

HW 2

Q1

1.1

By the law of total expectation and law of total variance,

$$\begin{aligned}
E[(Y - f(X))^2] &= E(E[(Y - f(X))^2 | X]) \\
&= E(E(Y^2 | X) - 2f(X)E(Y | X) + E(f(X)^2)) \\
&= E(Var(Y | X) + [E(Y | X)]^2 - 2f(X)E(Y | X) + E(f(X)^2)) \\
&= E(Var(Y | X) + (E[Y | X] - f(X))^2)
\end{aligned}$$

Since $E(Var(Y | X))$ is independent of $f(X)$, and $(E[Y | X] - f(X))^2 \geq 0$, $E[(Y - f(X))^2]$ is minimized when $E[Y | X] = f(X)$. So, $\$ E[Y | X] = \text{argmin}_f E[(Y - f(X))^2] \$$.

1.2

With $LR[Y | X] = \alpha + \beta X$, we have $E[(Y - f(X))^2] = E[(Y - \alpha - \beta X)^2]$.

Taking derivative wrt α and setting it to 0 to get the optimizer, $-E[Y - \alpha - \beta X] = 0$ such that $\alpha = E(Y) - \beta E(X)$.

Taking derivative wrt β and setting it to 0 to get the optimizer, $E((Y - \alpha - \beta X)X) = 0$ such that $E(XY) - \alpha E(X) - \beta E(X^2) = 0$. So, $E(XY) - [E(Y) - \beta E(X)]E(X) - \beta E(X^2) = 0$. Therefore, $\beta = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E(X)^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$.

Therefore, $LR[Y | X] = \alpha + \beta X = \text{argmin}_{f \in \text{linear}} E[(Y - f(X))^2]$ when $\alpha = E(Y) - \beta E(X)$ and $\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$.

1.3

Since

$$\begin{aligned}
E[(E(Y | X) - f(X))^2] &= E(E(Y^2 | X) - 2f(X)E(Y | X) + E(f(X)^2)) \\
&= E(E[(Y - f(X))^2 | X]) \\
&= E[(Y - f(X))^2]
\end{aligned}$$

Therefore, $LR[Y | X] = \alpha + \beta X = \text{argmin}_{f \in \text{linear}} E[(Y - f(X))^2] = \text{argmin}_{f \in \text{linear}} E[(E(Y | X) - f(X))^2]$ when $\alpha = E(Y) - \beta E(X)$ and $\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$.

The fact that we can estimate the regression of Y on X using aggregate data instead of individual level data means that we can now use the expectation of Y, expectation of X, covariance of X, Y, and the variance of X to estimate the regression of Y on X. In other words, we only need those aggregate level information to be able to estimate the optimizers α, β and no longer need the actual data points to perform regression analysis.

1.4

Since $E[Y | X] = \text{argmin}_f E[(Y - f(X))^2]$ and $LR[Y | X] = \alpha + \beta X = \text{argmin}_{f \in \text{linear}} E[(Y - f(X))^2]$, when CEF is linear, $E[Y | X] = \text{argmin}_f E[(Y - f(X))^2] = \text{argmin}_{f \in \text{linear}} E[(Y - f(X))^2] = \alpha + \beta X = LR[Y | X]$.

Q2

2.1

a Since $P(x, y) = \frac{3(x^2+y)}{11}$, $p(x) = \int_0^1 \frac{3(x^2+y)}{11} dy = \frac{6x^2+3}{22}$. $CEF = E(Y|X) = \int_0^1 \frac{\frac{3(x^2+y)}{11}}{\frac{6x^2+3}{22}} y dy = \frac{x^2+\frac{2}{3}}{2x^2+1}$

$$b E(X) = \int_0^2 \int_0^1 x \frac{3(x^2+y)}{11} dy dx = \int_0^2 \frac{3x^3 + \frac{3x}{2}}{11} dx = \frac{15}{11}.$$

$$E(X^2) = \int_0^2 \int_0^1 x^2 \frac{3(x^2+y)}{11} dy dx = \int_0^2 \frac{3x^4 + \frac{3x^2}{2}}{11} dx = \frac{116}{55}.$$

$$E(Y) = \int_0^1 \int_0^2 y \frac{3(x^2+y)}{11} dx dy = \int_0^1 \frac{8y + 6y^2}{11} dy = \frac{6}{11}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{116}{55} - \frac{225}{121} = \frac{151}{605}$$

$$E(XY) = \int_0^2 \int_0^1 xy \frac{3(x^2+y)}{11} dy dx = \int_0^2 \frac{\frac{3}{2}x^3 + x}{11} dx = \frac{8}{11}$$

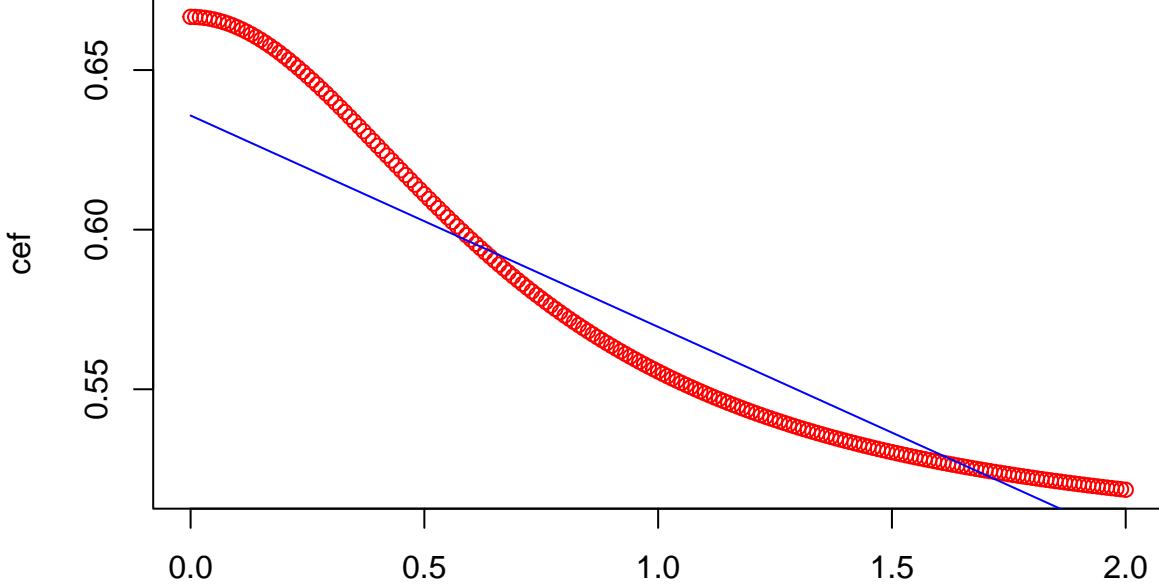
$$Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{8}{11} - \frac{15}{11} * \frac{6}{11} = -\frac{2}{121}$$

$$\beta = \frac{Cov(X, Y)}{Var(X)} = -\frac{10}{151}$$

$$\alpha = E(Y) - \beta E(X) = \frac{6}{11} + \frac{10}{151} * \frac{15}{11} = \frac{96}{151}$$

```
x = seq(0,2,0.01)
cef = (x^2+2/3)/(2*x^2 +1)
reg = 96/151+(-10/151 * x)
```

```
plot(x,cef, col = 'red')
lines(x,reg, col = 'blue')
```



c

No,
the CEF and the linear regression are not the same. The BLP(blue) approximates CEF(red) better after $x>0.5$.

