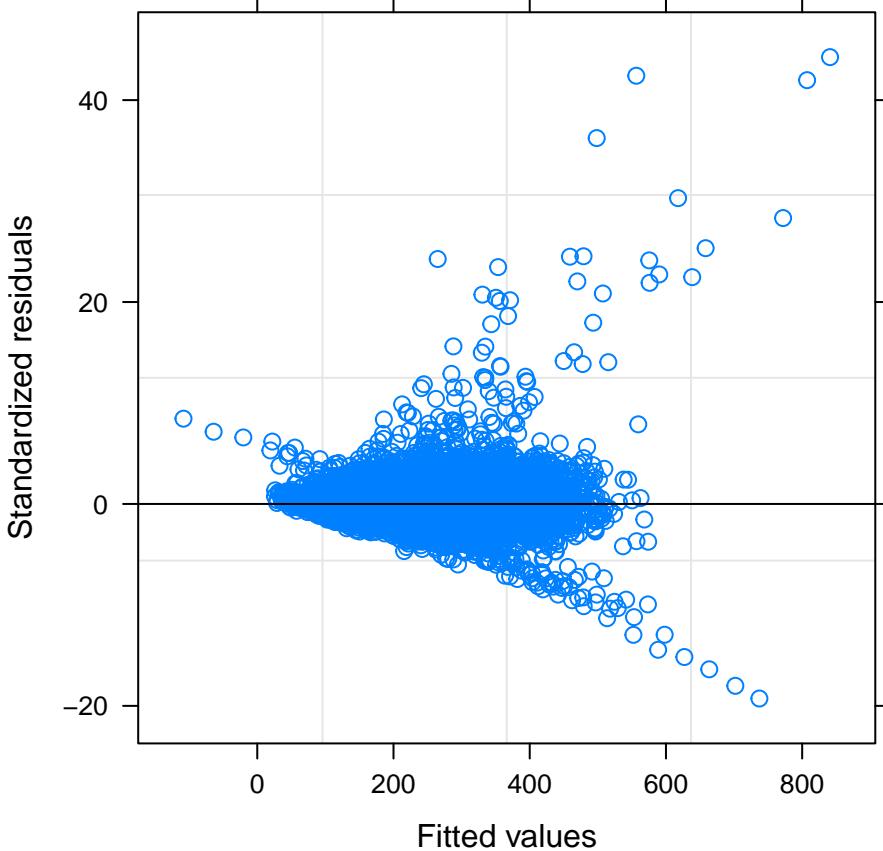
```

anova(fit1, fit.1)

##      Model df     AIC     BIC   logLik   Test L.Ratio p-value
## fit1     1 5 1100692 1100740 -550340.8
## fit.1    2 7 1077630 1077698 -538808.2 1 vs 2 23065.11 <.0001

#Consider higher order terms
fit.2 <-lme(Avg.Fare ~ Quarter + Avg.Pass+ I(Quarter^2) + I(Avg.Pass^2),
             random = ~ Quarter|Route, method="REML",
             data = rout)
plot(fit.2)

```



```

qqnorm(resid(fit.2)) # Not much change
#Random Route effect is important

#We will consider joining the quarters together
x <- 1:26 %% 4
x[x==0] <- 4
rout$Y.Q <- rep(x, 4177)

sub3 <- groupedData(formula = Avg.Fare ~ Y.Q | Route, data = rout)
sub4 <- sub3[1:max(which(rout[,1] == "50")),] #First 50 routes
#Subset for each Quarters
xyplot(sub4$Avg.Fare ~ sub4$Y.Q,
       type="b",
       group=sub4$Route,
       data=sub4,
       auto.key =F
)
fit.var <- lme(Avg.Fare ~ Quarter + Avg.Pass,
               random = ~ 1 | Route,
               weights = varIdent(form= ~ 1 | Y.Q),

