Consider a Generalized Roy model:

$$\begin{aligned}
[A] &: Y_1 = \alpha_1' X + \beta + U_1 \\
[B] &: Y_0 = \alpha_0' X + U_0 \\
[C] &: D = 1\{Y_1 - Y_0 - C(Z) \geq 0\}
\end{aligned}$$

where $C(Z) = \gamma Z + U_Z$ and $(U_0, U_1, U_Z) \perp (X, Z)$ with distribution $N(0, \Sigma_U)$. Observed Y is $Y = DY_1 + (1-D)Y_0$.

1 Defining Things

Problem 1.1. What is the causal effect of D on Y at the individual level?

Solution. Writing out the observed outcome as a switching regression:

$$\begin{aligned}
Y &= DY_1 + (1-D)Y_0 \\
&= D(\alpha_1' X + \beta + U_1) + (1-D)(\alpha_0' X + U_0) \\
&= ((\alpha_1 - \alpha_0)' X + \beta + U_1 - U_0)D + \alpha_0' X + U_0
\end{aligned}$$

The individual causal effect of D is therefore given as:

$$(\alpha_1 - \alpha_0)' X + \beta + U_1 - U_0$$

■

Problem 1.2. What is the causal effect of D on Y at the aggregate level?

Solution. The *aggregate* effect of D on Y is ambiguous since we may be interested in ettects such as ATE or TT. In this case, we would have to specify exactly what we are interested in. For example, if we are interested in the ATE, then the causal effect at the aggregate level would be

$$\begin{aligned}
ATE(X) &\equiv \mathbb{E}[Y_1 - Y_0 | X] \\
&= (\alpha_1 - \alpha_0)' X + \beta
\end{aligned}$$

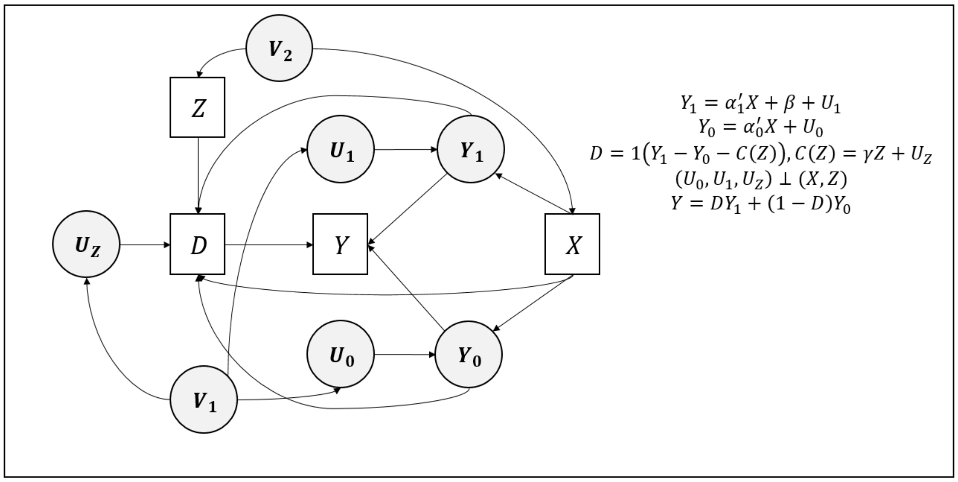
which implies

$$ATE = (\alpha_1 - \alpha_0)' \mathbb{E}[X] + \beta$$

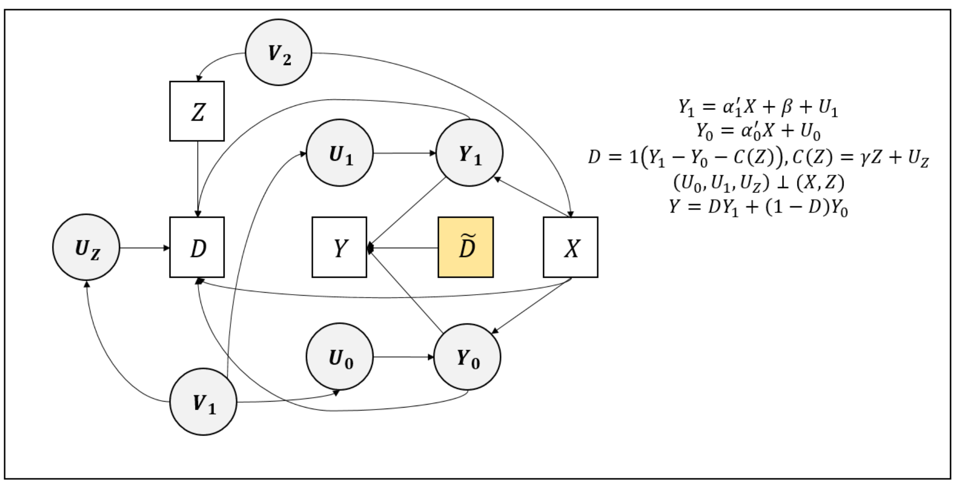
■

Problem 1.3. Using the hypothetical model framework of Heckman and Pinto, define $\tilde{D}$ and define the parents of D and of Y; of (Y_0, Y_1).

Solution. A graphical representation of the provided model would be of the following:



In the hypothetical framework, we can define $\tilde{D}$ as the following:



The parents of D, Y, Y_0, Y_1 are defined as the following:

- ▷ $Pa(D) = \{Y_1, Y_0, Z, U_Z\}$
- ▷ $Pa(Y) = \{Y_1, Y_0, \tilde{D}\}$
- ▷ $Pa(Y_1) = \{U_1, X\}$
- ▷ $Pa(Y_0) = \{U_0, X\}$

■

Problem 1.4. Write the model (A), (B), and (C) in structural equation form.

Solution. The models can be written as the following sets of structural equations for the empirical model:

$$
\begin{aligned}
Y &= f_Y(Y_1, Y_0, D) \\
D &= f_D(Y_1, Y_0, Z, U_Z) \\
Z &= f_Z(V_2, \epsilon_Z) \\
X &= f_X(V_2, \epsilon_X) \\
U_1 &= f_{U_1}(V_1, \epsilon_{U_1}) \\
U_0 &= f_{U_0}(V_1, \epsilon_{U_0}) \\
U_Z &= f_{U_z}(V_1, \epsilon_{U_z}) \\
\epsilon &= \{\epsilon_X, \epsilon_Z, \epsilon_{U_1}, \epsilon_{U_2}, \epsilon_{U_Z}\}
\end{aligned}
$$

where the exogenous variables are $V_1, V_2, \tilde{D}$. For the hypothetical model, the only equation we need to modify the structural equation for Y:

$$Y = f_Y(Y_1, Y_0, \tilde{D})$$

■

2 Question 2

Using the posted data sets I, II, and III, and assuming observations are independent, for the Generalized Roy model,

Problem 2.1. Estimate the identified parameters of $Pr\left(D=1 \mid X, Z\right)$. Be explicit about what is identified and what is not.

Solution. We can use the probit regression model to estimate the identified parameters of $Pr\left(D=1 \mid X, Z\right)$. See the results below.

```
Probit regression                               Number of obs     =      1,000
                                                Wald chi2(2)      =     170.29
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -589.34219               Pseudo R2         =     0.1485
```

d	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
z1	-.517625	.0485464	-10.66	0.000	-.6127742	-.4224758
z2	-.4268661	.0463063	-9.22	0.000	-.5176249	-.3361073
_cons	.2455417	.1302083	1.89	0.059	-.009662	.5007454

Figure 1: Probit Estimation for Sample 1

```
Probit regression                               Number of obs     =      1,000
                                                Wald chi2(2)      =     124.08
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -160.31334               Pseudo R2         =     0.7676
```

d	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
z1	-3.173196	.2957287	-10.73	0.000	-3.752813	-2.593578
z2	-3.207634	.289349	-11.09	0.000	-3.774747	-2.64052
_cons	.5474425	.2186192	2.50	0.012	.1189568	.9759282

Note: 73 failures and 113 successes completely determined.

Figure 2: Probit Estimation for Sample 2

```
Probit regression                               Number of obs     =      1,000
                                                Wald chi2(2)      =     138.56
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -608.37793               Pseudo R2         =     0.1114
```

d	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
x1	.395129	.0423183	9.34	0.000	.3121866	.4780714
x2	-.3695718	.0449015	-8.23	0.000	-.457577	-.2815665
_cons	-.2693645	.20649	-1.30	0.192	-.6740774	.1353484

Figure 3: Probit Estimation for Sample 3

Probit model is based on the latent variable

$$\begin{aligned} Pr\left(D=1 \mid X, Z\right) &= Pr\left(D^{\star}>0 \mid X, Z\right) \\ &= Pr\left[Y_1-Y_0-C\left(Z\right)>0 \mid X, Z\right] \\ &= Pr\left[\left(\alpha_1-\alpha_0\right)' X+\beta-\gamma Z+U_1-U_0-U_Z>0 \mid X, Z\right] \\ &= Pr\left[U_0-U_1+U_Z<\left(\alpha_1-\alpha_0\right)' X-\gamma Z+\beta \mid X, Z\right] \\ &= F\left(\left(\alpha_1-\alpha_0\right)' X-\gamma Z+\beta\right) \end{aligned}$$

Since we assume that (U_0, U_1, U_Z) are normally distributed, we have that

$$Pr\left(D=1 \mid X=x, Z=z\right)=\Phi\left(\frac{\left(\alpha_1-\alpha_0\right)' x-\gamma z+\beta}{\sigma}\right)$$

where Φ is the standard normal c.d.f. and

$$\sigma^2=\left(\begin{array}{ccc} -1 & 1 & 1 \end{array}\right) \Sigma_U\left(\begin{array}{c} -1 \\ 1 \\ 1 \end{array}\right)$$

Thus, in samples 1 and 2 ($X=1$), $\frac{\alpha_1-\alpha_0+\beta}{\sigma}$ and $\frac{\gamma}{\sigma}$ are identified. In sample 3 ($Z=1$), $\frac{\alpha_1-\alpha_0}{\sigma}$ and $\frac{\beta-\gamma}{\sigma}$ are identified.

$\blacksquare$

Problem 2.2. What is the sample support of $Pr\left(D=1 \mid X, Z\right)$? Plot the density of $Pr\left(D=1 \mid X, Z\right)$.