2.2

a $E(Y|X) = \Sigma Y P(Y|X) = P(Y = 1|X) = P(Y = 1|X = 1)I(X = 1) + P(Y = 1|X = 0)I(X = 0) = (P(Y = 1|X = 1) - P(Y = 1|X = 0))X + P(Y = 1|X = 0) = (E[Y|X = 1] - E[Y|X = 0])X + E[Y|X = 0]$
 So, $E(Y|X)$ is linear on X.

b Since $Y|X$ only takes on 0, 1, it follows a Bernoulli distribution, with $p = P(Y = 1|X)$. Thus, $Var(Y|X) = P(Y = 1|X)(1 - P(Y = 1|X))$.

c No, it is not true in this case. As shown in a, $E(Y|X)$ is linear on X. Therefore, a linear regression can be used to estimate the CEF. The meaning of the regression coefficients would be: for intercept a, the probability of Y being 1 when X is 0; for slope b, the difference in probability when X is 1 compared with when X is 0.

2.3

a By properties of the bivariate normal distribution, $E(Y|X) = \alpha + \beta X = \mu_y - \frac{\sigma_{xy}}{\sigma_x^2} \mu_x + \frac{\sigma_{xy}}{\sigma_x^2} X$, $E(X|Y) = \alpha' + \beta' Y = \mu_x - \frac{\sigma_{xy}}{\sigma_y^2} \mu_y + \frac{\sigma_{xy}}{\sigma_y^2} Y$.

b The linear regressions will be exactly the same terms as the CEF's in (a), i.e., $BLP(Y|X) = \alpha + \beta X = \mu_y - \frac{\sigma_{xy}}{\sigma_x^2} \mu_x + \frac{\sigma_{xy}}{\sigma_x^2} X$, $BLP(X|Y) = \alpha' + \beta' Y = \mu_x - \frac{\sigma_{xy}}{\sigma_y^2} \mu_y + \frac{\sigma_{xy}}{\sigma_y^2} Y$.

c By properties of the bivariate normal distribution, $Var(Y|X) = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^4} \sigma_x^2 = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}$. It is not directly a function of X, but if X varies such that its covariance with Y or its variance changes, then $Var(Y|X)$ will also change.

d β denotes the marginal change in Y with one unit change of X. α is simply the amount Y will take if X takes on 0.

2.4

a $E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = x^2$

b $Cov(X, Y) = Cov(X, X^2 + \epsilon) = E(X^3 + \epsilon X) - E(X)E(X^2 + \epsilon) = E(X^3) = 0$

$$\beta = \frac{Cov(X, Y)}{Var(X)} = 0$$

$$\alpha = E(Y) - \beta E(X) = E(X^2 + \epsilon) = E(X)^2 + Var(X) = 1$$

```
set.seed(42)
x = rnorm(1000, 0, 1)
e = rnorm(1000, 0, 1)
x2 = x^2
y = x2+e
lm(y ~ x)

c

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##       0.99821     -0.04561
```

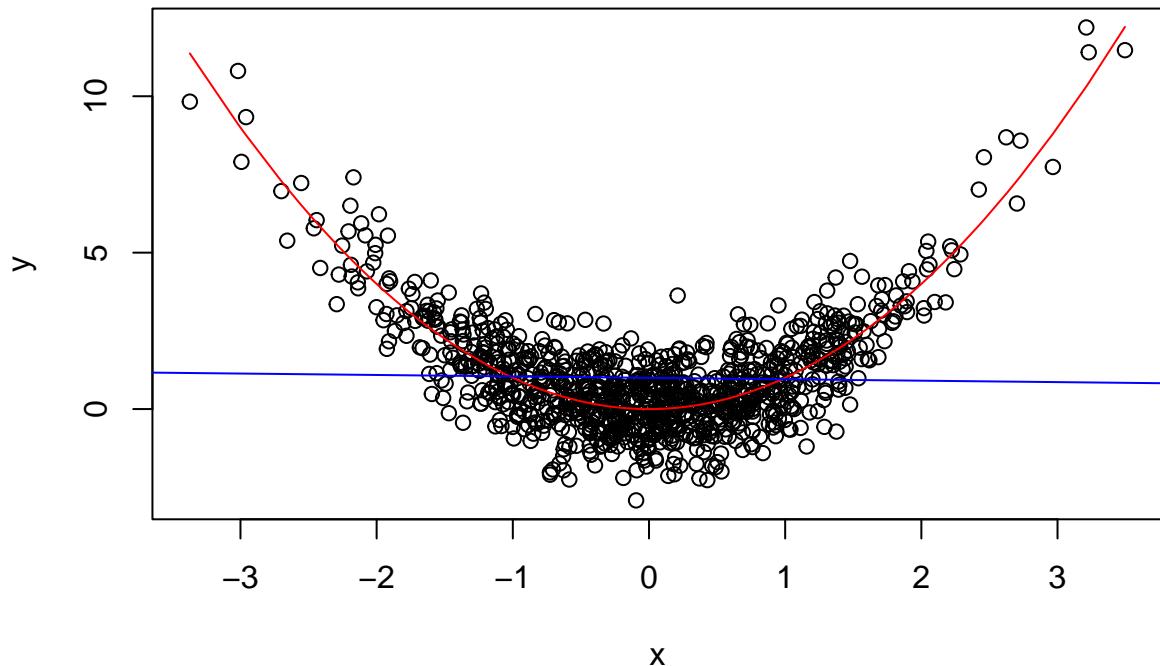
```

plot(x,y)

# CEF
lines(smooth.spline(x, x2, spar=0.35), col = 'red')

# BLP
abline(lm(y~ x), col = 'blue')

```



```
summary(lm(y~ x))
```

```

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min      1Q  Median      3Q     Max 
## -3.9220 -1.1024 -0.2886  0.7430 11.3470 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.99821   0.05631 17.726   <2e-16 ***
## x          -0.04561   0.05618 -0.812    0.417    
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.78 on 998 degrees of freedom
## Multiple R-squared:  0.0006599, Adjusted R-squared:  -0.0003415 
## F-statistic: 0.659 on 1 and 998 DF,  p-value: 0.4171

