Problem Set 6

1. (a) Instruments should be used when one of the regressors is endogenous. The IV estimation works well, if the instrument is highly correlated with the endogenous regressor x (relevance) and uncorrelated with u (exogeneity). The partial effect $\frac{dy}{dx}$ is still of interest and can be estimated by means of IV:

$$\beta_{IV} = \frac{dy/dz}{dx/dz} = \frac{dy}{dx},$$

Then,

$$\hat{\beta}_{IV} = \frac{(Z'Z)^{-1}Z'y}{(Z'Z)^{-1}Z'X} = (Z'X)^{-1}Z'y$$

If the instrument fulfills all necessary assumptions, $\hat{\beta}_{IV}$ is consistent in the presence of an endogenous regressor. If the instrument is equal to the regressor itself, $Z = X$, the IV estimator is equivalent to the OLS estimator.

(b)

$$\begin{aligned} \hat{\beta}_{OLS} &= (X'X)^{-1}X'y = \begin{pmatrix} 20 & 10 \\ 10 & 20 \end{pmatrix}^{-1} \begin{pmatrix} 12.5 \\ 10.5 \end{pmatrix} = \begin{pmatrix} 0.4833 \\ 0.2833 \end{pmatrix} \\ \hat{\beta}_{IV} &= (Z'X)^{-1}Z'y = \begin{pmatrix} 10 & 15 \\ 10 & 20 \end{pmatrix}^{-1} \begin{pmatrix} 8.5 \\ 10.5 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.4 \end{pmatrix} \end{aligned}$$

$$Z = \begin{pmatrix} z_1 & x_{2,1} \\ \vdots & \vdots \\ z_N & x_{2,N} \end{pmatrix} \quad \Rightarrow \quad \hat{\beta}_{IV} = \begin{pmatrix} \sum z_i x_{1,i} & \sum z_i x_{2,i} \\ \sum x_{2,i} x_{1,i} & \sum x_{2,i}^2 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum z_i y_i \\ \sum x_{2,i} y_i \end{pmatrix}$$

2. (a) Two reasons are at hand. Maybe people with certain (chronic) diseases are not able to exercise. Hence, there is a problem of simultaneity. If we assume that chronically ill people are not in the sample (or if we can control for this), there remains a possibility that people who exercise more have a healthier lifestyle in general. Accordingly, *health* and *exercise* would be jointly driven by a latent unobservable factor 'healthy lifestyle' (containing, for example, nutrition).

(b) Those variables would be endogenous if explanatory variables for *health*, that are not included in this regression, are also explanatory for the location of your work or home. But it is rather unlikely that this is the case (for example, if smokers often choose flats in the North of Kiel).

(c) $\boldsymbol{\beta}$ is identified if we can express it in population moments. In general, $\boldsymbol{\beta} = \mathrm{E}[\mathbf{z}'\mathbf{x}]^{-1}\,\mathrm{E}[\mathbf{z}'\mathbf{y}]$. If the $\mathrm{E}[\mathbf{z}'\mathbf{x}]$ is not invertible, β cannot be written in terms of population moments and is thereby not identified. $\mathrm{E}[\mathbf{z}'\mathbf{x}]$ is not invertible if it is not of full rank. Let us build $\mathrm{E}[\mathbf{z}'\mathbf{x}]$ for our example using *disthome* as an instrument: $x_6 = \textit{exercise}$, $z_1 = \textit{disthome}$

$$\mathbf{z}' = \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ z_1 \end{pmatrix} \qquad \mathbf{x} = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{pmatrix}$$

$$\mathbf{z}'\mathbf{x} = \begin{pmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ x_1 & x_1^2 & x_1x_2 & x_1x_3 & x_1x_4 & x_1x_5 & x_1x_6 \\ x_2 & x_2x_1 & x_2^2 & x_2x_3 & x_2x_4 & x_2x_5 & x_2x_6 \\ x_3 & x_3x_1 & x_3x_2 & x_3^2 & x_3x_4 & x_3x_5 & x_3x_6 \\ x_4 & x_4x_1 & x_4x_2 & x_4x_3 & x_4^2 & x_4x_5 & x_4x_6 \\ x_5 & x_5x_1 & x_5x_2 & x_5x_3 & x_5x_4 & x_5^2 & x_5x_6 \\ z_1 & z_1x_1 & z_1x_2 & z_1x_3 & z_1x_4 & z_1x_5 & z_1x_6 \end{pmatrix}$$