Solution. The support of $Pr\left(D=1 \mid X, Z\right)$ and the density are summarized below.

▷ For sample 1:

Variable	Obs	Mean	Std. Dev.	Min	Max
v1	523	486.9159	293.5966	1	1000
y	523	1.047874	.9470031	-1.529474	3.949654
z1	523	1.702987	.8675155	-.4684068	4.109538
z2	523	-2.255818	.948127	-5.458233	.6565941
d	523	1	0	1	1

Figure 4: Sample Support (Sample 1)

* The range of Z_1 is $[-0.46, 4.11]$
* The range of Z_2 is $[-5.46, 0.66]$
* The density is below:

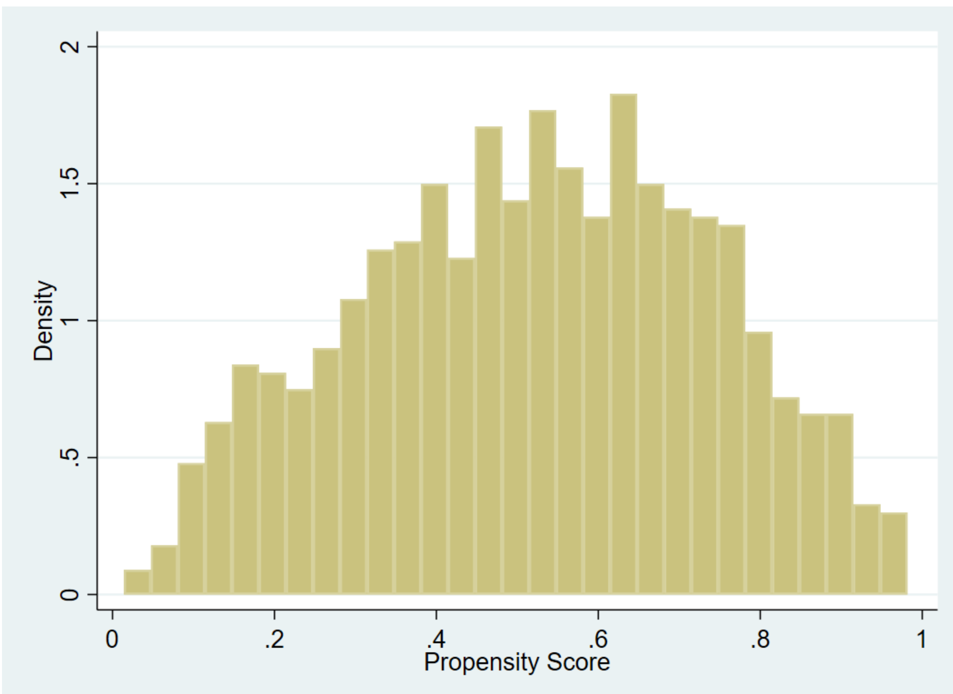


Figure 5: Density of the Propensity Score (Sample 1)