```

Good approximation happens near $x=-1$ and $x=1$.

d By CEF, $Q = a^2 - a^2 = 0$. Also, $Q' = 1 - 1 = 0$. Thus, $Q' = Q$ and thus is a good approximation of Q.

e CEF will not change because its formula is still $E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = x^2$. But for BLP, $Cov(X, Y) = Cov(X, X^2 + \epsilon) = E(X^3 + \epsilon X) - E(X)E(X^2 + \epsilon) = E(X^3)$, so as X shifts in centrality, covariance might change, leading to a change in β , the slope of the linear regression. Similarly, if changes are about the variance, β is affected as long as covariance between X,Y is not zero.

$$f \quad E(Y) = E(X^2) = 1$$

$$E(X^2) = Var(X) + E(X)^2 = 1$$

$$Var(X^2) = E(X^4) - E(X^2)^2 = 3 - 1 = 2$$

$$Cov(X^2, Y) = Cov(X^2, X^2 + \epsilon) = Var(X^2) = 2$$

$$\alpha = 1 - 1 * 1 = 0$$

$$\beta = \frac{2}{2} = 1$$

Thus, the BLP is $Y = X^2$.

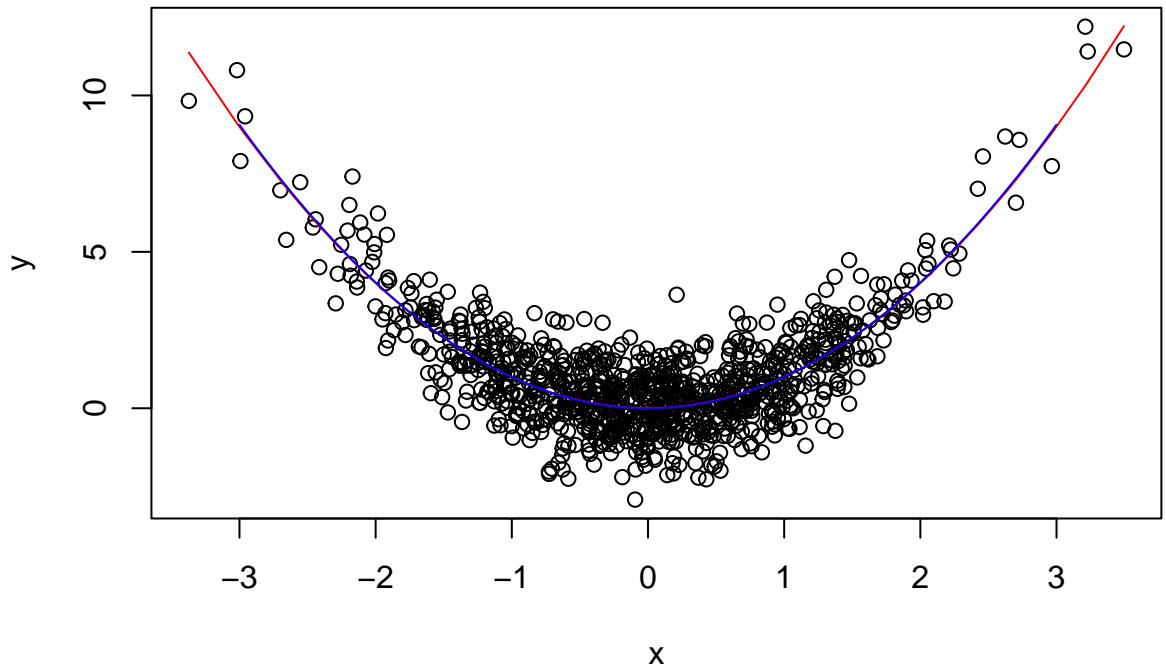
```
set.seed(42)
x = rnorm(1000, 0, 1)
e = rnorm(1000, 0, 1)
x2 = x^2
y = x2+e
lm(y~ x2)

##
## Call:
## lm(formula = y ~ x2)
##
## Coefficients:
## (Intercept)          x2
##      -0.01422       1.00886
plot(x,y)

# CEF
lines(smooth.spline(x, x2, spar=0.35), col = 'red')

# BLP
model_x = lm(y~ x2)

# can use prediction
xx = seq(-3,3,0.01)
xx2 = xx^2
y_pred = predict(model_x, list(x2 = xx2))
lines(xx, y_pred, col = 'blue')
```



```
summary(lm(y ~ x2))

##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -2.9143 -0.6545 -0.0042  0.6669  3.5985 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.01422   0.03780  -0.376   0.707    
## x2          1.00886   0.02125  47.478 <2e-16 ***
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 0.9865 on 998 degrees of freedom
## Multiple R-squared:  0.6931, Adjusted R-squared:  0.6928 
## F-statistic: 2254 on 1 and 998 DF,  p-value: < 2.2e-16
```

Q3

3.1

a $E(\epsilon|X) = E(Y - f(X)|X) = E(Y|X) - E(Y|X) = 0$

$$E(\epsilon) = E(E(\epsilon|X)) = E(0) = 0$$

b $Var(\epsilon|X) = E(\epsilon^2|X) - E(\epsilon|X)^2 = E((Y - f(X))^2|X) - 0 = E((Y - E(Y|X))^2|X) = Var(Y|X)$

$$Var(\epsilon) = Var(E(\epsilon|X)) + E(Var(\epsilon|X)) = Var(0) + E(Var(\epsilon|X)) = E(Var(\epsilon|X))$$

c $E(h(X)\epsilon) = E[E(h(X)\epsilon|X)] = E[h(X)E(0)] = 0$

The meaning of $f(X)$ is that it is the part of Y that is explained by X . The meaning of ϵ is that the part is residual and not explained and uncorrelated with X .

These are properties about the error term ϵ and not assumption.

3.2

a $E(e) = E(Y - a - bX) = E(Y) - E(a) - E(bX) = E(Y) - E(Y) + E(bX) - E(bX) = 0$

b $E(Xe) = E(E(Xe|X)) = E(X(e|X)) = E(x * 0) = 0$

c $Var(e) = Var(Y - a - bX) = Var(Y) - Var(a + bX) = Var(Y) - b^2Var(X)$

The meaning of $a_b X$ is the part of Y that is explained by a linear approximation/regression of X . e is where Y is not explained by such a linear model. These are properties of e rather than assumptions. The interpretation of b is that it is the slope of the linear regression.

3.3

No, this is not a contradiction. The ϵ and e are defined differently. ϵ is not reducible but e is the residual of information in Y where the linear model cannot capture. Therefore, it is possible we have a quadratic model, for example, to capture the information in the CEF.

