PROPOSED SOLUTION TO GROUP EXAM
GRA6039 AUTUMN 2020

EMIL A. STOLTENBERG

Exercise 1

(a). In the plot in Fig. 1 we see that the data is slightly curved as x increases. Therefore, the quadratic function $g_2(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ probably gives a good model.

(b). The design matrix corresponding to this model is

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \vdots & \\ 1 & x_n & x_n^2 \end{pmatrix},$$

and the matrix $H = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}$ ensures that $\widehat{\beta} = HY$ is the least squares estimator.

(c). Clearly, $HX = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}X = I_{K+1}$, where I_{K+1} is the $(K+1)\times(K+1)$ identity matrix. Since X consists of fixed numbers, and $\mathrm{E}\,[Y] = X\beta$, we have $\mathrm{E}\,\widehat{\beta} = \mathrm{E}\,[HY] = H\mathrm{E}\,[Y] = HX\beta = \beta$.

(d). To do this, we can use Matlab code from Homework 8. Here is a table with estimates and the estimated standard errors of these estimators.

	Estimates	Standard errors
β_0	-0.478	0.031
β_1	3.400	0.143
β_2	-2.427	0.138

The Matlab code for making this table is here

```
data = readtable("ex1_data.txt")
x = data.x ; y = data.y; n = length(y);
X = [1 + zeros(n,1),x,x.^2]; % The design matrix
p = length(X(1,:)); % Get dimension
betahat = inv(transpose(X)*X)*transpose(X)*y;
sigma2hat = sum((y - X*betahat).^2)/(n - p);
sebetahat = sqrt(diag(sigma2hat*inv(transpose(X)*X)));
out= round([betahat,sebetahat],3); out = array2table(out);
out.Properties.VariableNames = {'betahat' 'se'};
```

(e). The plot asked for is given in Figure 1.

Date: December 1, 2020.

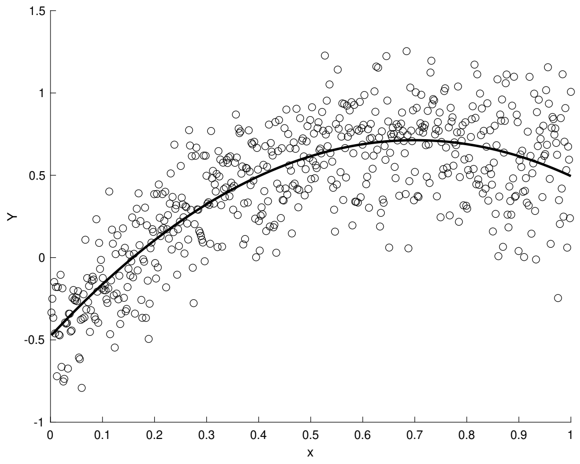


FIGURE 1. The plot from Ex. 1(a). The data points from the `ex1_data.txt` and the fitted quadratic function $\widehat{g}_2(x)$.

(f). The spread of the data points around the fitted line appears to be increasing with x. This indicates that the variance of the $\varepsilon_1, \dots, \varepsilon_n$ might not be constant. The estimated standard errors presented in the table in (d) are based on the assumption that $\mathrm{Var}(\varepsilon_i)$ are the same for all i. Since this assumption appears to be untenable, the estimated standard errors in (d) cannot be trusted.

1. Exercise 2

The pdf of X is

$$f_\mu(x) = \frac{1}{2\mu}\left(\frac{x}{2}\right)^{1/\mu - 1}, \quad \text{for } x \in [0,2],$$

with $\mu > 0$.

(a). Find $\mathrm{E}\, X^k$, for $k = 1, 2$, then use that $\mathrm{Var}(X) = \mathrm{E}\, X^2 - (\mathrm{E}\, X)^2$.