* The range of the propensity score is $[0.014, 0.981]$

▷ For sample 2:

```
su if d==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v1	541	504.1275	289.2416	1	998
y	541	.7503515	.9374318	-2.162541	3.6126
z1	541	1.501185	.7901364	-1.077855	3.998987
z2	541	-2.465618	.8825081	-5.514617	.3301109
d	541	1	0	1	1

Figure 6: Sample Support (Sample 2)

* The range of Z_1 is $[-1.07, 4.00]$
* The range of Z_2 is $[-5.51, 0.33]$
* The density is below:

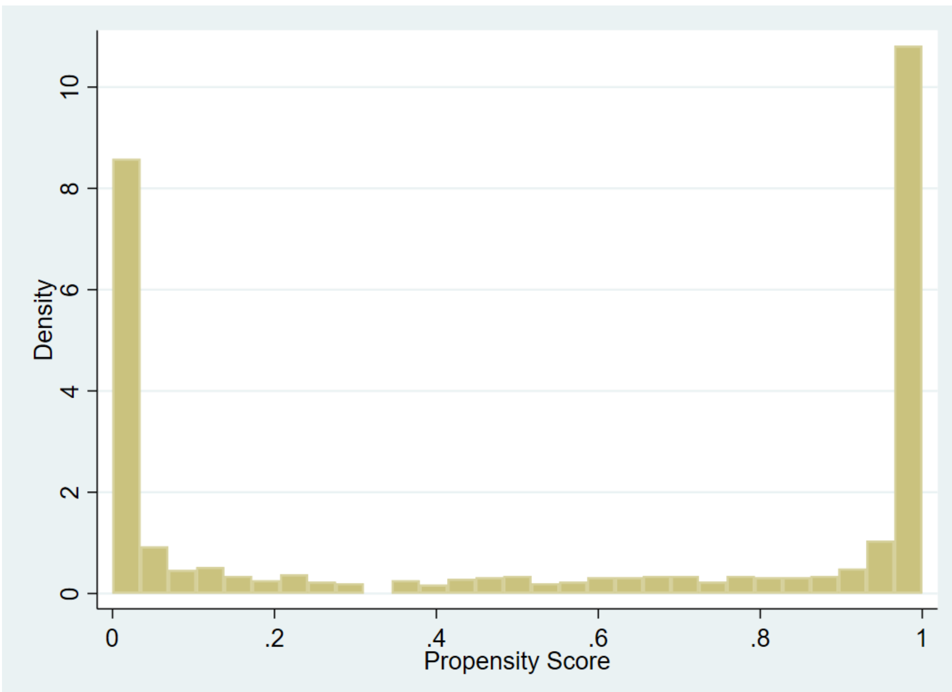


Figure 7: Density of the Propensity Score (Sample 2)

* The range of the propensity score is $[0, 1]$

▷ For sample 3:

Variable	Obs	Mean	Std. Dev.	Min	Max
v1	565	516.2903	286.6948	2	999
y	565	2.531423	1.22057	-1.056918	6.416952
x1	565	4.214199	1.006129	1.498377	6.989486
x2	565	2.792502	.9566659	.1941044	5.955626
d	565	1	0	1	1

Figure 8: Sample Support (Sample 3)

* The range of X_1 is $[1.50, 6.99]$
* The range of X_2 is $[0.19, 5.96]$
* The density is below:

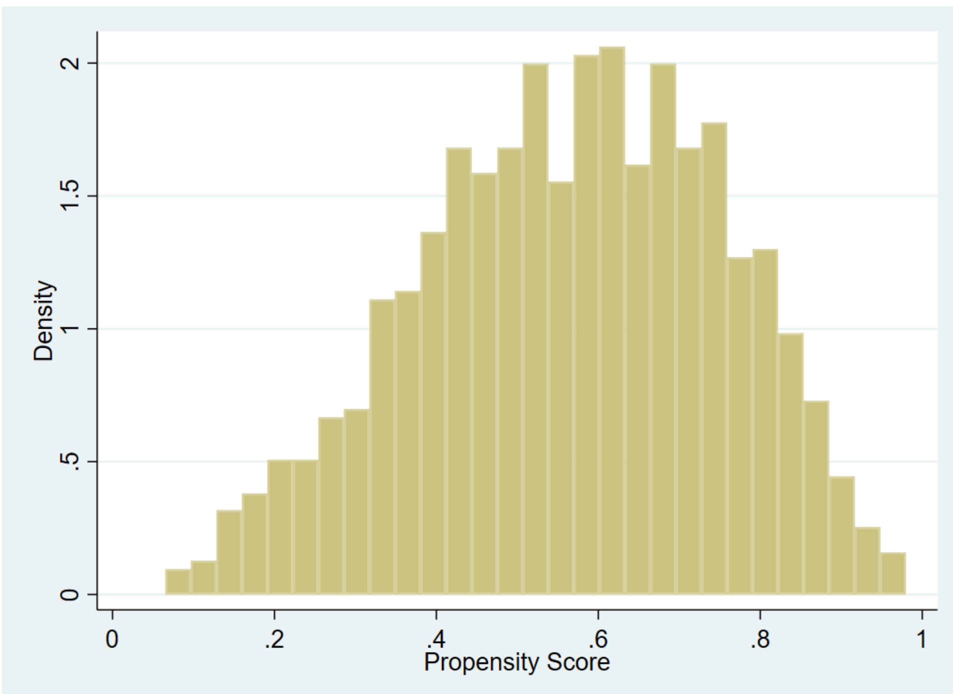


Figure 9: Density of the Propensity Score (Sample 3)

* The range of the propensity score is $[0.065, 0.980]$

Problem 2.3. Estimate the parameters of (A), (B) and (C) by maximum likelihood. Write out the likelihood and discuss what parameters are identified and what parameters are not.

Solution. Let $\theta = (\alpha_0, \alpha_1, \beta, \gamma, \Sigma_U)$. The likelihood of observing a specific data point $(Y, D = i, Z, X)$ is

$$\begin{aligned}
f(Y, X, D, Z; \theta) &= f(Y, D|X, Z; \theta) \cdot f(X, Z; \theta) \\
&= f(D = i|Y_i, X, Z; \theta) \cdot f(Y_i|X, Z; \theta) \cdot f(X, Z; \theta) \\
&\propto f(D = i|Y_i, X, Z; \theta) \cdot f(Y_i|X, Z; \theta) \\
&= D(f(D = 1|Y_1, X, Z; \theta) \cdot f(Y_1|X, Z; \theta)) + (1 - D)(f(D = 0|Y_0, X, Z; \theta) \cdot f(Y_0|X, Z; \theta)),
\end{aligned}$$

where the second line follows from Bayes rule and the third line invokes the fact that $f(X, Z; \theta)$ does not depend on θ and so can be left out of the MLE problem. Since $(U_0, U_1, U_Z) \perp (X, Z)$, we have

$$\begin{aligned}
Y_1|X, Z &\sim N(\alpha_1' X + \beta, \sigma_1^2) \\
Y_0|X, Z &\sim N(\alpha_0' X, \sigma_0^2).
\end{aligned}$$

Let $f_1(x)$ and $f_0(x)$ be the PDFs of these two distributions, respectively. We also have

$$\begin{aligned}
f(D = 1|Y_1, X, Z; \theta) &= Pr((\alpha_1 - \alpha_0)' X + \beta - \gamma Z + U_1 - U_0 - U_Z > 0|Y_1, X, Z) \\
&= Pr((\alpha_1 - \alpha_0)' X + \beta - \gamma Z > -U_1 + U_0 + U_Z|U_1, X, Z).
\end{aligned}$$

Note that

$$V = U_0 - U_1 + U_Z \sim N(0, \sigma_V^2)$$

where

$$\sigma_V^2 = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \Sigma_U \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Then,

$$\begin{pmatrix} U_1 \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{1V} \\ \sigma_{1V} & \sigma_V^2 \end{pmatrix} \right),$$

where

$$\sigma_{1V} = Cov(U_0 - U_1 + U_Z, U_1) = \sigma_{01} - \sigma_1^2 + \sigma_{21}.$$

By our standard formula for the conditional normal density:

$$V|U_1 \sim N\left(\frac{\sigma_{1V} U_1}{\sigma_1^2}, \sigma_v^2 - \frac{\sigma_{1V}}{\sigma_1^2} \right).$$

Thus,

$$f(D = 1|Y_1, X, Z; \theta) = \Phi\left(\frac{(\alpha_1 - \alpha_0)' X + \beta - \gamma Z - \frac{\sigma_{1V} U_1}{\sigma_1^2}}{\sqrt{\sigma_V^2 - \frac{\sigma_{1V}}{\sigma_1^2}}} \right),$$

where $U_1 = Y_1 - (\alpha_1' X + \beta)$. By the same argument, we find

$$V|U_0 \sim N\left(\frac{\sigma_{0V} U_0}{\sigma_0^2}, \sigma_v^2 - \frac{\sigma_{0V}}{\sigma_0^2}\right)$$
$$\sigma_{0V} = \sigma_0^2 - \sigma_{01} + \sigma_{0Z},$$

where $U_0 = Y_1 - \alpha_0' X$. Hence,

$$\begin{aligned} f(D=0|Y_1, X, Z; \theta) &= Pr((\alpha_1 - \alpha_0)'X + \beta - \gamma Z + U_1 - U_0 - U_Z < 0 | Y_1, X, Z) \\ &= Pr((\alpha_1 - \alpha_0)'X + \beta - \gamma Z < -U_1 + U_0 + U_Z | U_1, X, Z), \end{aligned}$$

and so

$$f(D=0|Y_1, X, Z; \theta) = 1 - \Phi\left(\frac{(\alpha_1 - \alpha_0)'X + \beta - \gamma Z - \frac{\sigma_{0V} U_0}{\sigma_0^2}}{\sqrt{\sigma_V^2 - \frac{\sigma_{0V}}{\sigma_0^2}}}\right).$$

Therefore, the likelihood for a particular observation is

$$\begin{aligned} f(Y, X, D, Z; \theta) = D\Bigg(& f_1(Y)\Phi\left(\frac{(\alpha_1 - \alpha_0)'X + \beta - \gamma Z - \frac{\sigma_{1V} U_1}{\sigma_1^2}}{\sqrt{\sigma_V^2 - \frac{\sigma_{1V}}{\sigma_1^2}}}\right).\Bigg) \\ & + (1-D)\Bigg(f_2(Y)\Bigg(1 - \Phi\left(\frac{(\alpha_1 - \alpha_0)'X + \beta - \gamma Z - \frac{\sigma_{0V} U_0}{\sigma_0^2}}{\sqrt{\sigma_V^2 - \frac{\sigma_{0V}}{\sigma_0^2}}}\right)\Bigg)\Bigg) \end{aligned}$$

So the log-likelihood of the entire data set is

$$\mathcal{L}(\{Y_i, X_i, D_i, Z_i\}_{i=1}^N; \theta) = \sum_{i=1}^N \ln f(Y, X, D, Z; \theta).$$

Computationally, we minimize the negative log-likelihood. Even though the MLE will find values for all parameters $(\alpha_0, \alpha_1, \beta, \gamma, \Sigma_U)$, note that not all parameters are independently identified. In samples 1 and 2, $\alpha_1 - \alpha_0 + \beta, \gamma$, and Σ_U are identified. In sample 3, $\alpha_1 - \alpha_0, \beta - \gamma$, and Σ_U are identified. Tables 1, 2, 3 display the MLE results.

	MLE Estimate
α_0	-0.122423
α_1	0.195291
β	0.954786
σ_0	1.50271
σ_1	0.933935
σ_z	1.75111
σ_{10}	-3.82533
σ_{1z}	4.99822
σ_{0z}	-0.596146
γ	[0.9123627467940012, 0.8541583433255893]

Table 1: Sample 1 MLE estimates

	MLE Estimate
α_0	-0.0289302
α_1	-0.0770294
β	0.433607
σ_0	1.18315
σ_1	0.95799
σ_z	2.31704
σ_{10}	-1.73764
σ_{1z}	2.4339
σ_{0z}	-3.11575
γ	[1.0963450936869301, 0.9388084651044197]

Table 2: Sample 2 MLE estimates

	MLE Estimate
α_0	[-0.03980103411375041, 0.44283801162204917]
α_1	[0.5509370161585607, -0.28665278554538587]
β	1.03928
σ_0	1.32901
σ_1	1.04234
σ_z	0.428502
σ_{10}	0.91371
σ_{1z}	0.332449
σ_{0z}	2.97832
γ	-0.220292

Table 3: Sample 3 MLE estimates

Problem 2.4. In terms of the model, write out the expressions for

$$E\left(Y \mid D=1, X, Z\right)$$

and

$$E\left(Y \mid D=0, X, Z\right)$$

Estimate the parameters of the model using regression analysis (use your estimates of $Pr\left(D=1 \mid X, Z\right)$ from (a)). What parameters of the full model are identified from these regressions?

Solution. The expressions are

$$\begin{aligned}
E\left(Y \mid D=1, X, Z\right) &= E\left(Y_1 \mid D=1, X, Z\right) \\
&= E\left(\alpha_1' X + \beta + U_1 \mid D=1, X, Z\right) \\
&= \alpha_1' X + \beta + E\left(U_1 \mid Y_1 - Y_0 - \gamma Z - U_Z > 0, X, Z\right) \\
&= \alpha_1' X + \beta + E\left(U_1 \mid \left(\alpha_1 - \alpha_0\right)' X - \gamma Z + \beta + U_1 - U_0 - U_Z > 0, X, Z\right) \\
&= \alpha_1' X + \beta + E\left(U_1 \mid U_0 - U_1 + U_Z < \left(\alpha_1 - \alpha_0\right)' X - \gamma Z + \beta, X, Z\right)
\end{aligned}$$

As in part c, let $V = U_0 - U_1 + U_Z$. Then

$$V \sim N(0, \sigma_V^2),$$