```

```
    data=sub4)

fit.sl <- lme(Avg.Fare ~ Quarter + Avg.Pass,
               random = ~ Quarter| Route,
               data=rand.samp)
#Different variance at each time points with only random intercept
plot(fit.var)

plot(sub4) #Does not indicate a similar pattern throughout all routes

# Create yearly data
temp <- lapply(1:6, rep, 4)
temp <- c(temp[[1]],temp[[2]],temp[[3]],temp[[4]],temp[[5]],temp[[6]], 7, 7)
rout$Year <- temp

sub4 <- groupedData(formula = Avg.Fare ~ Year| Route, data = rout)
sub5 <- sub4[1:max(which(rout[,1] == "50")),] #First 50 routes
#Subset for each Quarters
xyplot(sub5$Avg.Fare ~ sub5$Year,
       type="b",
       group=sub5$Route,
       data=sub5,
       auto.key =F
)
# route one individual,
#nested within reigon,
# Time is quarters
# route i within reigon

#Consider a Yearly model
fit3 <-lme(Avg.Fare ~ Year + Avg.Pass,
            random = ~ 1|Route, method="REML",
            data = rout)

fit4 <- gls(Avg.Fare ~ Year + Avg.Pass, method="REML",
             data = rout)
anova(fit3,fit4)

##      Model df      AIC      BIC logLik   Test L.Ratio p-value
## fit3     1 5 1100348 1100395 -550168.7
## fit4     2 4 1237668 1237706 -618829.9 1 vs 2 137322.4 <.0001

#Random Rout effect on Route is significant
plot(fit3) #Still does not indicate a good fit

#model 3
fit5 <-lme(Avg.Fare ~ Year + Avg.Pass,
            random = ~ 1|Route, method="ML",
            data = rout)

fit6 <-lme(Avg.Fare ~ Y.Q + Avg.Pass,
```

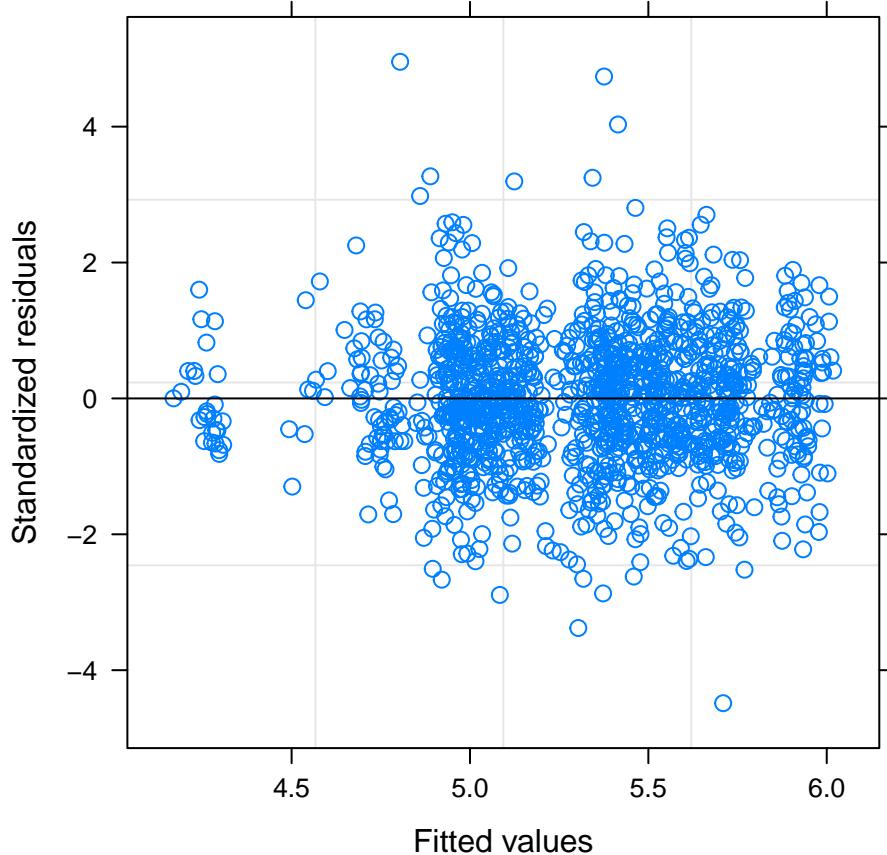
```

random = ~ 1|Route, method="ML",
data = rout)
anova(fit5, fit6) #Almost the same

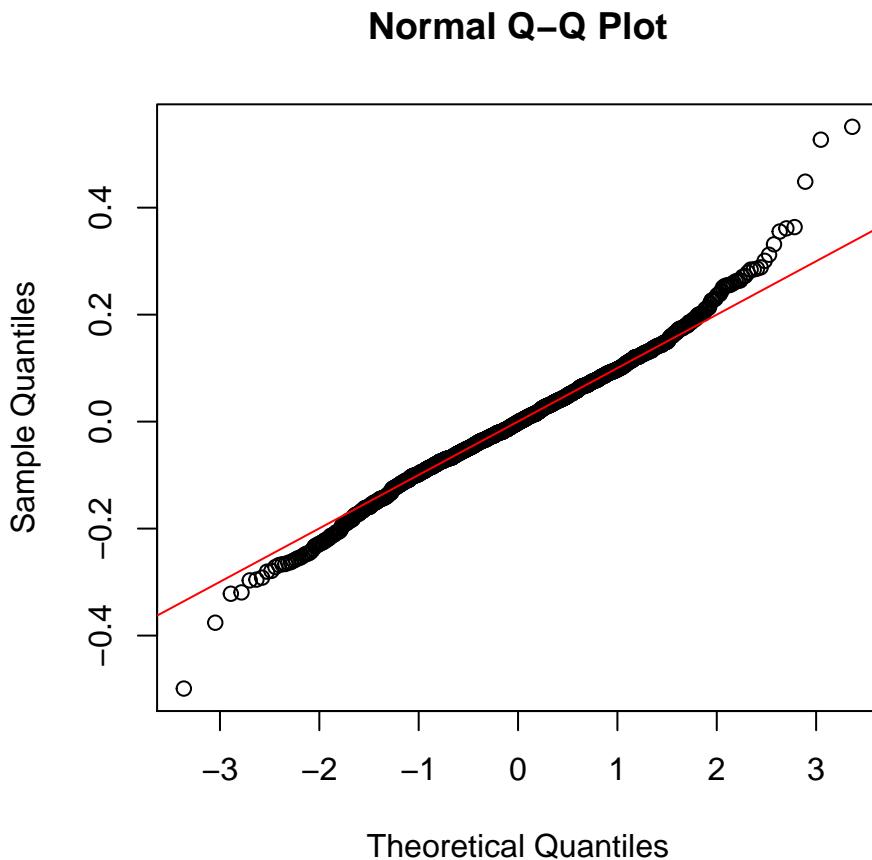
##      Model df     AIC     BIC   logLik
## fit5     1 5 1100333 1100381 -550161.7
## fit6     2 5 1100988 1101036 -550489.1

log.lme <- lme(log(Avg.Fare) ~ Quarter + I(Quarter^2)+ Avg.Pass + I(Avg.Pass^2),
                 data = sub2, random = ~1|Route)
plot(log.lme)
par(mfrow=c(1,1))