$$\begin{aligned} \mathrm{E}\, X^k &= \int_0^2 x^k f_\mu(x)\,\mathrm{d}x = \frac{1}{2\mu}\int_0^2 x^k \frac{x^{1/\mu-1}}{2^{1/\mu-1}}\,\mathrm{d}x = \frac{1}{2^{1/\mu}\mu}\int_0^2 x^{1/\mu+k-1}\,\mathrm{d}x \\ &= \frac{1}{2^{1/\mu}\mu}\frac{1}{1/\mu+k}\bigg|_0^2 x^{1/\mu+k} = \frac{1}{2^{1/\mu}}\frac{1}{1+\mu k}\bigg|_0^2 x^{1/\mu+k} = \frac{1}{2^{1/\mu}}\frac{1}{1+\mu k}2^{1/\mu+k} = \frac{2^k}{1+\mu k}, \end{aligned}$$

which gives that

$$\mathrm{E}\, X = \frac{2}{1+\mu}, \quad \text{and} \quad \mathrm{Var}(X) = \frac{4}{1+2\mu} - \frac{4}{(1+\mu)^2} = \frac{4\mu^2}{(1+2\mu)(1+\mu)^2}.$$

(b). For $x \in [0, 2)$,

$$F_\mu(x) = \frac{1}{\mu 2^{1/\mu}}\int_0^x y^{1/\mu-1}\,\mathrm{d}y = \frac{1}{2^{1/\mu}}\bigg|_0^x y^{1/\mu} = \left(\frac{x}{2}\right)^{1/\mu},$$

while $F_\mu(x) = 0$ for $x < 0$, and $F_\mu(x) = 1$ for $x \geq 2$.

(c). The natural logarithm of the pdf is

$$\log f_\mu(x) = -\log \mu - \log 2 + (1/\mu - 1)\log(x/2),$$

so the log-likelihood function is

$$\ell_n(\mu) = \sum_{i=1}^n \log f_\mu(X_i) = -n\log\mu - n\log 2 + (1/\mu - 1)\sum_{i=1}^n \log(X_i/2).$$

To find the maximum likelihood estimator we differentiate with respect to μ,

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\ell_n(\mu) = -\frac{n}{\mu} - \frac{1}{\mu^2}\sum_{i=1}^n \log(X_i/2),$$

then set $\mathrm{d}\ell_n(\mu)/\mathrm{d}\mu = 0$, and solve for μ to find

$$\widehat{\mu}_n = -\frac{1}{n}\sum_{i=1}^n \log(X_i/2).$$

(d). With $Y_1 = -\log(X_1/2)$, since $0 < X_1/2 < 1$, we see that Y_1 takes its values in $[0,\infty)$. So for $y > 0$ we have

$$\begin{aligned}\Pr(Y_1 \le y) &= \Pr(-\log(X_1/2) \le y) = \Pr(\log(X_1/2) \ge -y)\\ &= \Pr(X_1 \ge 2\exp(-y)) = 1 - \Pr(X_1 \le 2\exp(-y)) = 1 - F_\mu(2\exp(-y))\\ &= 1 - \left(\frac{2\exp(-y)}{2}\right)^{1/\mu} = 1 - \exp(-y/\mu),\end{aligned}$$

while $\Pr(Y_1 \le y) = 0$ for $y < 0$. We thus see that Y_1 has an exponential distribution, so that $\mathrm{E}\,Y_1 = \mu$ and $\mathrm{Var}(Y_1) = \mu^2$ (see Homework 2 Ex. 5, and also Homework 5 Ex. 3).

(e). Since $\widehat{\mu}_n = -(1/n)\sum_{i=1}^n \log(X_i/2) = (1/n)\sum_{i=1}^n Y_i$, and $\mathrm{E}\,[Y_i] = \mu$ for each i, we have $\mathrm{E}\,\widehat{\mu}_n = (1/n)\sum_{i=1}^n \mathrm{E}\,Y_i = \mu$, using the linearity of expectation. The $Y_1,\ldots,Y_n$ are i.i.d. random variables with mean μ and variance μ^2. Write $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i$. From the Central limit theorem (see Theorem 5.5 in the Lecture notes, or Wooldridge (2019, [C.12], p. 724)), we have that

$$\frac{\sqrt{n}(\widehat{\mu}_n - \mu)}{\mu} = \frac{\sqrt{n}(\bar{Y}_n - \mu)}{\mu} \xrightarrow{d} Z,$$

where $Z \sim \mathrm{N}(0,1)$. But by the definition of convergence in distribution (see the Lecture notes p. 22, or Wooldridge (2019, [C.11], p. 723), or handwritten notes from Lecture 5) this means that

$$\Pr\{\sqrt{n}(\widehat{\mu}_n - \mu)/\mu \le x\} \to \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}}\exp(-z^2/2)\,\mathrm{d}z,$$

for all x.