3.4

a The part of Y that is explained by a linear model.

b The part of Y that is explainable by some model by not the linear model, or we can call it the non-linearity error.

c The part of Y that is the irreducible error.

3.5

The phrasing “We make two important assumptions concerning the errors” is inaccurate. We have both of those as properties instead of assumptions.

Extra credit: On page 155 of Aronow and Miller. Foundations of Agnostic Statistics, “Suppose we assume that $E[\epsilon|X[1], X[2], \dots, X[K]] = 0$ and $V[\epsilon|X[1], X[2], \dots, X[K]] = \sigma^2$ for all possible values of $X[1], X[2], \dots, X[K]$.” The authors also used the word “assume” which is not quite accurate.

Q4

4.1

$$\begin{aligned} Var(Y) &= E(Y^2) - (E(Y))^2 \\ &= E(E(Y^2|X)) - (E(Y))^2 \\ &= E(Var(Y|X) + E(Y|X)^2) - (E(Y))^2 \\ &= E(Var(Y|X)) + E(E(Y|X)^2) - E(E(Y|X))^2 \\ &= E(Var(Y|X)) + Var(E(Y|X)) \end{aligned}$$

4.2

a

$$\begin{aligned}
\eta_{Y \sim X}^2 &= 1 - \frac{E(Var(Y|X))}{Var(Y)} \\
&= \frac{Var(E(Y|X))Var(E(Y|X))}{Var(Y)Var(E(Y|X))} \\
&= \left(\frac{E(E(Y|X) - E(Y))^2}{\sigma(Y)\sigma(E(Y|X))}\right)^2 \\
&= \left(\frac{E(E(Y^2|X) + E(Y)^2 - 2E(Y|X)E(Y))}{\sigma(Y)\sigma(E(Y|X))}\right)^2 \\
&= \left(\frac{Cov(E(Y|X), Y)}{\sigma(Y)\sigma(E(Y|X))}\right)^2 \\
&= Cor^2(E(Y|X), Y)
\end{aligned}$$

b No, not necessarily. For example, the binary example as covered in class will generate different correlations, for $(Y, E(Y|X))$ and $(X, E(X|Y))$.

4.3

$$\begin{aligned}
Var(Y) &= Var(a + bX + e) \\
&= Var(a + bX) + Var(e) + 2Cov(e, a + bX) \\
&= Var(a + bX) + Var(e) + 2bCov(e, X) \\
&= Var(a + bX) + Var(e) + 2b(E(eX) - E(e)E(X)) \\
&= Var(a + bX) + Var(e) + 2b(0 - 0) \\
&= Var(a + bX) + Var(e)
\end{aligned}$$

4.4

$$\begin{aligned}
R_{Y \sim X}^2 &= \frac{Var(a + bX)}{Var(Y)} \\
&= b^2 \frac{Var(X)}{Var(Y)} \\
&= \left(\frac{Cov(X, Y)}{Var(X)}\right)^2 \frac{Var(X)}{Var(Y)} \\
&= \frac{(Cov(X, Y))^2}{Var(X)Var(Y)} \\
&= Cor^2(X, Y)
\end{aligned}$$

Also note that

$$\begin{aligned}
R_{Y \sim X}^2 &= Cor^2(X, Y) \\
&= \frac{(Cov(X, Y))^2}{Var(X)Var(Y)} \\
&= \frac{b^2(Cov(X, Y))^2}{b^2Var(X)Var(Y)} \\
&= \frac{(Cov(bX, Y))^2}{Var(bX)Var(Y)} \\
&= \frac{(Cov(a + bX, Y))^2}{Var(a + bX)Var(Y)} \\
&= Cor^2(a + bX, Y)
\end{aligned}$$

By derivation above, without loss of generality, $R_{X \sim Y}^2 = Cor^2(Y, X)$. Since by definition of correlation $Cor(X, Y) = Cor(Y, X)$. So, $R_{Y \sim X}^2 = Cor^2(X, Y) = Cor^2(Y, X) = R_{X \sim Y}^2$. Thus, linear R^2 is symmetric.

4.5

a $\eta_{Y \sim X}^2 - R_{Y \sim X}^2 = Cor^2(E(Y|X), Y) - Cor^2(\alpha + \beta X, Y) = Cor^2(\alpha + \beta X + u, Y) - Cor^2(\alpha + \beta X, Y) = (Cor(\alpha + \beta X, Y) + Cor(u, Y))^2 - Cor^2(\alpha + \beta X, Y)$. Since correlation is greater or equal to 0, $Cor(\alpha + \beta X, Y) + Cor(u, Y) \geq Cor^2(\alpha + \beta X, Y)$, it's same for square terms. Thus, $(Cor(\alpha + \beta X, Y) + Cor(u, Y))^2 \geq Cor^2(\alpha + \beta X, Y)$ such that $\eta_{Y \sim X}^2 \geq R_{Y \sim X}^2$.

b

$$\begin{aligned}
\eta_{Y \sim X}^2 - R_{Y \sim X}^2 &= \frac{Var(E(Y|X))}{Var(Y)} - \frac{(Cov(X, Y))^2}{Var(X)Var(Y)} \\
&= \frac{(Cov(X, Y))^2 + Var(Y - e)Var(X)}{Var(X)Var(Y)} \\
&= Cor^2(Y, X) + \frac{Cov(u, e)}{Var(e)Var(u)} \frac{Var(e)}{Var(Y)} \\
&= (1 - R_{Y \sim X}^2)Cor^2(u, e)
\end{aligned}$$

So, $\eta_{Y \sim X}^2 = R_{Y \sim X}^2 + 1 - R_{Y \sim X}^2)Cor^2(u, e)$

Q5

5.1

$$\beta' = \frac{Cov(Y, X')}{Var(X')} = \frac{Cov(Y, a + bX)}{Var(a + bX)} = \frac{bCov(Y, X)}{b^2Var(X)} = \frac{\beta}{b}$$

$$\alpha' = E(Y) - \beta'E(X') = E(Y) - \frac{\beta}{b}(a + bE(X)) = E(Y) - \beta E(X) - \frac{a\beta}{b} = \alpha - \frac{a\beta}{b}$$

$$\begin{aligned}
R_{Y \sim X'}^2 &= R_{Y \sim a+bX}^2 = Cor^2(Y, a + bX) = \frac{Cov^2(Y, a + bX)}{Var(Y)Var(a + bX)} = \frac{b^2Cov^2(Y, X)}{b^2Var(Y)Var(X)} = \frac{Cov^2(Y, X)}{Var(Y)Var(X)} = \\
&Cor^2(Y, X) = R_{Y \sim X}^2
\end{aligned}$$

This transformation will not change the R^2 .