Reduced form for *exercise* (x_6)

$$x_6 = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5 + \gamma_6 z_1 + r$$

Now assume that $\gamma_6 = 0$ in a reduced form, meaning, the instrument is uncorrelated with x_6, plug it in for x_6 in the matrix above and take the expectations $\mathrm{E}[\mathbf{z}'\mathbf{x}]$:

$$\mathrm{E}[\mathbf{z}'\mathbf{x}] = \begin{pmatrix} 1 & \mathrm{E}[x_1] & \dots & \mathrm{E}[x_5] & \gamma_0 + \gamma_1 \mathrm{E}[x_1] + \cdots + \gamma_5 \mathrm{E}[x_5] \\ \mathrm{E}[x_1] & \mathrm{E}[x_1^2] & \dots & \mathrm{E}[x_1x_5] & \gamma_0 \mathrm{E}[x_1] + \gamma_1 \mathrm{E}[x_1^2] + \cdots + \gamma_5 \mathrm{E}[x_1x_5] \\ \mathrm{E}[x_2] & \mathrm{E}[x_2x_1] & \dots & \mathrm{E}[x_2x_5] & \gamma_0 \mathrm{E}[x_2] + \gamma_1 \mathrm{E}[x_2x_1] + \cdots + \gamma_5 \mathrm{E}[x_2x_5] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{E}[x_5] & \mathrm{E}[x_5x_1] & \dots & \mathrm{E}[x_5^2] & \gamma_0 \mathrm{E}[x_5] + \gamma_1 \mathrm{E}[x_5x_2] + \cdots + \gamma_5 \mathrm{E}[x_5^2] \\ \mathrm{E}[z_1] & \mathrm{E}[z_1x_1] & \dots & \mathrm{E}[z_1x_5] & \gamma_0 \mathrm{E}[z_1] + \gamma_1 \mathrm{E}[z_1x_1] + \cdots + \gamma_5 \mathrm{E}[z_1x_5] \end{pmatrix}$$

The last column is a linear combination of the first $k-1$ columns. Therefore, $\mathrm{E}[\mathbf{z}'\mathbf{x}]$ is not of full rank and is singular. By that, $\boldsymbol{\beta}$ is not identified. Identification conditions is thereby $\gamma_6 \neq 0$. In order to check such condition, run a t-test for the hypothesis $H_0 : \gamma_6 = 0$ based on an OLS regression of the reduced form.

(d) Identification conditions from the reduced form with both instruments

$$x_6 = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5 + \gamma_6 z_1 + \gamma_7 z_2 + r$$

would constitute that least one of the γ_6, γ_7 must be nonzero. We could run an F-test (or a Wald test) for the joint hypothesis $H_0; \gamma_6 = \gamma_7 = 0$ based on an OLS regression of the reduced form.

3. (a) We have to use instruments for the endogenous variable x_K to get consistent estimates of the β's. Since $z_1, \dots, z_M$ are redundant in the equation above, they are uncorrelated with the error v (exogeneity). Since each x_j is uncorrelated with v as well, we can use 2SLS with the instruments $(1, x_1, \dots, x_{K-1}, z_1, \dots, z_M) = Z$. Given all zero-correlation assumptions, what we need for identification is that at least one

of the z_i appears in the reduced form of x_K. More formally, in the linear projection

$$x_K = z\Pi + r = \Pi_0 + \Pi_1 x_1 + \ldots + \Pi_{K-1} x_{K-1} + \Pi_K z_1 + \ldots + \Pi_{K+M-1} z_M + r$$

at least one of the $\Pi_K, \ldots, \Pi_{K+M-1}$ must be different from zero (relevance). To check for identification, conduct an F-Test (or a Wald test) with $H_0 : \Pi_K = \ldots = \Pi_{K+M-1} = 0$. We would have *over-identification* if $L > K + 1$. Here, $L = K + M$. We would have *exact identification* if If $M = 1$. Given $M > 1$, over-identification prevails. Tha means, $M-1$ over-identification restrictions are present. Their validity can be tested by means of a Sargan test.