where σ_V^2 is defined as in part c. Let

$$\rho_1 = \frac{\sigma_{1V}}{\sigma_1 \sigma_V}.$$

Then,

$$\begin{pmatrix} U_1/\sigma_1 \\ V/\sigma_v \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}\right),$$

Thus,

$$\frac{U_1}{\sigma_1} = \rho_1 \frac{V}{\sigma_V} + \epsilon, \epsilon \sim N(0, 1-\rho_1^2), \epsilon \perp V.$$

Then by exogeneity $U_1 \perp (X, Z)$:

$$\begin{aligned}
\mathbb{E}[U_1 | V < \beta + (\alpha_1 - \alpha_0)'X - \gamma Z] &= \mathbb{E}[\rho_1 \sigma_1 \frac{V}{\sigma_V} + \epsilon | V < \beta + (\alpha_1 - \alpha_0)'X - \gamma Z] \\
&= \frac{\rho_1 \sigma_1}{\sigma_V} \mathbb{E}[V | V < \beta + (\alpha_1 - \alpha_0)'X - \gamma Z], \text{ by exogeneity} \\
&= 0 - \frac{\rho_1 \sigma_1}{\sigma_V} \sigma_V \frac{\phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)}{\Phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)}
\end{aligned}$$

Since

$$\rho_1 \sigma_1 = \frac{\sigma_{1V}}{\sigma_V},$$

we have

$$\mathbb{E}[Y | D = 1, X, Z] = \alpha_1' X + \beta - \frac{\sigma_{1V}}{\sigma_V} \frac{\phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)}{\Phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)},$$

where the inverse Mill's ratio

$$\frac{\sigma_{1V}}{\sigma_V} \frac{\phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)}{\Phi\left(\frac{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z}{\sigma_V}\right)}$$

can be computed using the parameters learned from the probit regression.

Similarly, we have

$$\begin{aligned}
E(Y \mid D = 0, X, Z) &= E(Y_0 \mid D = 0, X, Z) \\
&= E(\alpha_0' X + U_0 \mid D = 0, X, Z) \\
&= \alpha_0' X + E(U_0 \mid Y_1 - Y_0 - \gamma Z - U_Z \leq 0, X, Z) \\
&= \alpha_0' X + E\left(U_0 \mid (\alpha_1 - \alpha_0)' X - \gamma Z + \beta + U_1 - U_0 - U_Z \leq 0, X, Z\right) \\
&= \alpha_0' X + E\left(U_0 \mid V \geq (\alpha_1 - \alpha_0)' X - \gamma Z + \beta, X, Z\right)
\end{aligned}$$

	Sample$_1$	Sample$_2$	Sample$_3$
$\alpha_1' X + \beta$	0.942247	0.721983	NaN
α_1	NaN	NaN	[0.7229993849920087, -0.18194897368158716]
β	NaN	NaN	-0.237486
α_0	0.602631	0.461175	[-0.12608107215312786, 0.6288068692441382]
σ_{1v}/σ_v	-0.164825	-0.173041	-0.374994
σ_{0v}/σ_v	0.818849	-0.00548664	0.907163

Table 4: Regression parameter estimates.

Let

$$\rho_0 = \frac{\sigma_{0V}}{\sigma_0 \sigma_V}.$$

Then,

$$\frac{U_0}{\sigma_0} = \rho \frac{V}{\sigma_V} + \eta, \eta \sim N(0, 1-\rho_0^2), \eta \perp V.$$

Hence,

$$\begin{aligned}
\mathbb{E}[U_0 | V > \beta + (\alpha_1 - \alpha_0)' X - \gamma Z] &= \mathbb{E}[\rho_0 \sigma_0 \frac{V}{\sigma_V} + \eta | V > \beta + (\alpha_1 - \alpha_0)' X - \gamma Z] \\
&= \frac{\rho_0 \sigma_0}{\sigma_V} \mathbb{E}[V | V > \beta + (\alpha_1 - \alpha_0)' X - \gamma Z], \text{ by exogeneity} \\
&= \frac{\rho_0 \sigma_0}{\sigma_V} \sigma_V \frac{\phi\left(\frac{\beta + (\alpha_1 - \alpha_0)' X - \gamma Z}{\sigma_V}\right)}{1 - \Phi\left(\frac{\beta + (\alpha_1 - \alpha_0)' X - \gamma Z}{\sigma_V}\right)}
\end{aligned}$$

Since

$$\rho_0 \sigma_0 = \frac{\sigma_{0V}}{\sigma_V},$$

we have

$$\mathbb{E}[Y | D = 0, X, Z] = \alpha_0' X + \frac{\sigma_{0V}}{\sigma_V} \frac{\phi\left(\frac{\beta + (\alpha_1 - \alpha_0)' X - \gamma Z}{\sigma_V}\right)}{1 - \Phi\left(\frac{\beta + (\alpha_1 - \alpha_0)' X - \gamma Z}{\sigma_V}\right)}.$$

Thus, in samples 1 and 2 ($X = 1$): $\alpha_1 + \beta, \alpha_0, \sigma_{1V}/\sigma_V, \sigma_{0V}/\sigma_V$ are identified. In sample 3 ($Z = 1$): $\alpha_1, \beta, \alpha_0, \sigma_{1V}/\sigma_V, \sigma_{0V}/\sigma_V$ are identified.

Problem 2.5. Express $E\left(U_1 \mid D=1, X, Z\right)$ as a function of $Pr\left(D=1 \mid X, Z\right)$. Express $E\left(U_0 \mid D=0, X, Z\right)$ as a function of $Pr\left(D=1 \mid X, Z\right)$

- ▷ What are the estimates of these parameters? How do they compare with the estimates you obtain from Part (c).
- ▷ Plot these expressions as functions of $Pr\left(D=1 \mid X, Z\right)$, for X set at the sample mean.

Solution. As proved in part d, we have

$$
\begin{aligned}
\mathbb{E}[U_1|D=1,X,Z] &= \mathbb{E}[U_1|V<\beta+(\alpha_1-\alpha_0)'X-\gamma Z] \\
&= -\frac{\sigma_{1V}}{\sigma_V}\frac{\phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)}{\Phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)} \\
&= -\frac{\sigma_{1V}}{\sigma_V}\frac{\phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)}{Pr(D=1|X,Z)} \\
\mathbb{E}[U_0|D=0,X,Z] &= \mathbb{E}[U_0|V>\beta+(\alpha_1-\alpha_0)'X-\gamma Z] \\
&= \frac{\sigma_{0V}}{\sigma_V}\frac{\phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)}{1-\Phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)} \\
&= \frac{\sigma_{0V}}{\sigma_V}\frac{\phi\left(\frac{\beta+(\alpha_1-\alpha_0)'X-\gamma Z}{\sigma_V}\right)}{1-Pr(D=1|X,Z)}
\end{aligned}
$$

Note that all parameters are identified by MLE. The probit regression identifies the argument to ϕ and Φ, and the regression from part d identifies the variance ratios multiplying the Mill's ratio terms. Figures 10, 11, and 12 display histograms of $\mathbb{E}[U_1|D=1,X,Z]$ and $\mathbb{E}[U_0|D=0,X,Z]$ calculated at all (X,Z) in the data. Note that many of these values are, even in expectation, non-zero.

Figures 13, 14, and 14 plot $\mathbb{E}[U_1|D=1,X,Z]$ and $\mathbb{E}[U_0|D=0,X,Z]$ as functions of $P(D=1|X,Z)$ (i.e. set X to the sample mean and vary Z to generate a range or propensity scores).

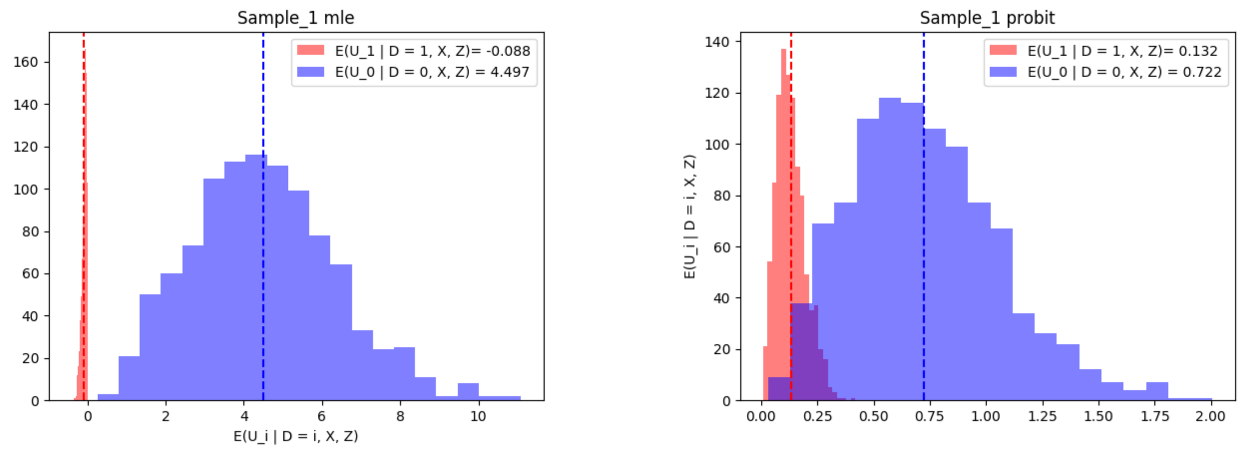


Figure 10: Sample 1 histogram of conditional expectations given (X, Z) observed in the data. Unconditional sample means $\mathbb{E}[U_1|D=1]$ and $\mathbb{E}[U_0|D=0]$ displayed in legends.

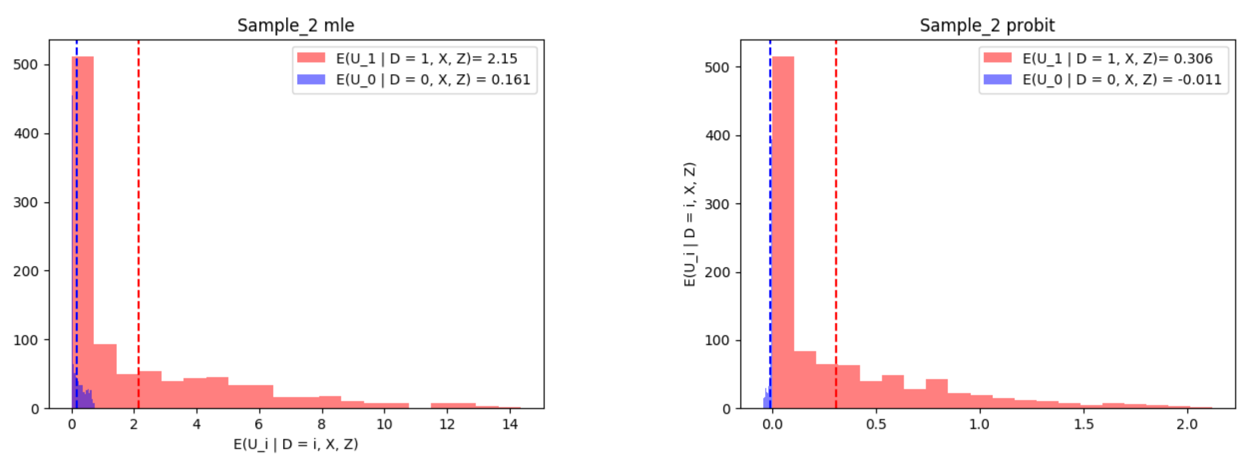


Figure 11: Sample 2 histogram of conditional expectations given (X, Z) observed in the data. Unconditional sample means $\mathbb{E}[U_1|D=1]$ and $\mathbb{E}[U_0|D=0]$ displayed in legends.

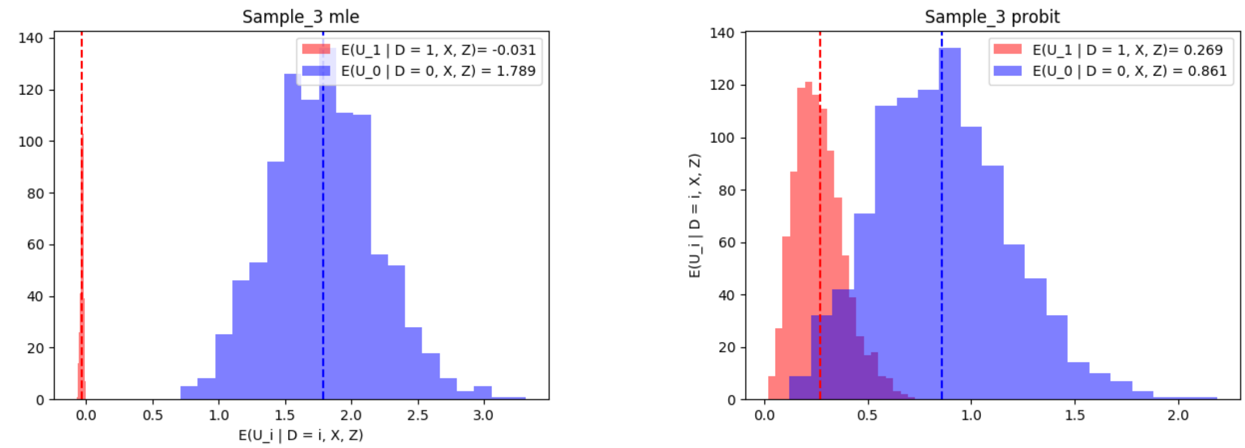


Figure 12: Sample 3 histogram of conditional expectations given (X, Z) observed in the data. Unconditional sample means $\mathbb{E}[U_1|D = 1]$ and $\mathbb{E}[U_0|D = 0]$ displayed in legends.

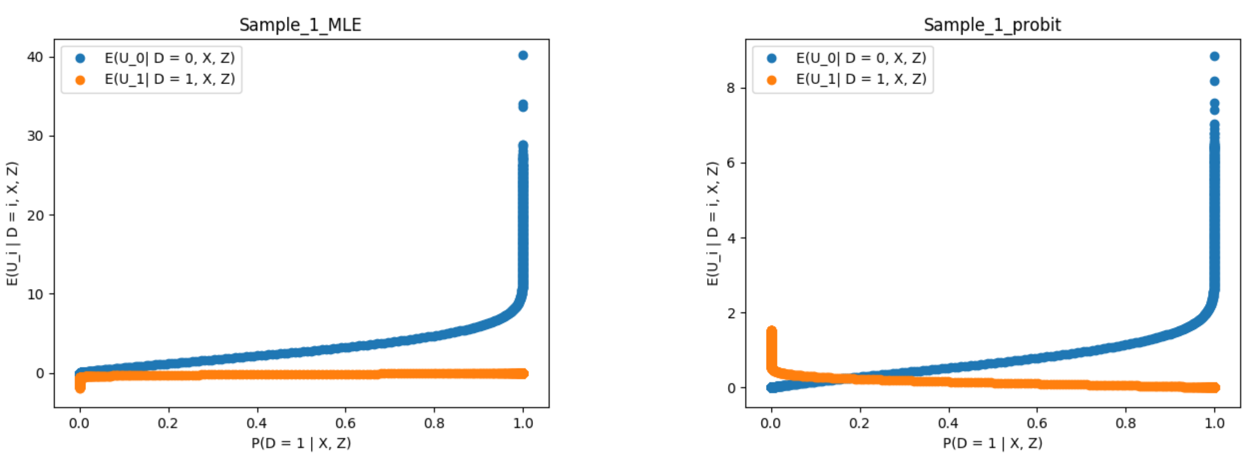


Figure 13: Sample 1 conditional expectations as functions of $P(D = 1|X, Z)$.

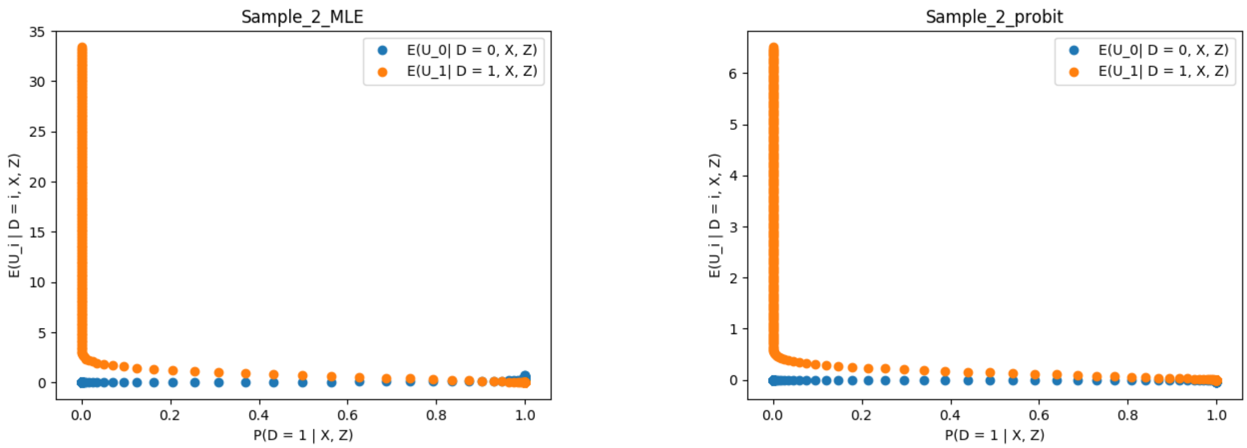


Figure 14: Sample 2 conditional expectations as functions of $P(D = 1|X, Z)$.

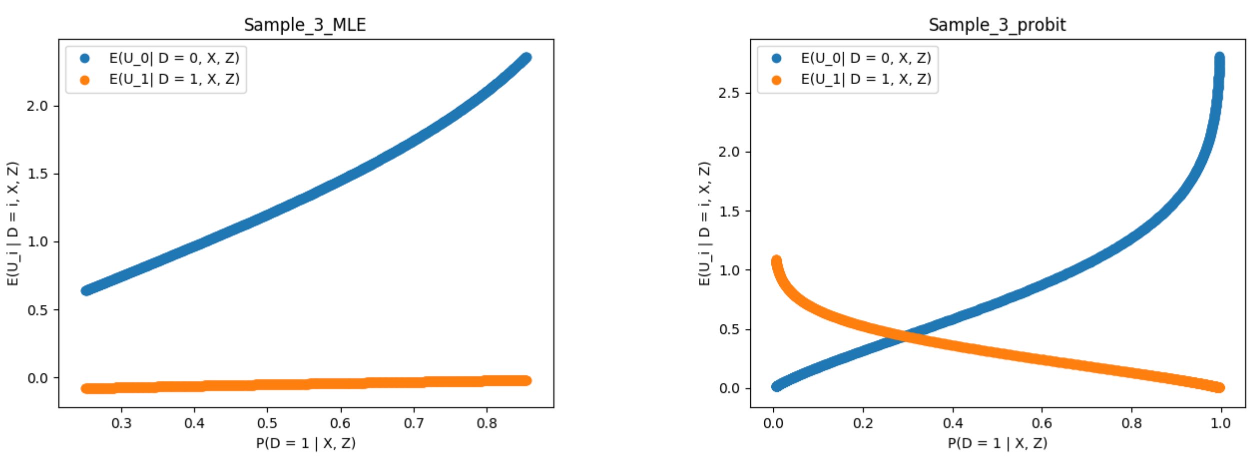


Figure 15: Sample 3 conditional expectations as functions of $P(D = 1|X, Z)$.

Problem 2.6. Estimate ATE using the "propensity score" you derived in (a). Plot the ATE as a function of the propensity score for each sample. What do these plots tell you about the appropriateness of the matching assumption for each sample?

Solution. ▷ For sample 1:

* The ATE estimated using propensity score is -0.2344:

```
Probit regression                               Number of obs     =      1,000
                                                LR chi2(1)        =     205.53
                                                Prob > chi2       =     0.0000
Log likelihood = -589.32546                     Pseudo R2         =     0.1485
```

d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d_hat_sample_1	2.879469	.2137283	13.47	0.000	2.460569	3.298369
_cons	-1.439859	.1190816	-12.09	0.000	-1.673255	-1.206463

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
y	Unmatched	1.04787356	1.17798679	-.130113232	.054884843	-2.37
	ATT	1.04787356	1.4724852	-.424611641	.07998632	-5.31
	ATU	1.17798679	1.15209088	-.025895911	.	.
	ATE			-.234424238	.	.

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	477	477
Treated	523	523
Total	1,000	1,000

Figure 16: ATE: Sample 1

* The graph of ATE as a function of propensity score (after smoothing):

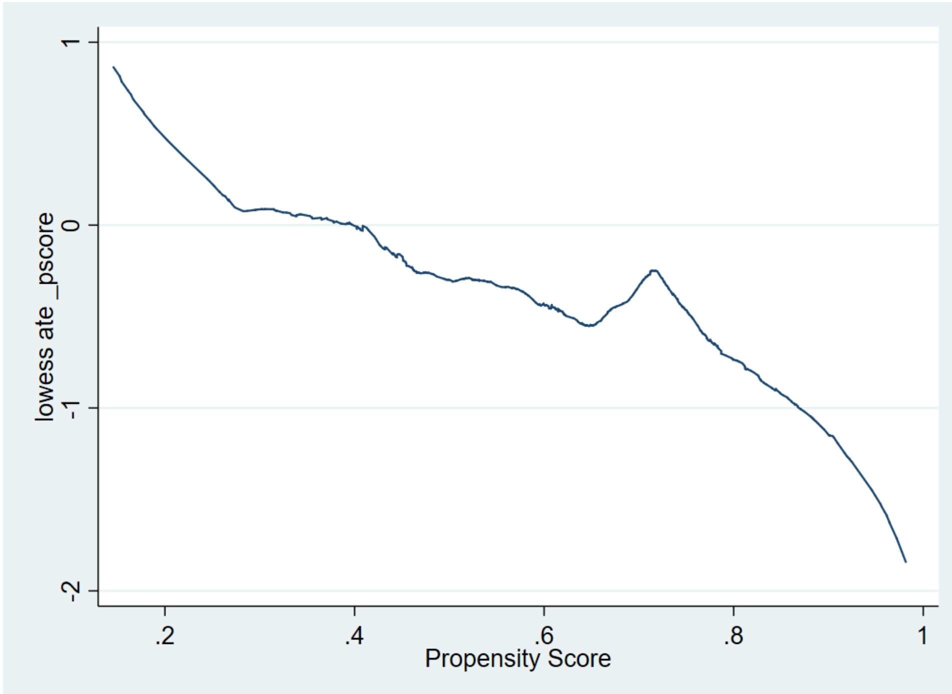


Figure 17: ATE as a function of P: Sample 1

* Since the local ATE is decreasing in propensity score, it is not consistent with our structural model, so the matching is inappropriate for sample 1.

▷ For sample 2:

* The ATE estimated using propensity score is 0.1486:

```
Probit regression                               Number of obs     =      1,000
                                                LR chi2(1)        =    1040.16
                                                Prob > chi2       =     0.0000
Log likelihood = -169.70161                     Pseudo R2         =     0.7540
```

d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d_hat_sample_2	4.398245	.2127816	20.67	0.000	3.981201	4.815289
_cons	-2.206514	.1346698	-16.38	0.000	-2.470462	-1.942566

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
y	Unmatched	.750351505	.460115107	.290236398	.063135137	4.60
	ATT	.750351505	.678422659	.071928846	.818852488	0.09
	ATU	.460115107	.699013384	.238898277	.	.
	ATE			.148567815	.	.

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	459	459
Treated	541	541
Total	1,000	1,000

Figure 18: ATE: Sample 2

* The graph of ATE as a function of propensity score (after smoothing):

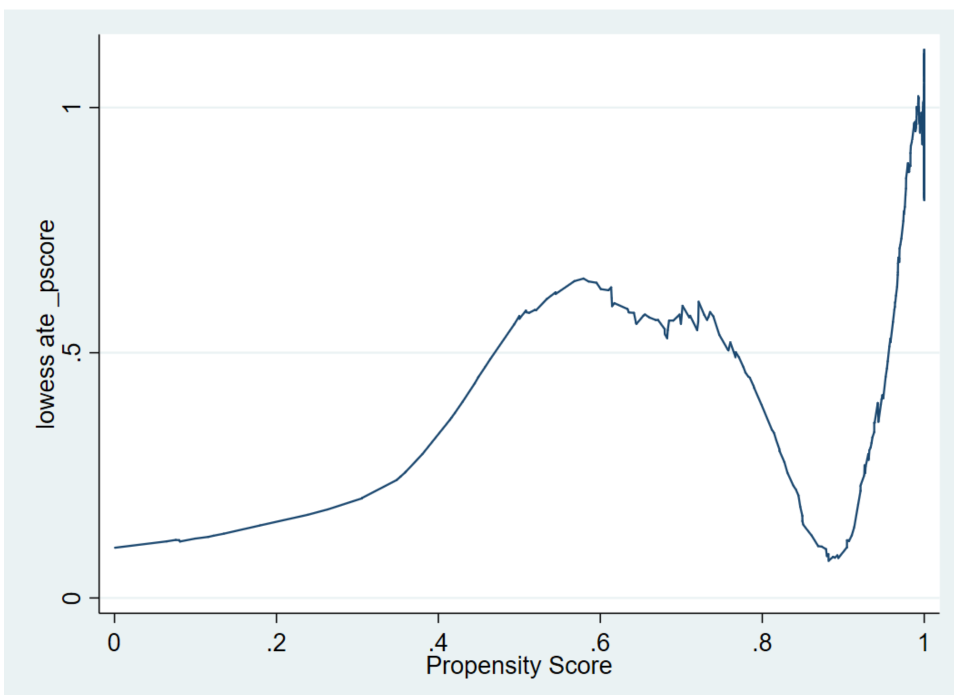


Figure 19: ATE as a function of P: Sample 2

* From previous result, we saw that the balancing is poor for sample 2, so matching is inappropriate for sample 2.

▷ For sample 3:

* The ATE estimated using propensity score is 0.1376:

Probit regression		Number of obs	=	1,000
		LR chi2(1)	=	150.07
		Prob > chi2	=	0.0000
Log likelihood = -609.63688		Pseudo R2	=	0.1096

d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d_hat_sample_3	2.785562	.2382459	11.69	0.000	2.318609	3.252515
_cons	-1.392983	.1389593	-10.02	0.000	-1.665338	-1.120628

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
y	Unmatched	2.53142337	2.32359371	.207829658	.073540634	2.83
	ATT	2.53142337	2.12536772	.406055642	.117119119	3.47
	ATU	2.32359371	2.11245262	-.211141088	.	.
	ATE			.137575064	.	.

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On suppor	Total
Untreated	435	435
Treated	565	565
Total	1,000	1,000

Figure 20: ATE: Sample 3

* The graph of ATE as a function of propensity score (after smoothing):

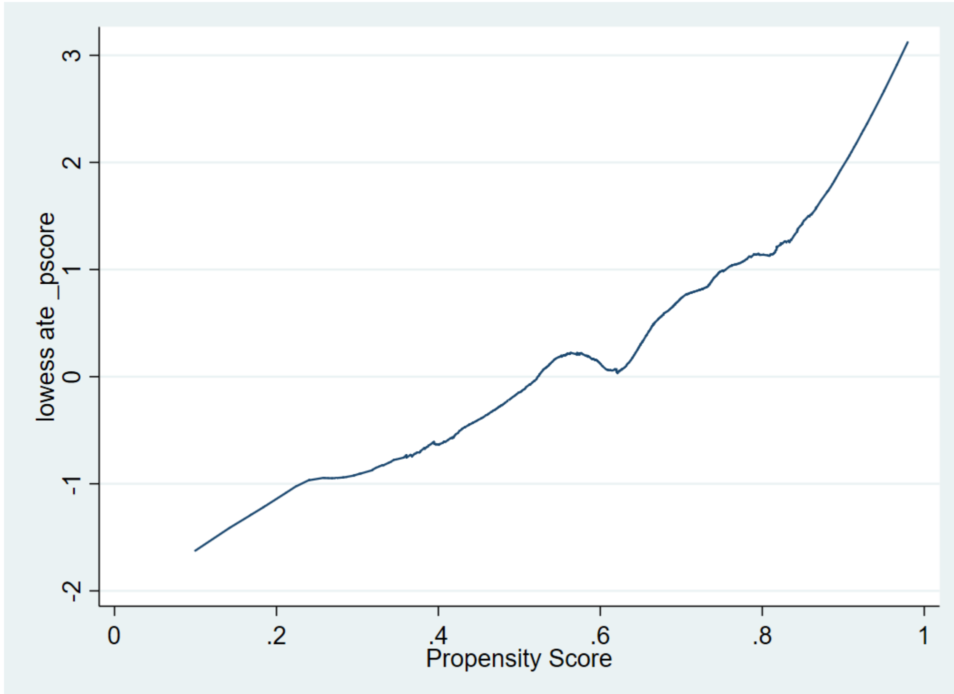


Figure 21: ATE as a function of P: Sample 3

* The local ATE is increasing in propensity score, which is consistent with our structural model, and there is no balancing or common support problem with sample 3, so matching is appropriate for sample 3.

Problem 2.7. Estimate ATE using a regression of Y on $P : X, Z$. Plot your estimate against P.

Solution. For sample 1:

▷ The ATE estimated using regression is:

```
Linear regression                       Number of obs     =      1,000
                                        F(3, 996)         =       4.94
                                        Prob > F          =     0.0021
                                        R-squared         =     0.0141
                                        Root MSE          =     .86403
```

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_pscore	2.076248	1.147934	1.81	0.071	-.1763982	4.328895
z1	.2686832	.2040203	1.32	0.188	-.1316758	.6690422
z2	.2897412	.1691989	1.71	0.087	-.0422859	.6217684
_cons	.0687576	.6753133	0.10	0.919	-1.256443	1.393958

Figure 22: ATE: Sample 1

* Since we can only obtain a point estimate for the ATE from the regression, it is not possible to plot the ATE estimate against any variable. Instead, we just report the graph of ATE as a function of propensity score, which is the same as in previous part. For exactly the same reason, the same corresponding graphs are reported below for sample 2 and 3.

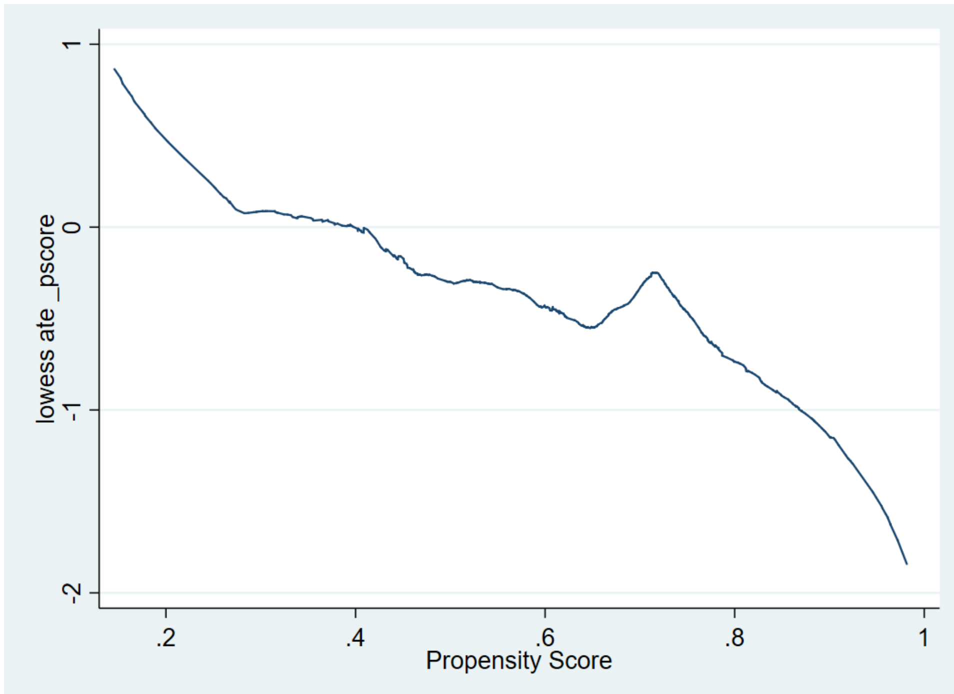


Figure 23: ATE as a function of P: Sample 1

▷ For sample 2:

* The ATE estimated using regression is:

Linear regression

Number of obs	=	1,000
F(3, 996)	=	4.65
Prob > F	=	0.0031
R-squared	=	0.0141
Root MSE	=	.99926

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_pscore	.345177	.1462616	2.36	0.018	.0581608	.6321933
z1	.0165885	.0531696	0.31	0.755	-.0877489	.1209258
z2	.0463074	.0515257	0.90	0.369	-.054804	.1474188
_cons	.4880723	.1254733	3.89	0.000	.24185	.7342947

Figure 24: ATE: Sample 2

* The graph of ATE as a function of propensity score:

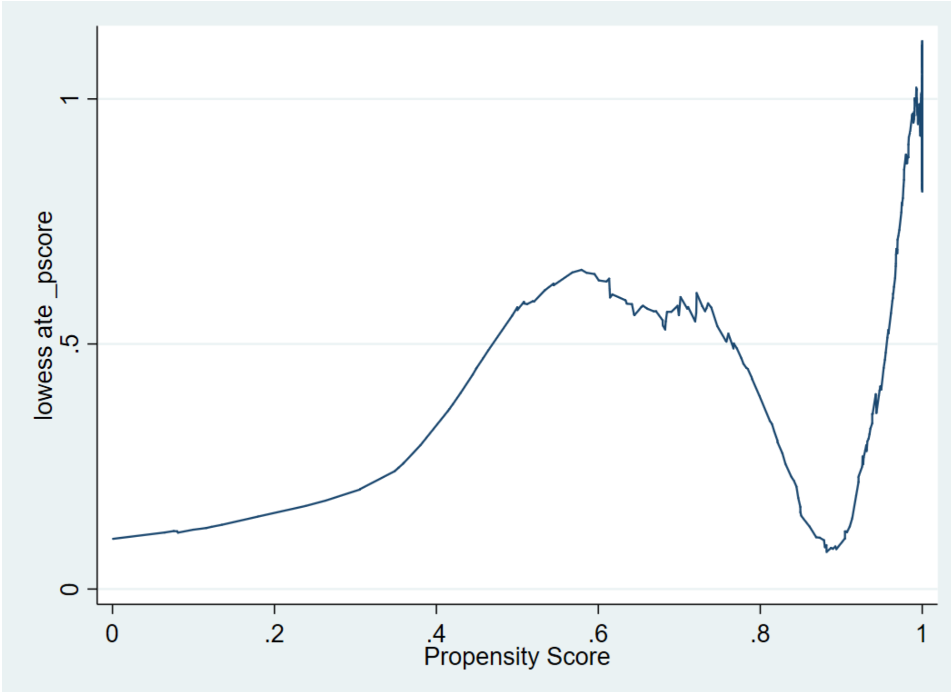


Figure 25: ATE as a function of P: Sample 2

▷ For sample 3:

* The ATE estimated using regression is:

Linear regression

Number of obs	=	1,000
F(3, 996)	=	53.89
Prob > F	=	0.0000
R-squared	=	0.1539
Root MSE	=	1.0658

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_pscore	-4.815707	1.812739	-2.66	0.008	-8.372933	-1.25848
x1	1.067232	.2517703	4.24	0.000	.573171	1.561293
x2	-.5029072	.2348416	-2.14	0.032	-.9637482	-.0420661
_cons	2.4484	.7526888	3.25	0.001	.9713627	3.925438

Figure 26: ATE: Sample 3

* The graph of ATE as a function of propensity score:

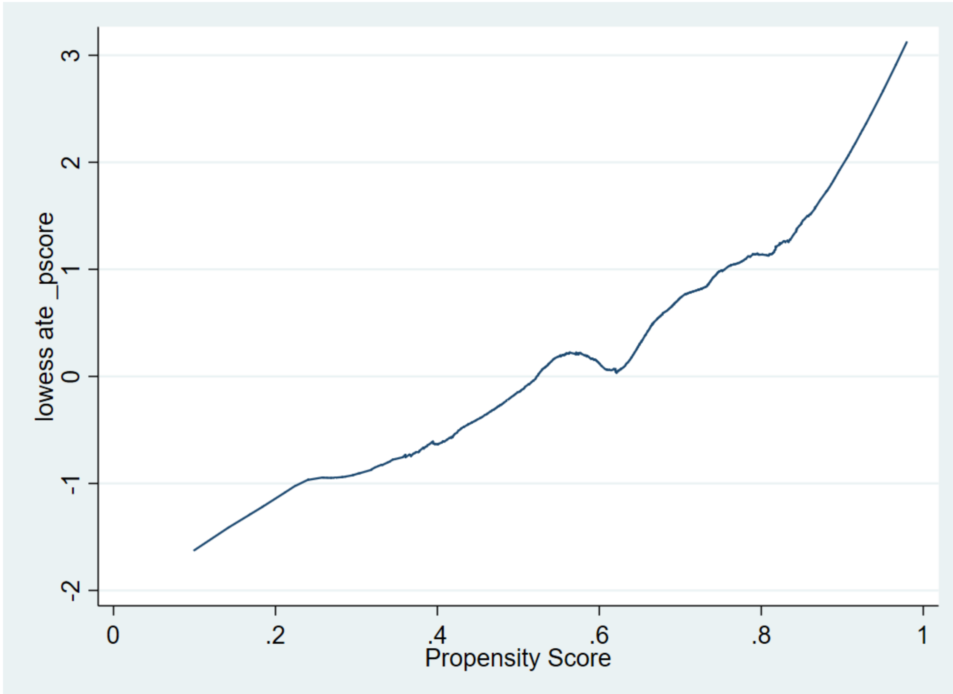


Figure 27: ATE as a function of P: Sample 3

3 Question 3

Using your estimates from MLE, regression and matching just obtained, compute:

Problem 3.1. The MTE as a function of $V = U_0 - U_1 + U_Z$.

Solution. Note that

$$\begin{aligned} ATE(X) &= \mathbb{E}[Y_1 - Y_0|X] \\ &= (\alpha_1 - \alpha_0)'X + \beta, \end{aligned}$$

by exogeneity. $ATE(X)$ can be identified from either MLE or the regression in part d. Then

$$\begin{aligned} MTE(X,Z,V) &= \mathbb{E}[Y_1 - Y_0|V,X,Z] \\ &= ATE(X) + \mathbb{E}[U_1 - U_0|V,X,Z] \\ &= ATE(X) + \mathbb{E}[U_1 - U_0|V], \text{ by exogneity} \end{aligned}$$

We have

$$\begin{pmatrix} U_1 - U_0 \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 + \sigma_0^2 - 2\sigma_{10} & \sigma_{1V} - \sigma_{0V} \\ \sigma_{1V} - \sigma_{0V} & \sigma_V^2 \end{pmatrix} \right).$$

Thus,

$$\mathbb{E}[U_1 - U_0|V] = \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V^2} V,$$

and so

$$MTE(X,Z,V) = (\alpha_1 - \alpha_0)'X + \beta + \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V} \frac{V}{\sigma_V}.$$

MLE identifies all of the variance parameters; in particular, MLE identifies $\sigma_{1V}, \sigma_{0V}, \sigma_V$ individually, so $MTE(X,Z,V)$ can be calculated for arbitrary (X,Z,V). However, the regression in part d only identifies σ_{1V}/σ_V and σ_{0V}/σ_V, so we can only calculate the MTE at arbitrary $(X, V/\sigma_V)$.

To obtain the MTE from the matching estimates, we can write the MTE as a function of X, Z, and the propenstiy score $P(X,Z) = p$ (since the marginal individual is he who is indifferent between taking treatement and not):

$$\begin{aligned} MTE(X,Z,p) &= \mathbb{E}\left[(Y_1 - Y_0)\,|P(X,Z) = p, X = x\right], \text{ where } P(X,Z) = p \text{ identifies the marginal individual} \\ &= ATE(X), \end{aligned}$$

where the second equality follows from $(U_1, U_0, U_Z) \perp (X,Z)$. Figure 28 displays all of the MTE estimates graphically.

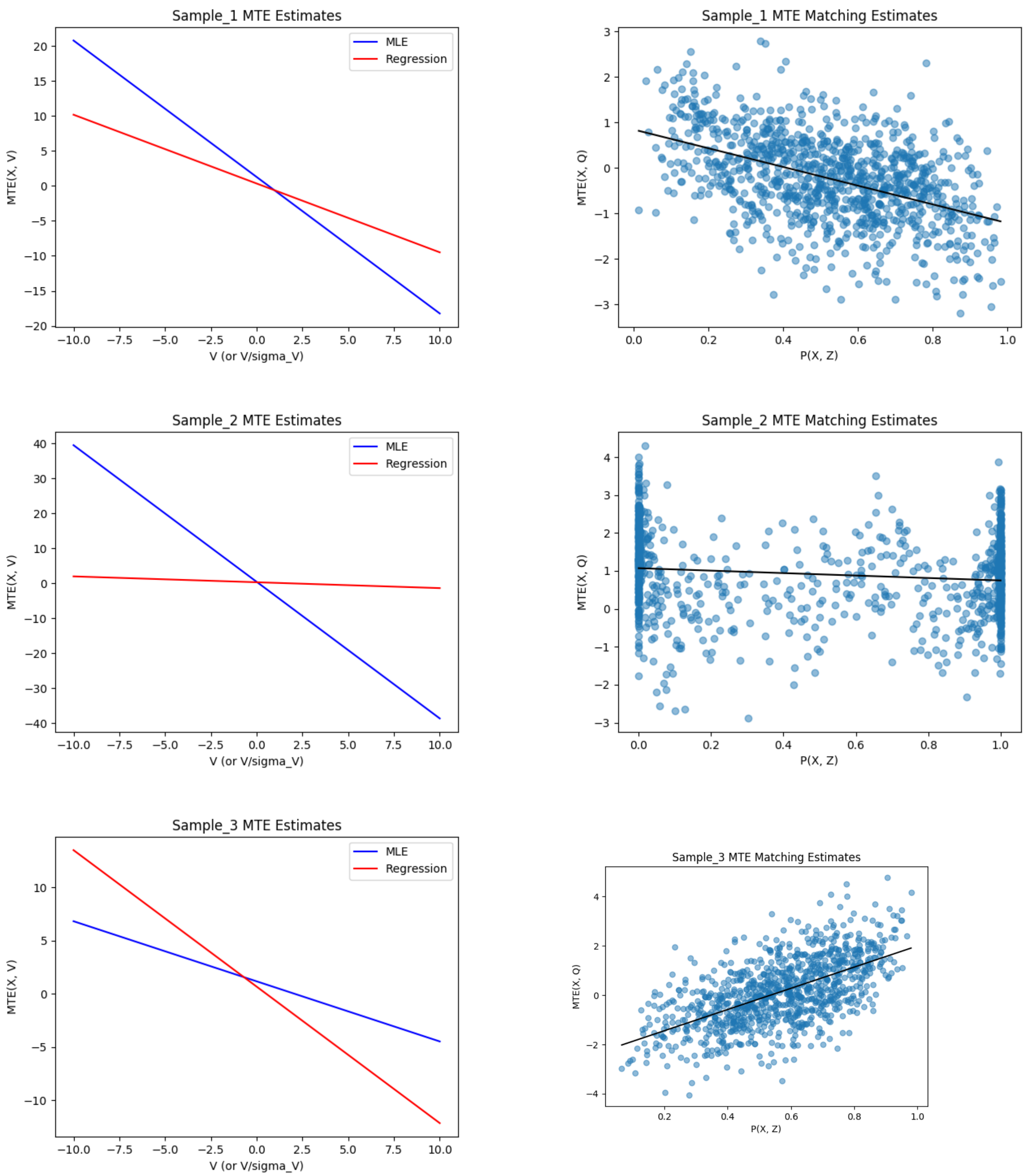


Figure 28: MTE estimates using parameters from MLE, the part d regression, and matching for Samples 1, 2, and 3.

Problem 3.2. The policy relevant treatment effect for a 10% upward shift in all arguments of Z.

Solution. We first derive an explicit expression for $PRTE(X, Z)$, the expected effect of this policy on an individual with observables (X, Z), and then calculate the unconditional $PRTE$ by taking the average of $PRTE(X, Z)$ in the sample.

Let

$$\begin{aligned} D^1 &= 1\{\beta + (\alpha_1 - \alpha_0)'X - \gamma Z > V\} \\ D^2 &= 1\{\beta + (\alpha_1 - \alpha_0)'X - 1.1\gamma Z > V\}. \end{aligned}$$

Assume $\gamma < 0$ (since the sample 3 MLE estimate is negative. This is a convenience assumption, if $\gamma >= 0$ nothing substantive changes, we just flip some signs). So D^1 and D^2 represent the treatments for this individual under the baseline and proposed polices, respectively. Let

$$Y^i = D^i(Y_1 - Y_0) + Y_0,$$

be the outcome for this individual under policy $i = 1, 2$.

Then, since it is impossible for $D^2 < D^1$, we can write

$$\begin{aligned} PRTE(X, Z) &= \mathbb{E}[Y^2 - Y^1|X, Z] \\ &= \mathbb{E}[D^2(Y_1 - Y_0) + Y_0 - D^1(Y_1 - Y_0) - Y_0|X, Z] \\ &= \mathbb{E}[(D^2 - D^1)(Y_1 - Y_0)|X, Z] \\ &= \mathbb{E}[Y_1 - Y_0|D^2 > D^1, X, Z]Pr(D^2 > D^1). \end{aligned}$$

Note that

$$Pr(D^2 > D^1) = Pr(\beta + (\alpha_1 - \alpha_0)'X - 1.1\gamma Z > V > \beta + (\alpha_1 - \alpha_0)'X - \gamma Z).$$

Letting $c_2 = \beta + (\alpha_1 - \alpha_0)'X - 1.1\gamma Z$ and $c_1 = \beta + (\alpha_1 - \alpha_0)'X - \gamma Z$, we get

$$Pr(D^2 > D^1) = \Phi(\frac{c_2}{\sigma_v}) - \Phi(\frac{c_1}{\sigma_v}).$$

On the other hand,

$$\begin{aligned} \mathbb{E}[Y_1 - Y_0|D^2 > D^1, X, Z] &= \mathbb{E}[(\alpha_1 - \alpha_0)'X + \beta + U_1 - U_0|c_2 > V > c_1], \text{ by exogeneity} \\ &= (\alpha_1 - \alpha_0)'X + \beta + \mathbb{E}[U_1 - U_0|c_2 > V > c_1]. \end{aligned}$$