(f). Using the convergence in distribution result from (e), we have that for some significance level $\alpha \in (0,1)$,

$$\begin{aligned}\Pr\{\Phi^{-1}(\alpha/2) \leq \sqrt{n}(\widehat{\mu}_n - \mu)/\mu \leq \Phi^{-1}(1-\alpha/2)\} &\approx \Phi\{\Phi^{-1}(1-\alpha/2)\} + 1 - \Phi\{\Phi^{-1}(\alpha/2)\} \\ &= 1 - \alpha/2 + 1 - \alpha/2 = \alpha,\end{aligned}$$

when n is sufficiently large. Moving things around, we see that the event

$$\{\Phi^{-1}(\alpha/2) \leq \frac{\sqrt{n}(\widehat{\mu}_n - \mu)}{\mu} \leq \Phi^{-1}(1-\alpha/2)\},$$

is the same as the event

$$\{\frac{\sqrt{n}\widehat{\mu}_n}{\sqrt{n} + \Phi^{-1}(1-\alpha/2)} \leq \mu \leq \frac{\sqrt{n}\widehat{\mu}_n}{\sqrt{n} + \Phi^{-1}(\alpha/2)}\},$$

and we get the $(1-\alpha) \times 100$ percent confidence interval for μ.

(g). In this Matlab script we check by way of simulations that $n = 53$ is sufficiently big for the normal approximation to kick in.

```
mu = 2; alpha = 0.05; n = 53; sims = 1000;
contains = zeros(1,sims);
for jj = 1:sims
    YY = exprnd(mu,1,53);
    muhat = mean(YY);
    upper = sqrt(n)*muhat/(sqrt(n) + norminv(alpha/2));
    lower = sqrt(n)*muhat/(sqrt(n) + norminv(1 -alpha/2));
    contains(jj) = (lower <= mu)&(mu <= upper);
end
mean(contains) % should be close to (1 - alpha) = 0.95
```

2. Exercise 2

(a). Here is the Matlab script

```
n = 123;
sigma2 = 1.208;
beta0 = 0.432;
beta1 = 1.234;
beta2 = 2.467;
rho = -0.567;

sims = 10^3;
beta1hats = 0.*(1:sims);
for uu = 1:sims
    eps = normrnd(0,1,1,n);
    eta = normrnd(0,1,1,n);
    xx = sqrt(sigma2).*eta;
    zz = rho*eta + (1 - rho^2)^(1/2).*normrnd(0,1,1,n);
    y = beta0 + beta1.*xx + beta2.*zz + eps;
    beta1hats(uu) = sum((xx - mean(xx)).*y)/sum((xx - mean(xx)).^2);
```

```
end

histogram(beta1hats,"Normalization","pdf")
xlim([-1,2])
hold on
plot([mean(beta1hats),mean(beta1hats)],[0,2.4],"Linewidth",2)
plot([beta1,beta1],[0,2.4],"Linewidth",2)
```

(b). The expression for $\widehat{\beta}_1$ follows because

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n) = \sum_{i=1}^{n}(X_i - \bar{X}_n)Y_i - \sum_{i=1}^{n}(X_i - \bar{X}_n)\bar{Y}_n = \sum_{i=1}^{n}(X_i - \bar{X}_n)Y_i,$$

because $\sum_{i=1}^{n}(X_i - \bar{X}_n)\bar{Y}_n = 0$.

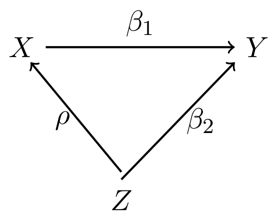
(c). We use that $\mathrm{E}\,[Y_i \mid X] = \beta_0 + \beta_1 X_i + (\rho/\sigma)\beta_2 X_i$ for each i. Then

$$\begin{aligned}\mathrm{E}\,[\widehat{\beta}_1 \mid X] &= \mathrm{E}\,[\frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)Y_i}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} \mid X] = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)\mathrm{E}\,[Y_i \mid X]}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} \\ &= \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)(\beta_0 + \beta_1 X_i + (\rho/\sigma)\beta_2 X_i)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2} = \beta_1 + \frac{\rho\beta_2}{\sigma}.\end{aligned}$$

(d). From the expression

$$\mathrm{E}\,[\widehat{\beta}_1 \mid X] = \beta_1 + \frac{\rho\beta_2}{\sigma},$$

we learn that, keeping $\sigma > 0$ constant for the moment, that $\widehat{\beta}_1$ is an unbiased estimator if $\rho = 0$ or $\beta_2 = 0$. This means that when we are interested in inference on β_1, we only need to control for Z_i (that is, include Z_i when estimating β_1) when Z_i is correlated with X_i *and* with Y_i. We can make a drawing of this.