(b) The family background variables have to be correlated with *education* and redundant in the log(*wage*) equation, once education (and some other factors) are controlled for. The idea is that family background may influence education but has no (direct) partial effect on log(*wage*).

4. (a) Derivation of AVar$(\hat\beta_{IV} - \beta)$.

$$\hat\beta_{IV} = (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta + U) = \beta + (Z'X)^{-1}Z'U$$

Rearrange the equality and multiply both sides by $\sqrt{N}$ to get

$$\sqrt{N}(\hat\beta_{IV} - \beta) = \sqrt{N}((Z'X)^{-1}Z'U) = \underbrace{\left(\frac{1}{N}Z'X\right)^{-1}}_{I.} \underbrace{N^{-\frac{1}{2}}Z'U}_{II.}$$

I. plim $\left((\frac{1}{N}(Z'X))^{-1}\right) = \mathrm{E}(z'x)^{-1} = A^{-1}$ due to WLLN and Slutsky's theorem

II. Note that $\mathrm{E}(z'u) = 0$. Then, due to CLT, $N^{-\frac{1}{2}}Z'U = N^{-\frac{1}{2}}\sum z_i'u_i \xrightarrow{d} N(0, B)$ with $B = \mathrm{AVar}(z_i'u_i) = \mathrm{E}(u^2z'z)$

Using Cramer's theorem we can combine I and II as following:

$$\sqrt{N}(\hat\beta_{IV} - \beta) \xrightarrow{d} N(0, A^{-1} \cdot BA'^{-1})$$

.

$$\Rightarrow \mathrm{AVar}(\sqrt{N}(\hat\beta_{IV} - \beta)) = \mathrm{E}(z'x)^{-1}\,\mathrm{E}(u^2z'z)\,\mathrm{E}(z'x)'^{-1} \overset{homosk.}{=} \sigma^2\,\mathrm{E}(z'x)^{-1}\,\mathrm{E}(z'z)\,\mathrm{E}(z'x)'^{-1}.$$

Now, back to the exercise. Denote $z = \begin{pmatrix}1 & z_1\end{pmatrix}$, $x = \begin{pmatrix}1 & x_1\end{pmatrix}$.

$$\mathrm{AVar}(\sqrt{N}(\hat\beta_{IV} - \beta)) = \sigma^2 \left(\mathrm{E}\begin{pmatrix}1 & x_1 \\ z_1 & x_1z_1\end{pmatrix}\right)^{-1} \mathrm{E}\begin{pmatrix}1 & z_1 \\ z_1 & z_1^2\end{pmatrix} \left(\mathrm{E}\begin{pmatrix}1 & x_1 \\ z_1 & x_1z_1\end{pmatrix}'\right)^{-1} =$$

$$= \sigma^2 \left(\mathrm{E}\begin{pmatrix}1 & x_1 \\ z_1 & x_1z_1\end{pmatrix}\right)^{-1} \mathrm{E}\begin{pmatrix}1 & z_1 \\ z_1 & z_1^2\end{pmatrix} \left(\mathrm{E}\begin{pmatrix}1 & z_1 \\ x_1 & x_1z_1\end{pmatrix}\right)^{-1} =$$

$$= \sigma^2 \underbrace{\frac{1}{1 \cdot \mathrm{E}(x_1 z_1) - \mathrm{E}(x_1)\,\mathrm{E}(z_1)}}_{\mathrm{Cov}(x_1,z_1)} \cdot \begin{pmatrix} \mathrm{E}(x_1 z_1) & -\mathrm{E}(x_1) \\ -\mathrm{E}(z_1) & 1 \end{pmatrix} \times$$

