## Problem Set 2

1. (a)

$$
\begin{aligned}
E(home) &= 0 \cdot \left(\frac{27}{380} + \frac{21}{380} + \frac{17}{380} + \frac{1}{38} + \frac{7}{380} + \frac{1}{380} + 0\right) \\
&+ 1 \cdot \left(\frac{1}{10} + \frac{1}{10} + \frac{6}{95} + \frac{3}{95} + \frac{7}{380} + \frac{1}{380} + \frac{1}{380}\right) \\
&+ 2 \cdot \left(\frac{33}{380} + \frac{32}{380} + \frac{17}{380} + \frac{3}{380} + \frac{1}{190} + 0 + 0\right) \\
&+ 3 \cdot \left(\frac{1}{20} + \frac{13}{190} + \frac{2}{95} + \frac{1}{190} + \frac{1}{190} + 0 + 0\right) \\
&+ 4 \cdot \left(\frac{11}{380} + \frac{1}{76} + \frac{3}{190} + \frac{1}{190} + 0 + 0 + 0\right) \\
&+ 5 \cdot \left(\frac{3}{380} + \frac{1}{190} + 0 + \ldots + 0\right) \\
&+ 6 \cdot \left(0 + \frac{1}{190} + 0 + \frac{1}{380} + 0 + 0 + 0\right) \\
&= 0.318 + 2 \cdot 0.229 + 3 \cdot 0.15 + 4 \cdot 0.063 + 5 \cdot 0.013 + 6 \cdot 0.008 \\
&= 1.591
\end{aligned}
$$

$$
\begin{aligned}
E(away) &= 0 \cdot \left(\frac{27}{380} + \frac{1}{10} + \frac{33}{380} + \frac{1}{20} + \frac{11}{380} + \frac{3}{380} + 0\right) \\
&+ 1 \cdot \left(\frac{21}{380} + \frac{1}{10} + \frac{32}{380} + \frac{13}{190} + \frac{1}{76} + \frac{1}{190} + \frac{1}{190}\right) \\
&+ 2 \cdot \left(\frac{17}{380} + \frac{6}{95} + \frac{17}{380} + \frac{2}{95} + \frac{3}{190} + 0 + 0\right) \\
&+ 3 \cdot \left(\frac{1}{38} + \frac{3}{95} + \frac{3}{380} + \frac{1}{190} + \frac{1}{190} + 0 + \frac{1}{380}\right) \\
&+ 4 \cdot \left(\frac{7}{380} + \frac{7}{380} + \frac{1}{190} + \frac{1}{190} + 0 + 0 + 0\right) \\
&+ 5 \cdot \left(\frac{1}{380} + \frac{1}{380} + 0 + \ldots + 0\right) \\
&+ 6 \cdot \left(0 + \frac{1}{380} + 0 + \ldots + 0\right) \\
&= 0.332 + 2 \cdot 0.189 + 3 \cdot 0.079 + 4 \cdot 0.047 + 5 \cdot 0.005 + 6 \cdot 0.003 \\
&= 1.178
\end{aligned}
$$

(b)

$$\begin{aligned}
\text{in general: } E(y|x) &= \sum_{i=1}^{6} y_i f_{y|x}(y_i|x) = \sum_{i=1}^{6} y_i \frac{f(y_i,x)}{f(x)} = \frac{1}{f(x)} \sum_{i=1}^{6} y_i f(y_i,x) \\
E(home|away=2) &= \frac{1}{0.189} \cdot \left(0 \cdot \frac{17}{380} + 1 \cdot \frac{6}{95} + 2 \cdot \frac{17}{380} + 3 \cdot \frac{2}{95} + 4 \cdot \frac{3}{190} + 5 \cdot 0 + 6 \cdot 0\right) \\
&= 1.4759 \\
E(away|home=6) &= \frac{1}{0.008} \cdot \left(0 \cdot 0 + 1 \cdot \frac{1}{190} + 2 \cdot 0 + 3 \cdot \frac{1}{380} + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot 0\right) \\
&= 1.6447
\end{aligned}$$

2. (a) simple partial differentiation gives:

$$\frac{\partial E(y|x_1,x_2)}{\partial x_1} = \beta_1 + \beta_4 x_2$$

and

$$\frac{\partial E(y|x_1,x_2)}{\partial x_2} = \beta_2 + 2\beta_3 x_2 + \beta_4 x_1$$

(b) By definition, $E(u|x_1,x_2)$=0. Because $x_2^2$ and $x_1x_2$ are functions of ($x_1$, $x_2$), it does not matter whether or not we also condition on them:

$$E(u|x_1,x_2,x_2^2,x_1x_2) = 0$$

(c) All we can say about $Var(u|x_1,x_2)$ is that it is nonnegative for all $x_1$ and $x_2$: $E(u|x_1,x_2)$ in no way restricts $Var(u|x_1,x_2)$.