Here Z_i is what is often called a confounder. If $\rho = 0$ or $\beta_2 = 0$ (in which case we would erase the associated arrow), then Z_i is no longer a confounder, and we do not need to worry about Z_i when estimating β_1.[1]

[1]An excellent popular science book on confounding and related matters is Pearl and Mackenzie (2018). In this book, Pearl says some things that I disagree with, so if you read it, do also read the blog post Gelman (2019) or pages Section 3 in the introduction (kappa) to my PhD-thesis, Stoltenberg (2020).

(e). The estimator $\widehat{b}_n$ is

$$\widehat{b}_n = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i^2} = \frac{(1/n)\sum_{i=1}^n W_i X_i}{(1/n)\sum_{i=1}^n W_i^2}.$$

Look at the numerator $(1/n)\sum_{i=1}^n W_i X_i$, where $W_1X_1, \ldots, W_nX_n$ are i.i.d. random variables with expectation

$$\mathrm{E}\,W_1X_1 = \mathrm{E}\,W_1(bW_1 + u_1) = b\mathrm{E}\,W_1^2 + \mathrm{E}\,u_1 = b,$$

since $\mathrm{E}\,W_1^2 = 1$ and $\mathrm{E}\,u_1 = 0$, and variance

$$\begin{aligned}\mathrm{Var}(W_1X_1) &= \mathrm{E}\,(W_1X_1)^2 - (\mathrm{E}\,W_1X_1)^2 = \mathrm{E}\,(W_1X_1)^2 - b^2\mathrm{E}\,W_1^2(bW_1+u_1)^2 - b^2 \\ &= b^2\,\mathrm{E}\,[W_1^4] + 2b\mathrm{E}\,[W_1^3]\,\mathrm{E}\,[u_1] + \mathrm{E}\,[W_1^2]\,\mathrm{E}\,[u_1^2] - b^2 = 3b^2 - b^2 = b^2,\end{aligned}$$

which is finite, so the Law of large numbers (LLN) yields

$$\frac{1}{n}\sum_{i=1}^n W_i X_i \xrightarrow{p} \mathrm{E}\,W_1X_1 = b.$$

In the denominator $(1/n)\sum_{i=1}^n W_i^2$, the $W_1^2, \ldots, W_n^2$ are i.i.d. random variables, with $\mathrm{E}\,W_1^2 = 1$, and

$$\mathrm{Var}(W_1^2) = \mathrm{E}\,W_1^4 - (\mathrm{E}\,W_1^2)^2 = 3 - 1 = 2.$$

So by the LLN, $(1/n)\sum_{i=1}^n W_i^2 \to_p \mathrm{E}\,W_1^2 = 1$. Using the PLIM.2 rules, we conclude that

$$\widehat{b}_n = \frac{(1/n)\sum_{i=1}^n W_i X_i}{(1/n)\sum_{i=1}^n W_i^2} \xrightarrow{p} \frac{b}{1} = b.$$

(f). Using the result from (b), and writing $\bar{W}_n = (1/n)\sum_{i=1}^n W_i$,

$$\begin{aligned}\widetilde{\beta}_1 &= \frac{\sum_{i=1}^n \{\widehat{X_i} - (1/n)\sum_{j=1}^n \widehat{X}_j\}Y_i}{\sum_{i=1}^n \{\widehat{X_i} - (1/n)\sum_{j=1}^n \widehat{X}_j\}^2} = \frac{1}{\widehat{b}_n}\frac{\sum_{i=1}^n (W_i - \bar{W}_n)Y_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} \\ &= \frac{1}{\widehat{b}_n}\frac{\sum_{i=1}^n (W_i - \bar{W}_n)\{\beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i\}}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} = \frac{1}{\widehat{b}_n}\Big\{\beta_1\frac{\sum_{i=1}^n (W_i - \bar{W}_n)X_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} + \beta_2 B_n + C_n\Big\} \\ &= \frac{1}{\widehat{b}_n}\Big\{\beta_1\frac{\sum_{i=1}^n (W_i - \bar{W}_n)(bW_i + u_i)}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} + \beta_2 B_n + C_n\Big\} = \frac{1}{\widehat{b}_n}(b\beta_1 + \beta_1 A_n + \beta_2 B_n + C_n),\end{aligned}$$

where

$$A_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n)u_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2}, \quad B_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n)Z_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2}, \quad C_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n)\varepsilon_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2}.$$