$$\times \begin{pmatrix} 1 & \mathrm{E}(z_1) \\ \mathrm{E}(z_1) & \mathrm{E}(z_1^2) \end{pmatrix} \cdot \underbrace{\frac{1}{1 \cdot \mathrm{E}(x_1 z_1) - \mathrm{E}(x_1)\,\mathrm{E}(z_1)}}_{\mathrm{Cov}(x_1,z_1)} \cdot \begin{pmatrix} \mathrm{E}(x_1 z_1) & -\mathrm{E}(z_1) \\ -\mathrm{E}(x_1) & 1 \end{pmatrix} =$$

$$= \sigma^2 \frac{1}{\mathrm{Cov}(x_1, z_1)^2} \cdot \begin{pmatrix} \mathrm{E}(x_1 z_1) - \mathrm{E}(x_1) \cdot \mathrm{E}(z_1) & \mathrm{E}(x_1 z_1) \cdot \mathrm{E}(z_1) - \mathrm{E}(x_1) \cdot \mathrm{E}(z_1^2) \\ 0 & \underbrace{\mathrm{E}(z_1^2) - \mathrm{E}(z_1)^2}_{\mathrm{Var}(z_1)} \end{pmatrix} \times$$

$$\times \begin{pmatrix} \mathrm{E}(x_1 z_1) & -\mathrm{E}(z_1) \\ -\mathrm{E}(x_1) & 1 \end{pmatrix}$$

Since we are only interested in $\mathrm{AVar}(\sqrt{N}(\hat{\beta}_{1,IV} - \beta_1))$, we only consider the lower right corner element of the final matrix:

$$\begin{aligned} \mathrm{AVar}(\sqrt{N}(\hat{\beta}_{1,IV} - \beta_1)) &= \frac{\sigma^2}{\mathrm{Cov}(x_1, z_1)^2} \cdot (0 \cdot (-\mathrm{E}(z_1)) + \mathrm{Var}(z_1) \cdot 1) = \\ = \frac{\sigma^2}{\mathrm{Cov}(x_1, z_1)^2} \cdot \mathrm{Var}(z_1) &= \frac{\sigma^2 \cdot Var(z_1)}{\mathrm{Corr}(x_1, z_1)^2 \cdot \mathrm{Var}(x_1) \cdot \mathrm{Var}(z_1)} = \\ = \frac{\sigma^2}{\rho_{x_1,z_1}^2 \cdot \mathrm{Var}(x_1)} & \end{aligned}$$

Compared to OLS,

$$\mathrm{AVar}(\sqrt{N}(\hat{\beta}_{OLS} - \beta)) = \sigma^2\, \mathrm{E}(x'x)^{-1} = \sigma^2 \begin{pmatrix} 1 & \mathrm{E}(x_1) \\ \mathrm{E}(x_1) & \mathrm{E}(x_1^2) \end{pmatrix}^{-1}$$

$$= \sigma^2 \underbrace{\frac{1}{\mathrm{E}(x_1^2) - \mathrm{E}(x_1)\,\mathrm{E}(x_1)}}_{\mathrm{Var}(x_1)} \cdot \begin{pmatrix} \mathrm{E}(x_1^2) & -\mathrm{E}(x_1) \\ -\mathrm{E}(x_1) & 1 \end{pmatrix}$$

$$\Rightarrow \mathrm{AVar}(\sqrt{N}(\hat{\beta}_{1,OLS} - \beta_1)) = \frac{\sigma^2}{\mathrm{Var}(x_1)} \qquad \text{(lower right corner)}$$

Since $0 < \rho_{x_1,z_1}^2 < 1$, it follows that $\mathrm{AVar}(\sqrt{N}(\hat{\beta}_{1,IV} - \beta_1)) \geq Avar(\sqrt{N}(\hat{\beta}_{1,OLS} - \beta_1))$. For β_1 OLS is more efficient than IV. But if OLS is inconsistent, efficiency is meaningless. Use OLS if it is consistent (be sure about that), but take IV if it's not.

(b) If σ^2 grows, the variance of the IV increases. If $\mathrm{Var}(x_1)$ grows, the variance of the IV decreases. Finally, if $\rho_{zx} \to 0$, the variance of the IV tends to infinity, meaning, the lower the correlation between the instrument and the endogenous regressor, the lower is the precision of IV estimator.