Note that

$$\begin{pmatrix} U_1 - U_0 \\ V \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 + \sigma_0^2 - 2\sigma_{10} & \sigma_{1V} - \sigma_{0V} \\ \sigma_{1V} - \sigma_{0V} & \sigma_V^2 \end{pmatrix}\right).$$

Then

$$U_1 - U_0 = \frac{\sigma_{1V} - \sigma_{0v}}{\sigma_V^2} V + \eta, \eta \perp V. \tag{1}$$

	MLE	Regression
Sample 1	0.00787489	0.0115326
Sample 2	0.00218072	0.00241555
Sample 3	0.00587851	0.00482535

Table 5: PRTE estimates

So

$$\begin{aligned}\mathbb{E}[U_1 - U_0 | c_2 > V > c_1] &= \frac{\sigma_{1V} - \sigma_{0v}}{\sigma_V^2}\mathbb{E}[V|c_2 > V > c_1] \\ &= \frac{\sigma_{1V} - \sigma_{0v}}{\sigma_V}\frac{\phi\left(\frac{c_2}{\sigma_V}\right) - \phi\left(\frac{c_1}{\sigma_V}\right)}{\Phi\left(\frac{c_1}{\sigma_V}\right) - \Phi\left(\frac{c_2}{\sigma_V}\right)}\end{aligned}$$

Therefore,

$$PRTE(X,Z) = \left[(\alpha_1 - \alpha_0)'X + \beta + \frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V}\frac{\phi\left(\frac{c_2}{\sigma_V}\right) - \phi\left(\frac{c_1}{\sigma_V}\right)}{\Phi\left(\frac{c_1}{\sigma_V}\right) - \Phi\left(\frac{c_2}{\sigma_V}\right)}\right] \cdot \left(\Phi\left(\frac{c_2}{\sigma_v}\right) - \Phi\left(\frac{c_1}{\sigma_v}\right)\right).$$

Lastly,

$$PRTE = \mathbb{E}[PRTE(X,Z)].$$

Table 5 displays the PRTE estimates. Note that we cannot estimate this quantity from the parameters yielded by matching, since the term

$$\frac{\sigma_{1V} - \sigma_{0V}}{\sigma_V}$$

is not estimated in the matching.

Problem 3.3. Estimate β from all 3 samples using IV (i) using Z_1 as an instrument, (ii) using Z_2 as an instrument and (iii) using $Pr(D = 1|X < Z)$ as an instrument. Interpret your estimates.

Solution. The second stage of the 2SLS we want to run is

$$\begin{aligned}Y &= D(Y_1 - Y_0) + Y_0 + \epsilon_2 \\ &= D((\alpha_1 - \alpha_0)'X + \beta + U_1 - U_0) + \alpha_0'X + U_0 + \epsilon_2 \\ &= (\alpha_1 - \alpha_0)'XD + \beta D + \alpha_0'X + \epsilon^*,\end{aligned}$$

where ϵ^* is correlated with D. So we use the following first stage

$$D = \Pi_1 X + \Pi_2 Z + \epsilon_1,$$

	Sample 1			Sample 2			Sample 3
	Z1	Z2	P(D=1| X, Z)	Z1	Z2	P(D=1| X, Z)	P(D=1| X, Z)
$(\alpha_1 - \alpha_0)'X + \beta$	0.535 (2.952)	0.04 (0.192)	0.356 (2.675)	0.286 (2.369)	0.171 (1.49)	0.258 (3.651)	-
α_0	0.83 (8.534)	1.089 (9.688)	0.924 (12.664)	0.463 (6.178)	0.525 (7.448)	0.478 (9.122)	[0.049 0.551]([0.285 7.251])
$\alpha_1 - \alpha_0$	-	-	-	-	-	-	[0.449 -0.601]([3.878 -4.209])
β	-	-	-	-	-	-	0.836 (1.904)

Table 6: 2SLS estimates (t-stats in parentheses).

where Z is the instrument under consideration.

Note that in samples 1 and 2, since there is no variation in $X = 1$, we will only be able to identify $(\alpha_1 - \alpha_0) + \beta$ as the second stage coefficient on D (LATE) and α_0 as the second stage intercept term.

In sample 3, however, we know $(\alpha_1 - \alpha_0)$ is the second stage coefficient on the interaction term XD, α_0 is the second stage coefficient on X, and β is the second stage coefficient on D (LATE). Since having XD in the second stage introduces two new endogenous variables (X_1D and X_2D), we use three instruments: $P(D = 1|X, Z), P(D = 1|X, Z) \cdot X_1, P(D = 1|X, Z) \cdot X_2$.

Since we have multiple levels (a continuum) of treatment here, the LATE (second stage coefficient on D) interpretation is an instrument-strength weighted average of the instrument-specific LATE values. So for example, for sample 1 and instrument Z_1 taking treatment increases Y by $.535$ "on average" (under a weird weighting scheme).

Table 6 displays these 2SLS estimates.

Problem 3.4. Compare your estimate LATE using $Pr(D = 1|X, Z)$ as an instrument with the derivation of $\mathbb{E}(Y_1 - Y_0|X, Z)$formed from your answer to 2(d) and 2(e) above.

Solution. Note that by exogeneity

$$\mathbb{E}(Y_1 - Y_0|X, Z)] = (\alpha_1 - \alpha_0)'X + \beta.$$

Table 7 displays the values of the IV LATE estimates from part b and the conditional potential outcome difference (averaged across the whole sample in the case of sample 3). Note that the values do not match, as we would expect. LATE using the propensity score is

$$\mathbb{E}(Y_1|D = 1, X, Z)] - \mathbb{E}(Y_0|D = 0, X, Z)] = (\alpha_1 - \alpha_0)'X + \beta + \mathbb{E}[U_1|D = 1, X, Z)] - \mathbb{E}[U_0|D = 0, X, Z)],$$

where the conditional error terms are as we calculated in part e. The difference here is due to the fact that the LATE accounts for selection while a simple conditional expectation does not.

	IV Difference	Parts e,d Difference
Sample_1	-0.568292	0.339616
Sample_2	-0.219878	0.260807
Sample_3	0.836495	-0.237486

Table 7: Table of LATE (IV Difference) and conditional potential outcome differences (Parts e,d Difference).

4 Question 4

Suppose that

$$\gamma \sim N(\bar{\gamma}, \Sigma_\gamma)$$

$$\gamma \perp (U_0, U_1, U_Z, X, Z), Z \perp (U_1 - U_0 - U_Z)$$

Problem 4.1. Does the Imbens-Angrist monotonicity condition hold? Prove or disprove.

Solution. The monotonicity condition in Imbens and Angrist (1994) states that $D_\omega(Z=0) \leq D_\omega(Z=1)$ for all individuals indexed by ω. We are assuming that we are checking the monotonicity condition given a sample.

▷ Recall that D is defined as

$$\begin{aligned} D &= 1\{Y_1 - Y_0 - C(Z) \geq 0\} \\ &= 1\{(\alpha_1 - \alpha_0)'X + \beta - \gamma Z + U_1 - U_0 - U_Z \geq 0\} \end{aligned}$$

▷ Suppose $D_\omega(Z=1) = 0$ for a given realization of γ, U_1, U_0, and U_Z. This is equivalent to asserting:

$$(\alpha_1 - \alpha_0)'X + \beta - \gamma + U_1 - U_0 - U_Z < 0$$

Then if γ is positive with magnitude $|\gamma| > |(\alpha_1 - \alpha_0)'X + \beta + U_1 - U_0 - U_Z|$, we have that

$$(\alpha_1 - \alpha_0)'X + \beta + U_1 - U_0 - U_Z \geq 0$$

in which case $D_\omega(Z=0) = 1$.

Thus, the case in which

$$1 = D_\omega(Z=0) \nleq D_\omega(Z=1) = 0$$

arises given the conditions, and this happens because γ can take large enough positive values given that its realization is drawn from a normal distribution. Therefore, the monotonicity condition **does not** hold. (Note that if we focus on the monotonicity condition for the population, then the condition will not hold for $\gamma > 0$.

■