To get this expression for $\widetilde{\beta}_1$ we use that $\sum_{i=1}^n (W_i - \bar{W}_n) = 0$, and that $\sum_{i=1}^n (W_i - \bar{W}_n)W_i = \sum_{i=1}^n (W_i - \bar{W}_n)^2$, which is what was shown in (b). Now, write,

$$A_n = \frac{\sum_{i=1}^n (W_i - \bar{W}_n)u_i}{\sum_{i=1}^n (W_i - \bar{W}_n)^2} = \frac{(1/n)\sum_{i=1}^n (W_i - \bar{W}_n)u_i}{(1/n)\sum_{i=1}^n (W_i - \bar{W}_n)^2}.$$

It is given in the exercise that $(1/n)\sum_{i=1}^n (W_i - \bar{W}_n)^2 \to_p 1$, so we only need to prove that the numerator tends to 0 in probability. Write

$$\frac{1}{n}\sum_{i=1}^n (W_i - \bar{W}_n)u_i = \frac{1}{n}\sum_{i=1}^n W_i u_i - \bar{W}_n \frac{1}{n}\sum_{i=1}^n u_i.$$

The $W_1u_1, \dots, W_nu_n$ are i.i.d. random variables with expectation $\mathrm{E}[W_iu_i] = \mathrm{E}[W_i]\,\mathrm{E}[u_i] = 0$, using independence, and variance $\mathrm{Var}(W_iu_i) = \mathrm{E}[W_i^2u_i^2] = \mathrm{E}[W_i^2]\,\mathrm{E}[u_i^2] = 1$. Therefore,

$$\frac{1}{n}\sum_{i=1}^n W_i u_i \xrightarrow{p} 0,$$

by the LLN. Since the $W_1, \dots, W_n$ are i.i.d. $\mathrm{N}(0,1)$, and the $u_1, \dots, u_n$ are i.i.d. $\mathrm{N}(0,1)$, the LLN gives

$$\bar{W}_n = \frac{1}{n}\sum_{i=1}^n W_i \xrightarrow{p} 0, \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n u_i \xrightarrow{p} 0$$

Therefore PLIM.2 (Lemma 5.2(ii) in the Lecture notes, or Property PLIM.2(ii) in Wooldridge (2019, p. 723)), gives

$$\bar{W}_n \frac{1}{n}\sum_{i=1}^n u_i \xrightarrow{p} 0.$$

We can now use PLIM.2(i) to conclude that

$$A_n = \frac{1}{n}\sum_{i=1}^n W_i u_i - \bar{W}_n \frac{1}{n}\sum_{i=1}^n u_i \xrightarrow{p} 0.$$

(g). We have that $\widehat{b}_n \to_p b \neq 0$, and that A_n, B_n and C_n tend in probability to zero. Using the expression we found above,

$$\widetilde{\beta}_1 = \frac{1}{\widehat{b}_n}(b\beta_1 + \beta_1 A_n + \beta_2 B_n + C_n) = \frac{b}{\widehat{b}_n}\beta_1 + \beta_1\frac{A_n}{\widehat{b}_n} + \beta_2\frac{B_n}{\widehat{b}_n} + \frac{C_n}{\widehat{b}_n}.$$

Now, we use PLIM.2(iii) to conclude that

$$\frac{b}{\widehat{b}_n}\beta_1 \xrightarrow{p} \beta_1, \quad \beta_1\frac{A_n}{\widehat{b}_n} \xrightarrow{p} 0, \quad \beta_2\frac{B_n}{\widehat{b}_n} \xrightarrow{p} 0, \quad \text{and} \quad \frac{C_n}{\widehat{b}_n} \xrightarrow{p} 0,$$

and the PLIM.2(i) to conclude that

$$\widetilde{\beta}_1 = \frac{b}{\widehat{b}_n}\beta_1 + \beta_1\frac{A_n}{\widehat{b}_n} + \beta_2\frac{B_n}{\widehat{b}_n} + \frac{C_n}{\widehat{b}_n} \xrightarrow{p} \beta_1,$$

which is rather cool, and which you can learn more about in econometrics courses that cover so-called ‘instrumental variables’.

	variance	bias^2	mse
$\widehat{\beta}_1$	0.0482	0.0623	0.1105
$\widetilde{\beta}_1$	0.2999	0.0011	0.3010

TABLE 1. Results of simulations as described in Ex. 3(h). The estimates of the variance, bias^2, and the mean squared error are based on 1000 simulated datasets.

