1. Numerical equivalence of the indirect least squares and two-stage least squares estimators

Consider the following canonical regression with one endogenous regressor D and one instrumental variable Z:

$$\begin{aligned} Y &= \tau D + X\beta + \varepsilon_1, \\ D &= \alpha Z + X\gamma + \varepsilon_2, \end{aligned}$$

where $Y, D, Z, \varepsilon_1, \varepsilon_2$ are $n \times 1$ vectors, and X is an $n \times p$ matrix. If D is endogenous, then the OLS fit for the first equation gives a biased estimator for τ. Instead, we can use either the indirect least squares or the two-stage least squares estimator.

The indirect least squares estimator has three steps: first, fit the OLS of Y on Z and X and get the coefficient of Z, denoted by $\hat{\theta}$; second, fit the OLS of D on Z and X and get the coefficient of Z, denoted by $\hat{\alpha}$; third, the indirect least squares estimator is the ratio $\hat{\tau}_{\text{ils}} = \hat{\theta}/\hat{\alpha}$.

The two-stage least squares estimator has two steps: first, fit the OLS of D on Z and X and obtain the fitted vector $\hat{D}$; second, fit the OLS of Y on $\hat{D}$ and X and obtain the coefficient of $\hat{D}$, denoted by $\hat{\tau}_{\text{tsls}}$.

Many textbooks claim that $\hat{\tau}_{\text{ils}} = \hat{\tau}_{\text{tsls}}$ without a formal proof. Note that this is a linear algebra fact without assuming any modeling assumptions. We can verify this using the following simple numerical example.

```
> n = 10^5
> u = rnorm(n)
> v = rnorm(n)
> x = matrix(rnorm(n*2), n, 2)
> z = rnorm(n)
> d = z + as.vector(x%*%c(1, 2)) + u
> y = d + as.vector(x%*%c(1, -1)) + u + v
> summary(lm(y ~ d + x))$coef[2]
```

```
[1] 1.500528
> summary(lm(y ~ z + x))$coef[2]/summary(lm(d ~ z + x))$coef[2]
[1] 1.001864
> dhat = lm(d ~ z + x)$fitted.values
> summary(lm(y ~ dhat + x))$coef[2]
[1] 1.001864
```

Now prove that $\hat{\tau}_{\text{ils}} = \hat{\tau}_{\text{tsls}}$.

Proof. Define $H_X = X(X'X)^{-1}X'$ as the hat matrix of X. We will use the following basic properties of the projection matrix $I - H_X$: $(I - H_X)X = 0$, $(I - H_X)' = (I - H_X)$ and $(I - H_X)^2 = (I - H_X)$.

First, we have

$$\begin{aligned}
\hat{\theta} &= \text{coefficient of } Z \text{ in the OLS fit of } Y \text{ on } Z \text{ and } X \\
&\overset{FWL}{=} \text{coefficient of } (I - H_X)Z \text{ in the OLS fit of } (I - H_X)Y \text{ on } (I - H_X)Z \\
&= \frac{Z'(I - H_X)Y}{Z'(I - H_X)Z},
\end{aligned}$$

and

$$\begin{aligned}
\hat{\alpha} &= \text{coefficient of } Z \text{ in the OLS fit of } D \text{ on } Z \text{ and } X \\
&\overset{FWL}{=} \text{coefficient of } (I - H_X)Z \text{ in the OLS fit of } (I - H_X)D \text{ on } (I - H_X)Z \\
&= \frac{Z'(I - H_X)D}{Z'(I - H_X)Z}.
\end{aligned}$$

The indirect least squares estimator is then

$$\hat{\tau}_{\text{ils}} = \frac{\hat{\theta}}{\hat{\alpha}} = \frac{Z'(I - H_X)Y}{Z'(I - H_X)D}.$$

The two-stage least squares estimator is

$$
\begin{aligned}
\hat{\tau}_{\text{ils}} &= \text{coefficient of } \hat{D} \text{ in the OLS fit of } Y \text{ on } \hat{D} \text{ and } X \\
&\overset{FWL}{=} \text{coefficient of } (I - H_X)\hat{D} \text{ in the OLS fit of } (I - H_X)Y \text{ on } (I - H_X)\hat{D} \\
&= \frac{\hat{D}'(I - H_X)Y}{\hat{D}'(I - H_X)\hat{D}}.
\end{aligned}
$$

The fitted vector $\hat{D}$ can be written as $\hat{D} = \hat{\alpha}Z + X\hat{\gamma}$ from the OLS of D on Z and X. Therefore, we can further write the two-stage least squares estimator as

$$
\begin{aligned}
\hat{\tau}_{\text{ils}} &= \frac{(\hat{\alpha}Z + X\hat{\gamma})'(I - H_X)Y}{(\hat{\alpha}Z + X\hat{\gamma})'(I - H_X)(\hat{\alpha}Z + X\hat{\gamma})} \\
&= \frac{\hat{\alpha}Z'(I - H_X)Y}{\hat{\alpha}^2 Z'(I - H_X)Z} \\
&= \frac{Z'(I - H_X)Y}{\hat{\alpha}Z'(I - H_X)Z},
\end{aligned}
$$

which equals the indirect least squares because $\hat{\alpha} = Z'(I - H_X)D/Z'(I - H_X)Z$ as I show above.