Model the relation between oil/gas production and monthly beta measurement

Longxiang Li

2018-10-16

Contents

Introduction

The primary goal of this study is to explore the correlation between monthly mean β and the monthly production of natural gas and oil within 25km of the RadNet monitor.

Background

- β level can also be used as a proxy to particulate radioactivity in Maggie's NAS paper.
- The relation between β and radon is not as direct as that between Pb210 and radon.
- We want to use β as another proxy to prove the influence of oil/gas production on local radon.
- Since we have monthly mean β measurment (original frequency is every 5-7 days), we have more data and power to detect the correlation.

Data

Study Period: from 2010 to 2016:

In this way, we can track the trend of oil/gas production influenced by the viberation of price.

Study Region:

Lower 48 states of the U.S

Data Sources:

Oil/Gas production data:

We colleced drilling information from drillinginfo.com. From the database, we extracted the monthly gas/oil production during the study period, both horizontal and vertical drillings are included. Then we have monthly gas/oil production, number of active gas/oil wells, categorical and uncategorical within a radius of 25km from the assumed location of RadNet monitors.

β radiation data:

We got β radiation from 139 RadNet monitors every 5-7 days. To match the frequency of oil/gas production data, we calculated the monthly mean.

USGS radiation potential map:

We also included USGS aeroradiometric data (including Potassium, Thorium and Uranium). These information can be applied as a proxy to the cosmic radiation source.

EPA radon zone map:

We downloaded the EPA radon map data and join it to the assumed location of RadNet monitor. All counties of the U,S are catergorized into three classes ranging from 1 with the highest radon level and 3 with the lowest. The background radon level is calculated based on soil type, weather and other information.

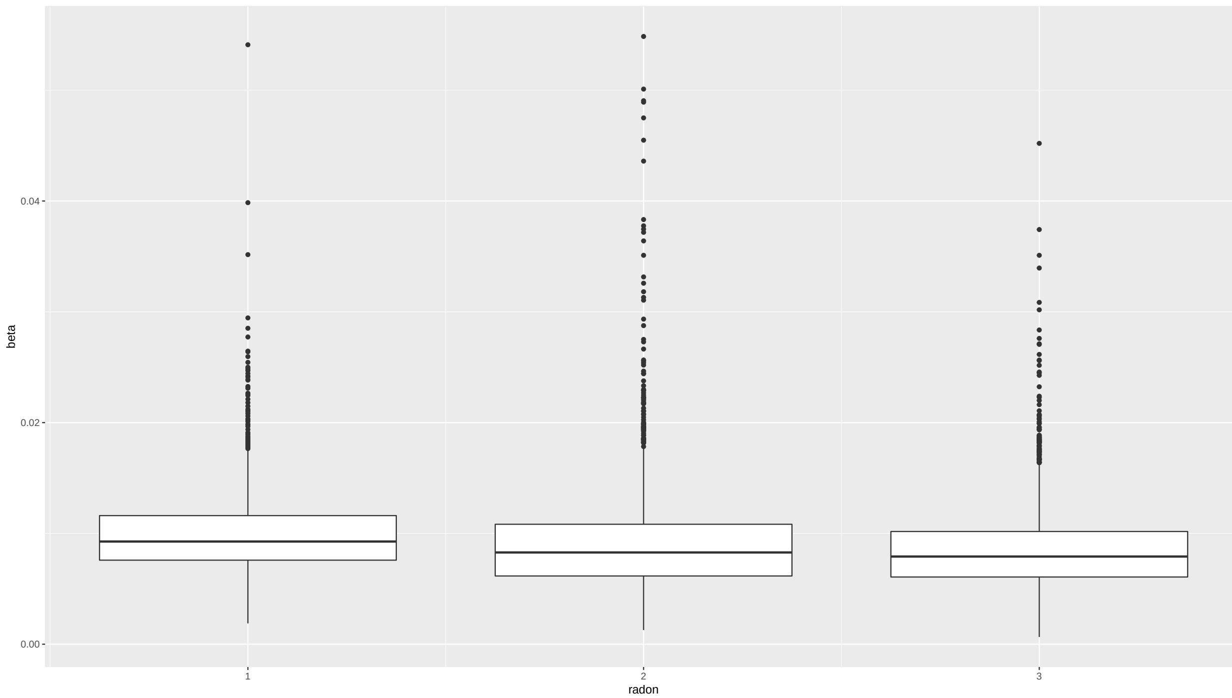
Descriptive Analysis of the data

From the boxplot of β against radon zones, we can find that β is also negatively related with radon zones. This's similar with Pb-210, but the variation of β is wilder.