5. (a)

$$\begin{aligned}
x &= \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} \\
z &= \begin{pmatrix} z_1 & z_2 & x_3 & x_4 \end{pmatrix} \\
\Rightarrow L(x|z) = z\Pi &= \begin{pmatrix} z_1 & z_2 & x_3 & x_4 \end{pmatrix} \begin{pmatrix} \Pi_{11} & \Pi_{12} & 0 & 0 \\ \Pi_{21} & \Pi_{22} & 0 & 0 \\ \Pi_{31} & \Pi_{32} & 1 & 0 \\ \Pi_{41} & \Pi_{42} & 0 & 1 \end{pmatrix}
\end{aligned}$$

If for one of the endogenous x's (for example x_1) z_1 and z_2 do not have any explanatory power (not relevant), then in the linear projection the parameters Π_{11} and Π_{21} are zero. Then the first column of Π is a linear combination of the third and the last columns of Π and therefore $rank(\Pi) < K$. Hence, for each endogenous regressor at least one instrument must be relevant, otherwise the rank condition is violated and IV is not feasible.

(b) The condition from (a) is only necessary, but not sufficient for the rank condition, since there are other reasons why $rank(\Pi) < K$. Suppose z_1 is relevant for both x_1 and x_2, so Π_{11} and Π_{12} are not zero, but z_2 is not relevant for explaining x_1 and x_2, so $\Pi_{21} = \Pi_{22} = 0$. Then the second row of Π is a zero-vector and therefore $rank(\Pi) < K$. Here, we have only one relevant instrument.

(c) For $L_1 = K_1$, this condition means that Π is a diagonal matrix with nonzero diagonal elements. This implies that $rank(\Pi) = K$. For our example with $L_1 = K_1 = 2$ it means that

$$\Pi = \begin{pmatrix} \Pi_{11} & 0 & 0 & 0 \\ 0 & \Pi_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

6. (a) Given $\mathrm{E}(e_{i,t}|x_{i,T-1}, x_{i,T}) = 0$. It follows from that

i. $\mathrm{E}(e_{i,T}e_{i,T-1}) \stackrel{iid}{=} \mathrm{E}(e_{i,T}) \cdot \mathrm{E}(e_{i,T-1}) = 0 \cdot 0 = 0$

ii.

$$\begin{aligned}
& \mathrm{E}(e_{i,T}y_{i,T-1}) = \mathrm{E}(e_{i,T} \cdot (\beta_0 + \beta_1 x_{i,T-1} + \beta_2 y_{i,T-2} + e_{i,T-1} + \gamma e_{i,T-2}) \\
= \; & \beta_0 \underbrace{\mathrm{E}(e_{i,T})}_{=0} + \beta_1 \underbrace{\mathrm{E}(e_{i,T}x_{i,T-1})}_{=0(LIE)} + \beta_2 \underbrace{\mathrm{E}(e_{i,T}y_{i,T-2})}_{=0(LIE)} + \underbrace{\mathrm{E}(e_{i,T}e_{i,T-1})}_{=0} \\
+ \; & \gamma \underbrace{\mathrm{E}(e_{i,T}e_{i,T-2})}_{=0} = 0
\end{aligned}$$

(b) For consistency we need $\mathrm{E}((x_{i,T}, y_{i,T-1})u_{i,T}) = 0$. $\mathrm{E}(x_{i,T}u_{i,T}) = 0$ due to strict exogeneity. $\mathrm{E}(y_{i,T-1}u_{i,T}) \neq 0$ since $y_{i,T-1} = f(e_{i,T-1})$ and $u_{i,T} = g(e_{i,T-1})$. Thus, the OLS of the given regression will be inconsistent due to regressor $y_{i,T-1}$.

(c) Consider the following: $y_{i,T-1}$ depends on $x_{i,T-1}$ and $y_{i,T-2}$. Hence, we can use one (or both) of these variables for consistent IV-type estimation, since $\mathrm{E}(x_{i,T-1}u_{i,T}) = \mathrm{E}(y_{i,T-2}u_{i,T}) = 0$. Both instruments are exogenous and relevant.