3. (a)

$$\frac{\partial E(y|x)}{\partial x} = \delta_1 + 2\delta_2(x-\mu)$$

The partial effect of x is a linear function in x.
⇒ For higher x the partial effect of x on y is higher!
⇒ Effect not constant across different x!
Example: the effect of income on the demand for the luxury goods. When people becomes richer, they are likely to have higher demand for the luxury goods. Therefore, the effect of income (x) on demand for the luxury goods (y) depends on the current income that the consumer already have.
⇒ partial effect is linear and not constant
⇒ cond. expectation is non-linear

(b)

$$E\left(\frac{\partial E(y|x)}{\partial x}\right) = E(\delta_1 + 2\delta_2(x-\mu)) = \delta_1 + 2\delta_2(\underbrace{E(x)}_{\mu} - \mu) = \delta_1$$

The average partial effect is the linear parameter $\delta_1$ !

(c)

$$\begin{aligned}
L(y|1,x) &= L(\delta_0 + \delta_1(x-\mu) + \delta_2(x-\mu)^2|1,x) \\
&= \delta_0 - \delta_1\mu + \delta_1 x + \delta_2 L((x-\mu)^2|1,x)
\end{aligned}$$

Calculate $L((x-\mu)^2|1,x)$: we have to find the linear projection $\gamma_0+\gamma_1 x$ which has minimum distance to the nonlinear function $(x-\mu)^2$:

$$\min_{\gamma_0,\gamma_1} E\left(((x-\mu)^2-\gamma_0-\gamma_1 x)^2\right)$$

FOC:

$$\begin{aligned}
\frac{\partial E(\ldots)}{\partial \gamma_0} &= E(2((x-\mu)^2-\gamma_0-\gamma_1 x)(-1)) \stackrel{!}{=} 0\\
&\Leftrightarrow E((x-\mu)^2-\gamma_0-\gamma_1 x)=0\\
&\Leftrightarrow E((x-\mu)^2)-\gamma_0-\gamma_1 \underbrace{E(x)}_{\mu}=0\\
&\Rightarrow \gamma_0 = E((x-\mu)^2)-\gamma_1\mu
\end{aligned}$$

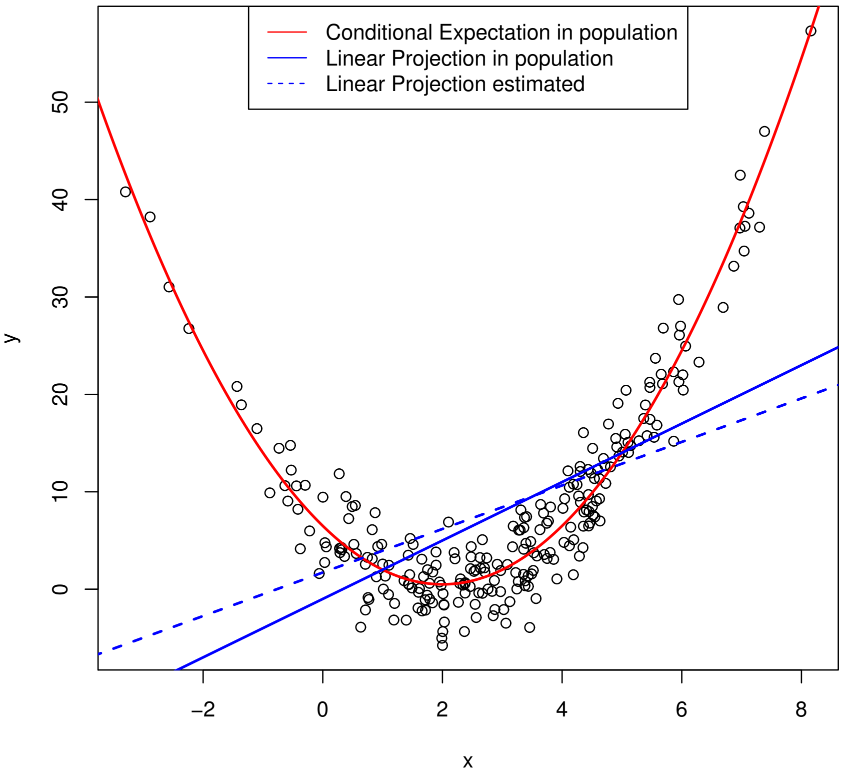
$$\begin{aligned}
\frac{\partial E(\ldots)}{\partial \gamma_1} &= E(2((x-\mu)^2-\gamma_0-\gamma_1 x)(-x)) \stackrel{!}{=} 0\\
&\Leftrightarrow E(-(x-\mu)^2 x+\gamma_0 x+\gamma_1 x^2)=0\\
&\Leftrightarrow -E((x-\mu)^2 x)+\gamma_0 \underbrace{E(x)}_{\mu} +\gamma_1 E(x^2)=0\\
&\stackrel{plugin \gamma_0}{\Leftrightarrow} -E((x-\mu)^2 x)+(E((x-\mu)^2)-\gamma_1\mu)\mu+\gamma_1 E(x^2)=0\\
&\Leftrightarrow -E(x^3-2x^2\mu+x\mu^2)+E(x^2\mu-2x\mu^2+\mu^3)-\gamma_1\mu^2+\gamma_1 E(x^2)=0\\
&\Leftrightarrow -\underbrace{E(x^3-3x^2\mu+3x\mu^2-\mu^3)}_{E((x-\mu)^3)=0} -\gamma_1(\mu^2-E(x^2))=0\\
&\Rightarrow \gamma_1=0\\
&\Rightarrow \gamma_0=E((x-\mu)^2)=Var(x)
\end{aligned}$$