From the boxplotof β against oil/gas field, we can find that β in oil/gas filed is higher than that non-field areas.

And similar with what we found in the Pb-210 dataset, the radon level of gas/oil field is always lower than that of non-field areas.

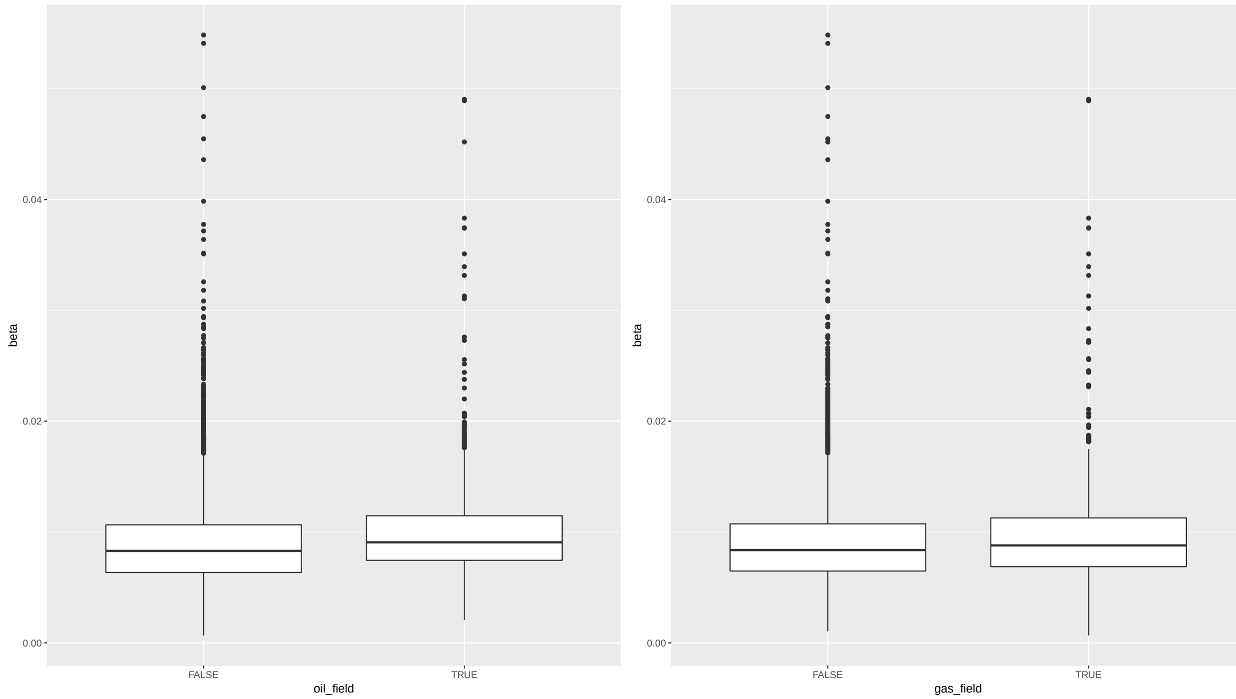
So, we can guess that the production of oil/gas may change the local relation between the EPA radon zone estimate and the β (a proxy to the real radon level).




```
##
## Call:
## lm(formula = beta ~ oil_field, data = rad_qs_zones)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.008215 -0.002512 -0.000607  0.001738  0.045978
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.872e-03  4.798e-05 184.913   <2e-16 ***
## oil_fieldTRUE 1.036e-03  1.166e-04   8.889   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00402 on 8449 degrees of freedom
## Multiple R-squared:  0.009265,   Adjusted R-squared:  0.009147
## F-statistic: 79.01 on 1 and 8449 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = beta ~ gas_field, data = rad_qs_zones)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.008764 -0.002506 -0.000614  0.001777  0.045881
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.969e-03  4.826e-05 185.844  < 2e-16 ***
## gas_fieldTRUE 4.515e-04  1.160e-04   3.891 0.000101 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004035 on 8449 degrees of freedom
## Multiple R-squared:  0.001789,   Adjusted R-squared:  0.001671
## F-statistic: 15.14 on 1 and 8449 DF,  p-value: 0.0001006
```