5.2

$$\beta' = \frac{Cov(Y', X)}{Var(X)} = \frac{Cov(a + bY, X)}{Var(X)} = \frac{bCov(Y, X)}{Var(X)} = b\beta$$

$$\alpha' = E(Y') - \beta'E(X) = a + bE(Y) - b\beta E(X) = a + b\alpha$$

$$R_{Y' \sim X}^2 = R_{a+bY \sim X}^2 = Cor^2(a + bY, X) = \frac{Cov^2(a+bY, X)}{Var(a+bY)Var(X)} = \frac{b^2Cov^2(Y, X)}{b^2Var(Y)Var(X)} = \frac{Cov^2(Y, X)}{Var(Y)Var(X)} = Cor^2(Y, X) = R_{Y \sim X}^2$$

This transformation will not change the R^2 .

5.3

a $E(X') = E\left(\frac{X-E(X)}{SD(X)}\right) = \frac{E(X)-E(X)}{E(SD(X))} = 0$

$$E(Y') = E\left(\frac{Y-E(Y)}{SD(Y)}\right) = \frac{E(Y)-E(Y)}{E(SD(Y))} = 0$$

b $Var(X') = E[(X' - E(X'))(X' - E(X'))] = E\left(\frac{(X-E(X))(X-E(X))}{SD(X)SD(X)}\right) = \frac{Var(X)}{Var(X)} = 1$

$$Var(Y') = E[(Y' - E(Y'))(Y' - E(Y'))] = E\left(\frac{(Y-E(Y))(Y-E(Y))}{SD(Y)SD(Y)}\right) = \frac{Var(Y)}{Var(Y)} = 1$$

c $Cov(X', Y') = E[(X' - E(X'))(Y' - E(Y'))] = E\left(\frac{(X-E(X))(Y-E(Y))}{SD(X)SD(Y)}\right) = \frac{Cov(X, Y)}{SD(X)SD(Y)} = Cor(X, Y)$

$$\beta' = \frac{Cov(X', Y')}{Var(X')} = \frac{Cov(X, Y)}{SD(X)SD(Y)} = \frac{SD(X)}{SD(Y)}\beta$$

$$\alpha' = E(Y') - \beta'E(X') = 0$$

$$R_{Y' \sim X'}^2 = Cor^2(Y', X') = \frac{Cov^2(X', Y')}{Var(Y')Var(X')} = \frac{Cor^2(X, Y)}{1} = Cor^2(X, Y) = R_{Y \sim X}^2$$

This transformation will not change the R^2 .

Q6

a

$$E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = X^2$$

$$E(X^2) = Var(X) + E(X)^2 = 1.2$$

$$E(Y) = E(X^2) = 1.2$$

Let Z denote standard normal distribution. By normalization on the normal distribution, $X = 0.4Z + 1$ and therefore $E(X^3) = 3 * 0.2E(Z^2) + 1 = 1.6$.

$$\beta = \frac{Cov(X, Y)}{Var(X)} = \frac{E(X^3 + X\epsilon) - E(X)E(X^2 + \epsilon)}{0.2} = \frac{1.6 - 1.2}{0.2} = 2$$

$$\alpha = 1.2 - 2 * 1 = -0.8$$

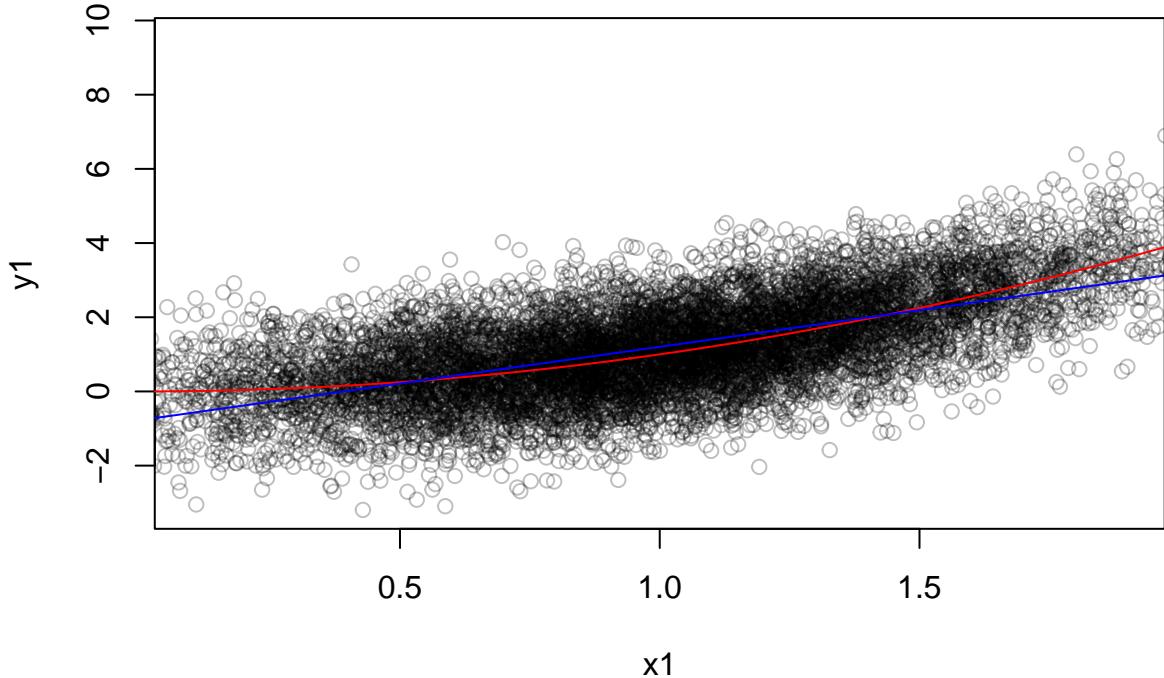
Thus, $BLP(Y|X) = -0.8 + 2X$.

```
library(scales)
set.seed(42)
x1 = rnorm(10000, 1, sqrt(0.2))
e = rnorm(10000, 0, 1)
x1_2 = x1^2
y1 = x1_2+e

model1 = lm(y1 ~ x1)
plot(x1,y1, xlim = c(.1,1.9), col = alpha('black', 0.25))

# CEF
lines(smooth.spline(x1, x1_2, spar=0.35), col = 'red')

# BLP
abline(model1, col = 'blue')
```



```
summary(model1)

##
## Call:
## lm(formula = y1 ~ x1)
##
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -3.6153 -0.7010 -0.0053  0.7010  5.4502 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.77312   0.02536 -30.49   <2e-16 ***
## x1          1.97628   0.02322  85.10   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.045 on 9998 degrees of freedom
## Multiple R-squared:  0.4201, Adjusted R-squared:  0.42 
## F-statistic: 7242 on 1 and 9998 DF,  p-value: < 2.2e-16
```

BLP is a good approximation to the CEF when X is in the range [0.5,1.5].