So the linear projection of $(x-\mu)^2$ is $Var(x)$:

$$\begin{aligned}
L((x-\mu)^2|1,x) &= Var(x)\\
\Rightarrow L(y|1,x) &= \delta_0-\delta_1\mu+\delta_1 x+\delta_2 Var(x)\\
&= \underbrace{\delta_0-\delta_1\mu+\delta_2 Var(x)}_{c} +\delta_1 x\\
&= c+\delta_1 x\\
\Rightarrow \frac{\partial L(y|1,x)}{\partial x} &= \delta_1 = E(\partial E(y|x)/\partial x) = \text{ av. part. effect}
\end{aligned}$$

The parameter in the linear projection $\delta_1$ is the same as the parameter $\delta_1$ in the cond. expectation. A linear approximation may be suited to do inference on the average partial effect.

(d) The population model gives us the true relationship between x and y in population. The linear projection linearizes the potentially nonlinear relationship between x and y in population under minimized squared differences across the distribution of x. The estimated linear model (for example estimated by OLS) based on a sample approximates the population model if it is linear and approximates the linear projection if the population model is nonlinear(given that the estimation is not biased). Have a look at the figure below which summarizes these entities for the given exercise.




4. (a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$, where $u$ has a zero mean given $x_1$ and $x_2$: $E(u|x_1, x_2) = 0$. We can say nothing further about $u$.

(b)
$$\frac{\partial E(y|x_1, x_2)}{\partial x_1} = \beta_1 + \beta_3 x_2$$

Because $E(x_2) = 0$:
$$E\left(\frac{\partial E(y|x_1, x_2)}{\partial x_1}\right) = \beta_1 + \beta_3 E(x_2) = \beta_1$$
$\Rightarrow$ $\beta_1$ is the average partial effect of $x_1$ on $y$.

Similarly:
$$\beta_2 = E\left(\frac{\partial E(y|x_1, x_2)}{\partial x_2}\right)$$

(c) If $x_1$ and $x_2$ are independent with zero mean, then $E(x_1 x_2) = E(x_1)E(x_2) = 0$.

Further:
$$\begin{aligned} Cov(x_1 x_2, x_1) &= E(x_1 x_2 x_1) - E(x_1 x_2)E(x_1) \\ &= E(x_1^2)\underbrace{E(x_2)}_{=0} - \underbrace{E(x_1 x_2)}_{=0}\underbrace{E(x_1)}_{=0} = 0 \end{aligned}$$

Similarly, $Cov(x_1 x_2, x_2) = 0$.

$$
\begin{aligned}
L(y|1,x_1,x_2) &= \beta_0+\beta_1x_1+\beta_2x_2+\beta_3L(x_1x_2|1,x_1,x_2)\\
L(x_1x_2|1,x_1,x_2) &= \min_{\gamma} E((x_1x_2-\gamma_0-\gamma_1x_1-\gamma_2x_2)^2)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E(\ldots)}{\partial\gamma_0} &= E(-2(x_1x_2-\gamma_0-\gamma_1x_1-\gamma_2x_2)) \overset{!}{=} 0\\
&\Leftrightarrow \underbrace{E(x_1x_2)}_{=0}-\gamma_0-\gamma_1\underbrace{E(x_1)}_{=0}-\gamma_2\underbrace{E(x_2)}_{=0}=0\\
&\Rightarrow \gamma_0=0\\
\frac{\partial E(\ldots)}{\partial\gamma_1} &= E(-2x_1(x_1x_2-\gamma_0-\gamma_1x_1-\gamma_2x_2)) \overset{!}{=} 0\\
&\Leftrightarrow \underbrace{E(x_1^2x_2)}_{=0}-\gamma_0\underbrace{E(x_1)}_{=0}-\gamma_1\underbrace{E(x_1^2)}_{=Var(x_1)}-\gamma_2\underbrace{E(x_1x_2)}_{=0}=0\\
&\Rightarrow \gamma_1Var(x_1)=0\\
&\Rightarrow \gamma_1=0\\
\frac{\partial E(\ldots)}{\partial\gamma_2} & \quad \text{similar to} \quad \frac{\partial E(\ldots)}{\partial\gamma_1}\\
&\Rightarrow \gamma_2=0
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow L(x_1x_2|1,x_1,x_2))=0\\
&\Rightarrow L(y|1,x_1,x_2))=\beta_0+\beta_1x_1+\beta_2x_2
\end{aligned}
$$