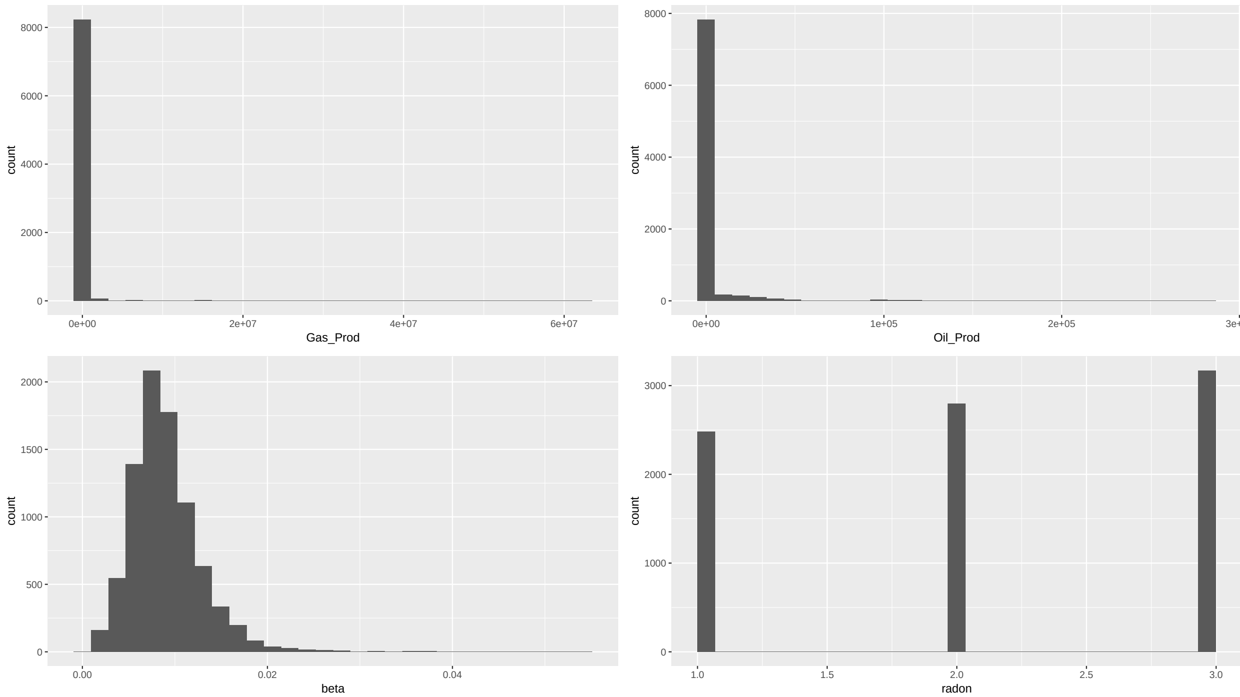



```
##
## Call:
## lm(formula = beta ~ oil_field, data = rad_qs_zones)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.008215 -0.002512 -0.000607  0.001738  0.045978
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.872e-03  4.798e-05 184.913   <2e-16 ***
## oil_fieldTRUE 1.036e-03  1.166e-04   8.889   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00402 on 8449 degrees of freedom
## Multiple R-squared:  0.009265,   Adjusted R-squared:  0.009147
## F-statistic: 79.01 on 1 and 8449 DF,  p-value: < 2.2e-16
```

```
## # A tibble: 2 x 2
##   oil_field mean_radon
##   <lgl>          <dbl>
## 1 FALSE           2.00
## 2 TRUE            2.48
```

```
## # A tibble: 2 x 2
##   gas_field mean_radon
```

```
##   <lgl>           <dbl>
## 1 FALSE           2.00
## 2 TRUE            2.50
```



Models

Mixed effects models are used in this report to model the correlation between our variable of interest and the β. Our variables of interest are always set as fixed effect while random intercepts are assigned to each RadNet monitor. In addition, radon zone is also set as fixed effect. To check the significance of our fixed effect, a bootstrp confidence interval is calculated. In addition, a likelihood-ratio test is also applied here.