b

$$E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = X^2$$

$$E(X^2) = \text{Var}(X) + E(X)^2 = 1.2$$

$$E(Y) = E(X^2) = 1.2$$

Let Z denote standard normal distribution. By normalization on the normal distribution, $X = 0.4Z - 1$ and therefore $E(X^3) = -3 * 0.2E(Z^2) - 1 = -1.6$.

$$\beta = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{E(X^3+X\epsilon)-E(X)E(X^2+\epsilon)}{0.2} = \frac{-1.6+1.2}{0.2} = -2$$

$$\alpha = 1.2 - 2 * 1 = -0.8$$

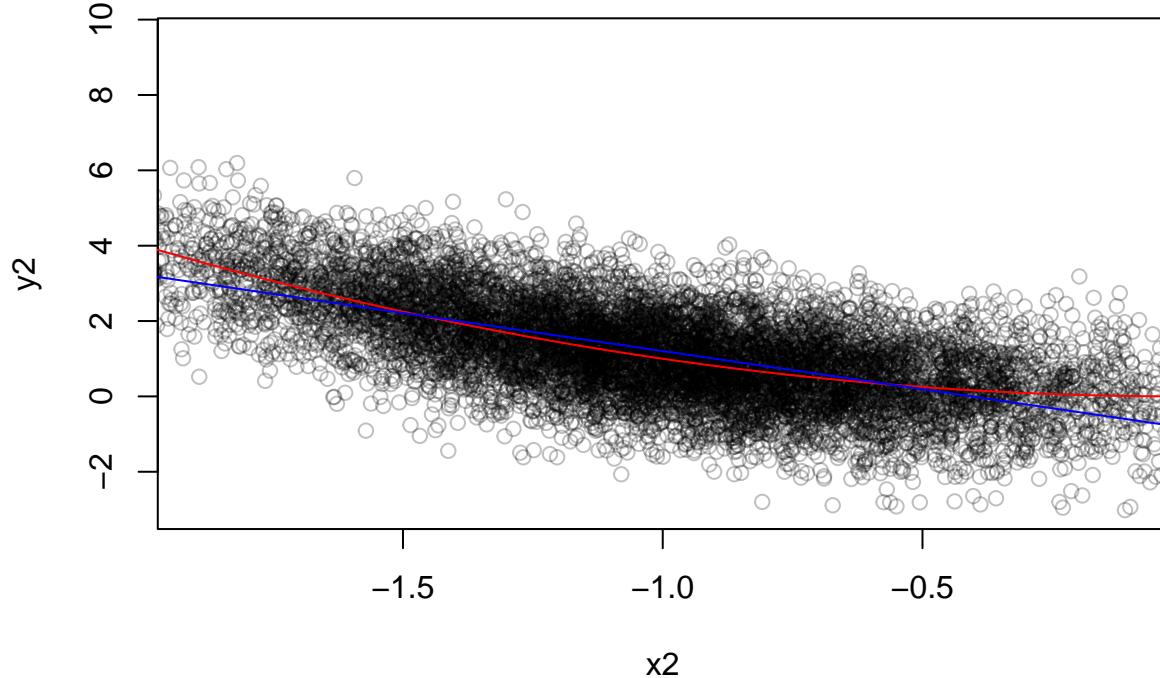
Thus, $BLP(Y|X) = -0.8 - 2X$. This is similar to (a) in the sense that only the slope has changed to a negative sign, and intercept does not change.

```
library(scales)
set.seed(42)
x2 = rnorm(10000,-1, sqrt(0.2))
e = rnorm(10000,0,1)
x2_2 = x2^2
y2 = x2_2+e

model2 = lm(y2~ x2)
plot(x2,y2, xlim = c(-1.9, -0.1), col = alpha('black', 0.25))

# CEF
lines(smooth.spline(x2, x2_2, spar=0.35), col = 'red')

# BLP
abline(model2, col = 'blue')
```



```
summary(model2)

##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6153 -0.7010 -0.0053  0.7010  5.4502
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.82055   0.02557 -32.09   <2e-16 ***
```

```

## x2           -2.02372    0.02322   -87.14    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.045 on 9998 degrees of freedom
## Multiple R-squared:  0.4317, Adjusted R-squared:  0.4316
## F-statistic:  7593 on 1 and 9998 DF,  p-value: < 2.2e-16

```

BLP is a good approximation to the CEF when X is in the range [-1.5,-0.5].

c

```

set.seed(996)
x3 = rnorm(10000,-1, sqrt(0.2))
e = rnorm(10000,0,1)
x3_2 = x3^2
y3 = x3_2+e

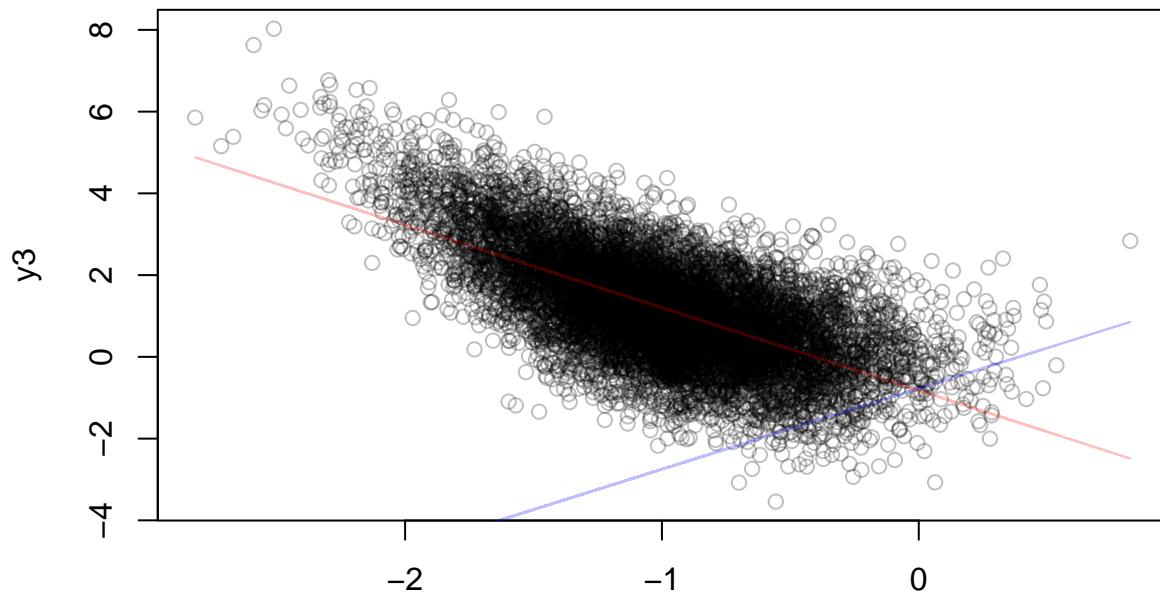
y3_pred1 <- predict(model1, newdata = data.frame(x1 = x3))
y3_pred2 <- predict(model2, newdata = data.frame(x2 = x3))

# BLP
plot(x3,y3, col = alpha('black', 0.25))

lines(x3,y3_pred1, col = alpha('blue', 0.25))

lines(x3,y3_pred2, col = alpha('red', 0.25))