```



```
qqnorm(resid(log.lme)); qqline(resid(log.lme), col = "red")
```



```

summary(log.lme)

## Linear mixed-effects model fit by REML
## Data: sub2
##      AIC      BIC   logLik
## -1654.593 -1618.429  834.2964
##
## Random effects:
## Formula: ~1 | Route
##          (Intercept) Residual
## StdDev:  0.3362234 0.111256
##
## Fixed effects: log(Avg.Fare) ~ Quarter + I(Quarter^2) + Avg.Pass + I(Avg.Pass^2)
##                  Value Std.Error DF t-value p-value
## (Intercept) 5.385867 0.04982064 1246 108.10514     0
## Quarter     0.015383 0.00172770 1246   8.90384     0
## I(Quarter^2) -0.000568 0.00006217 1246  -9.12952     0
## Avg.Pass    -0.001869 0.00016305 1246 -11.46461     0
## I(Avg.Pass^2) 0.000001 0.00000010 1246   6.68229     0
## Correlation:

```

```

##           (Intr) Quartr I(Q^2) Avg.Ps
## Quarter      -0.151
## I(Quarter^2)  0.124 -0.971
## Avg.Pass     -0.214 -0.130  0.146
## I(Avg.Pass^2)  0.156  0.142 -0.142 -0.852
##
## Standardized Within-Group Residuals:
##      Min       Q1       Med       Q3       Max
## -4.48583786 -0.60389065 -0.01403567  0.60718032  4.95395712
##
## Number of Observations: 1300
## Number of Groups: 50

p <- ggplot(sub2, aes(x=Quarter, y=Avg.Fare, colour=Route)) +
  geom_point(size=.5) +
  geom_line(aes(y=exp(fitted(log.lme)), group=Route), size = .5) +
  theme_bw(base_size=22)
print(p)

```