Gross Oil Production

Without log-transformation, gross oil is weakly correlated with local β level. After the log-tranformation, gross oil production is not significantly correlated with the local β level.

```
lm_basic<-lmer(beta~radon+Thmeans+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
lm_oil_prod<-lmer(beta~radon+Thmeans+Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
confint(lm_oil_prod,parm ="Oil_Prod",method="boot",boot.type="perc")
```

```
##                 2.5 %       97.5 %
## Oil_Prod 1.14897e-09 2.674407e-08
```

```
anova(lm_basic,lm_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
```

```
## lm_basic: beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
## lm_oil_prod: beta ~ radon + Thmeans + Oil_Prod + MONTH + YEAR + (1 | city_state)
##             Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_basic     7 -71726 -71676  35870   -71740
## lm_oil_prod  8 -71729 -71673  35873   -71745 5.5294      1     0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lm_log_basic<-lmer(log_beta~radon+Thmeans+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
lm_log_oil_prod<-lmer(log_beta~radon+Thmeans+Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
confint(lm_log_oil_prod,parm="Oil_Prod",method="boot")
```

```
##                  2.5 %       97.5 %
## Oil_Prod -5.364411e-07 1.808183e-06
```

```
anova(lm_log_basic,lm_log_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
## lm_log_basic: log_beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
## lm_log_oil_prod: log_beta ~ radon + Thmeans + Oil_Prod + MONTH + YEAR + (1 | city_state)
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_log_basic     7 5387.0 5436.3 -2686.5   5373.0
## lm_log_oil_prod  8 5388.1 5444.5 -2686.1   5372.1 0.8509      1     0.3563
```

Horizontal Oil Production

Without log-transformation, horizontal oil production is not correlated with local β level. After the log-tranformation, gross oil production is not significantly correlated with the local β level.

```
lm_h_oil_prod<-lmer(beta~radon+Thmeans+H_Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
confint(lm_h_oil_prod,parm ="H_Oil_Prod",method="boot",boot.type="perc")
```

```
##                    2.5 %       97.5 %
## H_Oil_Prod -2.846813e-08 3.312239e-08
```

```
anova(lm_basic,lm_h_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
## lm_basic: beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
## lm_h_oil_prod: beta ~ radon + Thmeans + H_Oil_Prod + MONTH + YEAR + (1 | city_state)
##               Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_basic       7 -71726 -71676  35870   -71740
## lm_h_oil_prod  8 -71724 -71667  35870   -71740 0.0071      1      0.933
```

```
lm_log_basic<-lmer(log_beta~radon+Thmeans+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
lm_log_oil_prod<-lmer(log_beta~radon+Thmeans+Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML
```

```
## Warning: Some predictor variables are on very different scales: consider
```

```
## rescaling
```

```r
confint(lm_log_oil_prod,parm="Oil_Prod",method="boot")
```

```
##                   2.5 %       97.5 %
## Oil_Prod -6.180325e-07 1.720671e-06
```

```r
anova(lm_log_basic,lm_log_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
## lm_log_basic: log_beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
## lm_log_oil_prod: log_beta ~ radon + Thmeans + Oil_Prod + MONTH + YEAR + (1 | city_state)
##                  Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_log_basic     7 5387.0 5436.3 -2686.5   5373.0
## lm_log_oil_prod  8 5388.1 5444.5 -2686.1   5372.1 0.8509      1     0.3563
```

Vertical Oil Production

Without log-transformation, vertical oil is significantly correlated with local β level. After the log-tranformation, gross oil production is not significantly correlated with the local β level.

```r
lm_v_oil_prod<-lmer(beta~radon+Thmeans+V_Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,REML = T)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```r
confint(lm_v_oil_prod,parm ="V_Oil_Prod",method="boot",boot.type="perc")
```

```
##                  2.5 %       97.5 %
## V_Oil_Prod 3.6687e-09 3.102379e-08
```

```r
anova(lm_basic,lm_v_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
## lm_basic: beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
## lm_v_oil_prod: beta ~ radon + Thmeans + V_Oil_Prod + MONTH + YEAR + (1 | city_state)
##                Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_basic       7 -71726 -71676  35870   -71740
## lm_v_oil_prod  8 -71730 -71674  35873   -71746 6.7226      1    0.00952 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lm_log_v_oil_prod<-lmer(log_beta~radon+Thmeans+V_Oil_Prod+MONTH+YEAR+(1|city_state),data=rad_qs_zones,R
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```r
confint(lm_log_v_oil_prod,parm="V_Oil_Prod",method="boot")
```

```
##                    2.5 %       97.5 %
## V_Oil_Prod -6.404005e-07 1.827545e-06
```

```r
anova(lm_log_basic,lm_log_v_oil_prod)
```

```
## Data: rad_qs_zones
## Models:
## lm_log_basic: log_beta ~ radon + Thmeans + MONTH + YEAR + (1 | city_state)
```

```
## lm_log_v_oil_prod: log_beta ~ radon + Thmeans + V_Oil_Prod + MONTH + YEAR + (1 |
## lm_log_v_oil_prod:     city_state)
##                   Df  AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## lm_log_basic       7 5387 5436.3 -2686.5     5373
## lm_log_v_oil_prod  8 5388 5444.3 -2686.0     5372 0.9734      1     0.3238
```