(h). The results from my simulations are summarised in Table 1. In this table we see that the estimator $\widetilde{\beta}_1$ is much less biased for β_1 than $\widehat{\beta}_1$. This is because $\rho = -0.123 \neq 0$, and $\beta_2 \neq 0$, and is what we would expect from our finding in (c). The variance of $\widetilde{\beta}_1$ is, however, much higher than the variance of $\widehat{\beta}_1$, leading to $\widehat{\beta}_1$ having a lower mean squared error than $\widetilde{\beta}_1$. So in terms of the mean squared error, $\widehat{\beta}_1$ is the better estimator.

The reason for the variance of $\widetilde{\beta}_1$ being higher than the variance of $\widehat{\beta}_1$ is twofold: First, the estimator $\widetilde{\beta}_1$ is based on the predicted values $\widehat{X}_i$ instead of X_i. The predicted values $\widehat{X}_i$ are less spread out than the X_i, and therefore contain less information about the relationshiop between X_i and Y_i. Second, in forming $\widetilde{\beta}_1$, we first estimate b. This estimating step also comes with its uncertainty (variance) which is then by $\widetilde{\beta}_1$.

The morale of all this is that if a confounder is present, but the confounding is not that strong, meaning that ρ or β_2 are close to 0, then we might want to accept some bias, because accepting some bias leads to less uncertain estimates, and perhaps a smaller mean squared error. In other words, the biased and inconsistent estimator $\widehat{\beta}_1$ might be a better estimator than the consistent estimator $\widetilde{\beta}_1$, even in the presence of a confounder.

Here is the Matlab code where I do the simulations that are asked for

```
n = 123;
beta0 = 0.432;
beta1 = 1.234;
beta2 = 2.467;
rho = -0.123;
b = 0.456;

sims = 10^3;
beta1hats = 0.*(1:sims);
beta1hatsIV = 0.*(1:sims) ;
for jj = 1:sims
    eps = normrnd(0,1,1,n) ;
    uu = normrnd(0,1,1,n);
    ww = normrnd(0,1,1,n);
    xx = b.*ww + uu;
    zz = rho.*uu + (1 - rho^2)^(1/2).*normrnd(0,1,1,n);
    y = beta0 + beta1.*xx + beta2.*zz + eps;
    beta1hats(jj) = sum((xx - mean(xx)).*y)/sum((xx - mean(xx)).^2);
    bhat = sum(xx.*ww)/sum(ww.^2);
    xhat = bhat.*ww;
```

```
    beta1hatsIV(jj) = sum((xhat - mean(xhat)).*y)/sum((xhat - mean(xhat)).^2);
end

% Make a table
vars = [var(beta1hats);var(beta1hatsIV)];
bias2 = [(mean(beta1hats) - beta1)^2;(mean(beta1hatsIV) - beta1)^2]
mse = [mean((beta1hats - beta1).^2);mean((beta1hatsIV - beta1).^2)]

out= round([vars,bias2,mse],3);
out = array2table(out);
out.Properties.VariableNames = {'variance' 'bias2' 'mse'};
out
```

References

Gelman, A. (2019). "The Book of Why" by Pearl and Mackenzie. https://statmodeling.stat.columbia.edu/2019/01/08/book-pearl-mackenzie/. Accessed: 6 January 2020.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York.

Stoltenberg, E. A. (2020). *Epidemiological, econometric, and decision theoretic applications of statistical inference*. PhD thesis, Department of Mathematics, University of Oslo. https://www.duo.uio.no/handle/10852/80949?show=full.

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach. Seventh Edition*. Cengage Learning, Boston, MA.

Department of Economics, BI Norwegian Business School
Email address: emil.a.stoltenberg@bi.no