```



The predictions will not be accurate. Basically, we are fitting a model based on upward trending data on data that is downward trending. The predictions given the upward trending dataset will predict y to increase as x increases, but this is not the case with our data.

d

$$\text{d.a } E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = X^2$$

$$E(X^2) = \text{Var}(X) + E(X)^2 = 1.2$$

$$E(Y) = E(X^2) = 1.2$$

$$\beta = \frac{\text{Cov}(X^2, Y)}{\text{Var}(X^2)} = \frac{E(X^4 + X^2\epsilon) - E(X^2)E(X^2 + \epsilon)}{\text{Var}(X^2)} = \frac{E(X^4) - E(X^2)E(X^2)}{\text{Var}(X^2)} = \frac{\text{Cov}(X^2, X^2)}{\text{Var}(X^2)} = 1$$

$$\alpha = 1.2 - 1 * 1.2 = 0$$

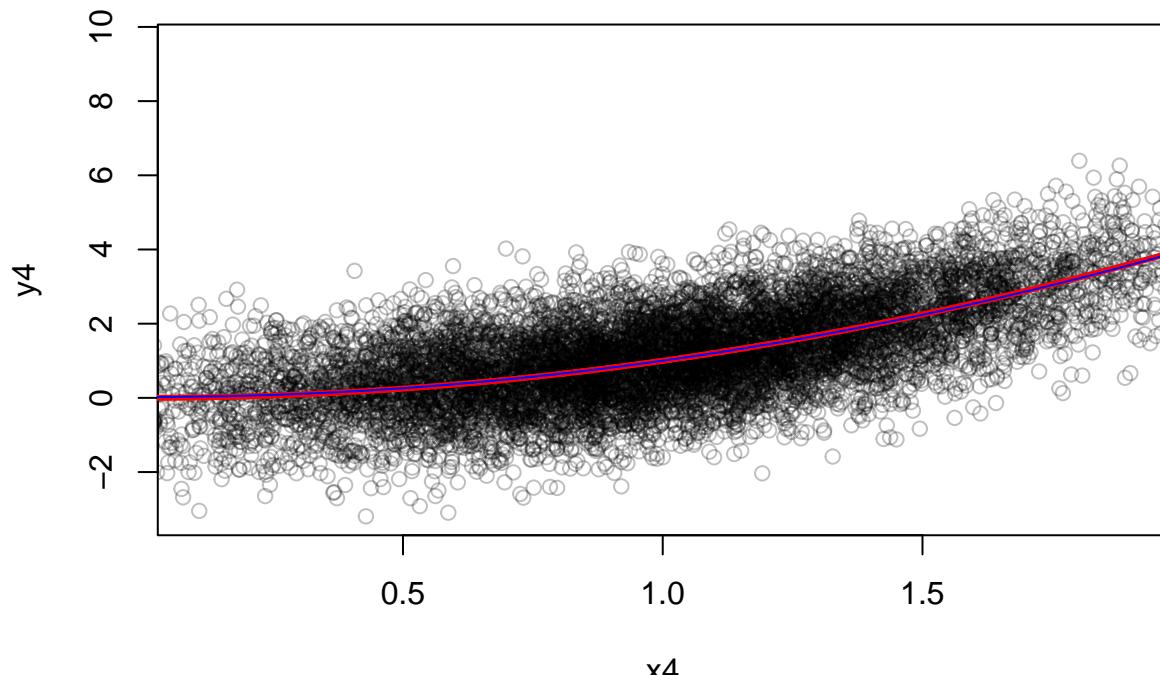
Thus, $BLP(Y|X) = X^2$.

```
library(scales)
set.seed(42)
x4 = rnorm(10000, 1, sqrt(0.2))
e = rnorm(10000, 0, 1)
x4_2 = x4^2
y4 = x4_2 + e

model4 = lm(y4 ~ x4_2)
plot(x4, y4, xlim = c(.1, 1.9), col = alpha('black', 0.25))

# CEF
lines(smooth.spline(x4, x4_2, spar=0.35), col = 'red', lwd = 3)

# BLP
y_pred4 = predict(model4, newdata = data.frame(x4_2 = x4_2))
lines(sort(x4), sort(y_pred4), col = 'blue')
```



```
summary(model4)
```

```
##
## Call:
## lm(formula = y4 ~ x4_2)
```

```

## 
## Residuals:
##   Min     1Q Median     3Q    Max
## -3.4518 -0.6755  0.0032  0.6773  3.6516
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  0.01325   0.01627   0.814   0.415    
## x4_2        0.98958   0.01071  92.397 <2e-16 ***  
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.008 on 9998 degrees of freedom
## Multiple R-squared:  0.4606, Adjusted R-squared:  0.4605 
## F-statistic:  8537 on 1 and 9998 DF,  p-value: < 2.2e-16

```

The BLP and the CEF actually are exactly the same and they overlap. BLP is a good approximation.

$$\text{d.b } E(Y|X) = E(X^2|X) + E(\epsilon|X) = E(X^2|X) = X^2$$

$$E(X^2) = \text{Var}(X) + E(X)^2 = 1.2$$

$$E(Y) = E(X^2) = 1.2$$

$$\beta = \frac{\text{Cov}(X^2, Y)}{\text{Var}(X^2)} = \frac{E(X^4 + X^2\epsilon) - E(X^2)E(X^2 + \epsilon)}{\text{Var}(X^2)} = \frac{E(X^4) - E(X^2)E(X^2)}{\text{Var}(X^2)} = \frac{\text{Cov}(X^2, X^2)}{\text{Var}(X^2)} = 1$$