$\Rightarrow$ The linear projection is able to estimate $\beta_0$, $\beta_1$ and $\beta_2$ correctly from the nonlinear conditional expectation if $x_1$ and $x_2$ are independent of one another.

Calculation rules: c is a constant

$$
\begin{aligned}
E(cx) &= cE(x)\\
E(x+c) &= E(x)+c\\
Var(x) &= E((x-\mu)^2)=E(x^2)-E(x)^2 \overset{\text{for } E(x)=0}{=} E(x^2),\ \text{where } \mu=E(x)\\
Var(x+c) &= Var(x)\\
Var(x+y) &= Var(x)+Var(y)+2Cov(x,y) \overset{iid}{=} Var(x)+Var(y)\\
Var(cx) &= c^2Var(x)\\
Cov(xy) &= E(xy)-E(x)E(y) \overset{E(x)=0 \text{ or } E(y)=0}{=} E(xy)
\end{aligned}
$$

5. $Var(u_1|x,z) \overset{E(u_1)=0}{=} E(u_1^2|x,z)=E((y-E(y|x,z))^2)=Var(y|x,z) \overset{ass.}{=} const.$
(independent of x and z)

$$\Rightarrow Var(u_1|x,z)=const.=Var(u_1)\overset{!}{=}\sigma_1^2$$

Similarly, $$Var(u_2|x)=Var(y|x)=const.$$

$$\Rightarrow Var(u_2|x) = const. = Var(u_2) \overset{!}{=} \sigma_2^2$$

Since $E(Var(y|x)) \geq E(Var(y|x,z))$ (see Wooldridge chapter appendix): $\sigma_2^2 \geq \sigma_1^2$.

This simple conclusion means that, whenever variances are constant, the error variance $\sigma^2$ falls every time as more explanatory variables are included.

6. Write the equation in error form as

$$y = \underbrace{g(x)}_{\text{nonlinear}} + \underbrace{z\beta}_{\text{linear}} + u \qquad (I)$$
$$E(u|x,z) = 0$$

Take the expected value of the first equation conditional only on x

$$E(y|x) = g(x) + (E(z|x))\beta \qquad (II)$$

and subtract this from the first equation to get:

$$(III) = (I) - (II): \underbrace{y - E(y|x)}_{\substack{\tilde{y}\text{: all variation in y} \\ \text{not explained by x}}} = \underbrace{(z - E(z|x))}_{\substack{\tilde{z}\text{: all variation in z} \\ \text{not explained by x}}} \beta + u$$
$$\tilde{y} = \tilde{z}\beta + u \quad \text{(linear model, easily estimable)}$$

Because $\tilde{z}$ is a function of $(x,z)$, $E(u|\tilde{z}) = 0$ (since $E(u|x,z) = 0$) and so $E(\tilde{y}|\tilde{z}) = \tilde{z}\beta$.

This basic result is fundamental in the literature estimating partial linear models. First, one estimates $E(y|x)$ and $E(y|z)$ using very flexible methods (typically nonparametric methods). Then, after obtaining residuals of the form $\tilde{y}_i \equiv y_i - \hat{E}(y_i|x_i)$ and $\tilde{z}_i \equiv z_i - \hat{E}(z_i|x_i)$, $\beta$ is estimated from an OLS regression $\tilde{y}_i$ on $\tilde{z}_i$, $i = 1, \dots, N$.
In the case where $E(y|x)$ and $E(z|x)$ are approximated as linear functions of a common set of functions say $\{f_1(x), \dots, f_Q(x)\}$ and $\{g_1(x), \dots, g_Q(x)\}$, the partialling out is equivalent to estimation a linear model:

$$y = a_0 + a_1 h_1(x) + \ldots + a_Q h_Q(x) + z\beta + \text{ error}$$

by OLS.

7. This is easily shown by using iterated expectations:

$$E(x'y) = E(E(x'y|x)) = E(x'E(y|x)) = E(x'\mu(x))$$

Therefore, $$\delta = (E(x'x))^{-1}E(x'y) = (E(x'x))^{-1}E(x'\mu(x))$$

and the latter equation is the vector of parameters in the linear projection of $\mu(x)$ on x.
LIE $E(u|x) \rightarrow E(u)$