$$\alpha = 1.2 - 1.2 * 1 = 0$$

Thus, $BLP(Y|X) = X^2$.

```

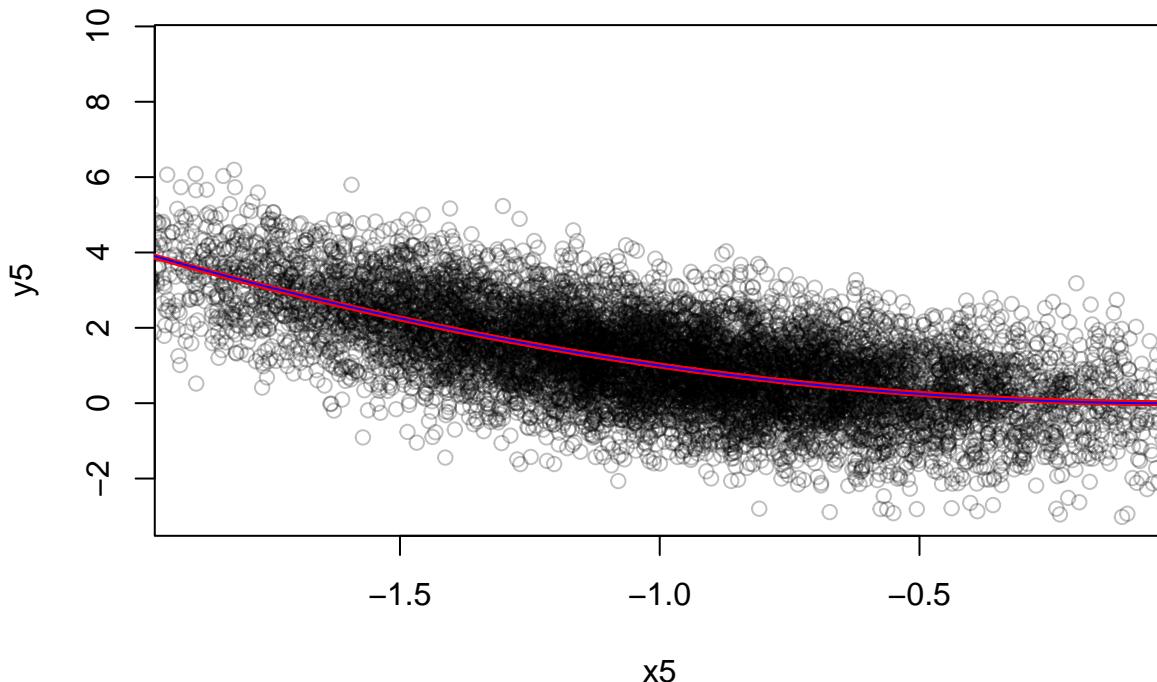
library(scales)
set.seed(42)
x5 = rnorm(10000,-1, sqrt(0.2))
e = rnorm(10000,0,1)
x5_2 = x5^2
y5 = x5_2+e

model5 = lm(y5 ~ x5_2)
plot(x5, y5, xlim = c(-1.9,-0.1), col = alpha('black', 0.25))

# CEF
lines(smooth.spline(x5, x5_2, spar=0.35), col = 'red', lwd =3)

# BLP
y_pred5 = predict(model5, newdata = data.frame(x5_2 = x5_2))
lines(smooth.spline(x5, y_pred5, spar=0.35), col = 'blue')

```



```
summary(model5)
```

```
##
## Call:
## lm(formula = y5 ~ x5_2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.4517 -0.6760  0.0027  0.6797  3.6415 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.004556   0.016370  -0.278   0.781    
## x5_2         1.004432   0.010639  94.414 <2e-16 *** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.008 on 9998 degrees of freedom
## Multiple R-squared:  0.4713, Adjusted R-squared:  0.4713 
## F-statistic: 8914 on 1 and 9998 DF,  p-value: < 2.2e-16
```

The BLP and the CEF actually are exactly the same and they overlap. BLP is a good approximation.

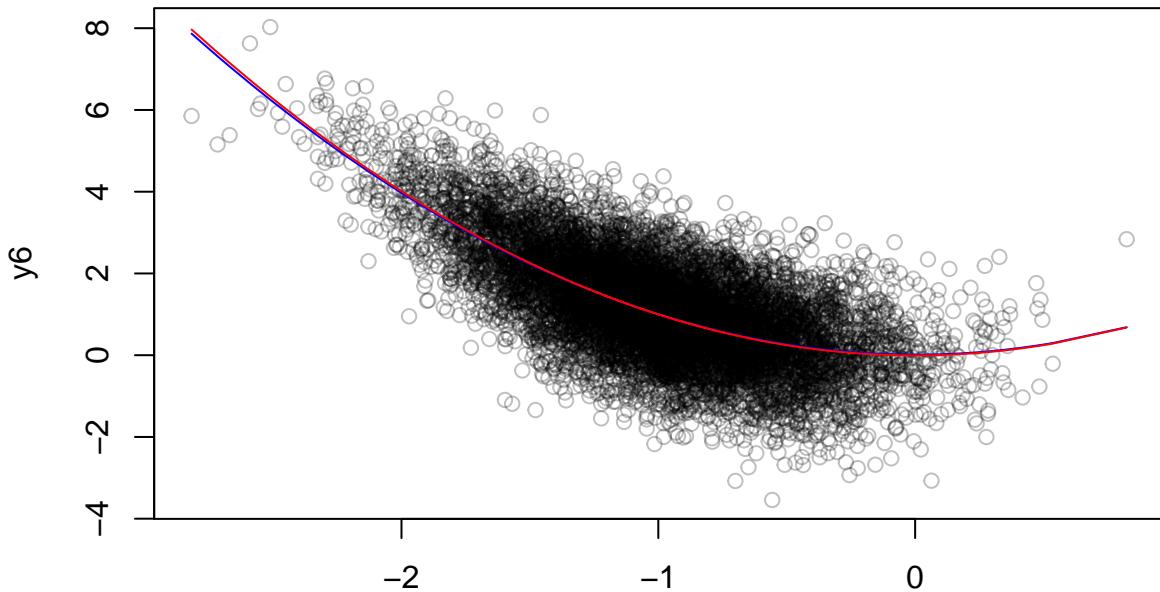
```
set.seed(996)
x6 = rnorm(10000,-1, sqrt(0.2))
e = rnorm(10000,0,1)
x6_2 = x6^2
y6 = x6_2+e

y6_pred1 <- predict(model4, newdata = data.frame(x4_2 = x6_2))
y6_pred2 <- predict(model5, newdata = data.frame(x5_2 = x6_2))
```

```
# BLP
plot(x6,y6, col = alpha('black', 0.25))

lines(smooth.spline(x6, y6_pred1, spar=0.35), col = 'blue')

lines(smooth.spline(x6, y6_pred2, spar=0.35), col = 'red')
```



d.c

x6

We can use one model to predict data from another distribution in this quadratic normal case. The predictions will be exactly the same because the two models have the same independent variables and same coefficients.

d.d There are two key changes. First, BLP now with the quadratic model is a better approximation of the CEF and does not fail with our data. This is because that they actually become the same. Second, now we can use the model in b to predict using data from a, and this is because we have had same expressions for BLP from a and b.

e

As simulated above along with the derivations, the outputs have shown that the theoretical and empirical results are quite similar, in terms of regression coefficients. The verification can further be seen at the end of each code chunk, where I put a summary function to return the coefficients, etc. and they appear close to the analytical calculations.