Command	Explanation	Notes
`np.log()`	takes logarithm	numpy as np
`np.log()`	takes logarithm	numpy as np
`df.dropna(subset=['var'])`	drops rows in `df` where `var` is NaN	pandas
`lmiv.IV2SLS.from_formula()`	2SLS regression	linearmodels.iv as lmiv
`ivReg.wu_hausman()`	test of exogeneity on `ivReg`	linearmodels.iv as lmiv
`ivReg.durbin()`	test of exogeneity on `ivReg`	linearmodels.iv as lmiv
`ivReg.wooldridge_regression`	test of exogeneity on `ivReg`	linearmodels.iv as lmiv
`ivRegOID.sargan`	overidentification test on `ivRegOID`	linearmodels.iv as lmiv
`ivRegOID.wooldridge_overid`	overidentification test on `ivRegOID`	linearmodels.iv as lmiv

Two-Stage Least Squares

```
1  ### Run the regression with instrument distance for educ
2  eqn = "np.log(wage) ~ 1 + [np.log(educ) ~ np.log(distance)]" \
3          "+ np.log(exper) + np.log(feduc) + np.log(meduc) + urban"
4  ivReg = lmiv.IV2SLS.from_formula(eqn, wages).fit()
5  print(ivReg)
6
7  ### Test to see if educ is really endogenous
8  ### Large p-value? Conclude educ is exogenous, IV unnecessary
9  print(ivReg.wu_hausman())
10 print(ivReg.durbin())
11 print(ivReg.wooldridge_regression)
12
13 ### Test to see if instruments are exogenous
14 ### Large p-value? Not overidentified, some instruments invalid
15 eqnOID = "np.log(wage) ~ 1 + [np.log(educ) ~ np.log(distance)" \
16           "+ np.log(sibs+.001)] + np.log(exper) + np.log(feduc)" \
17           "+ np.log(meduc) + urban"
18 ivRegOID = lmiv.IV2SLS.from_formula(eqn, wages).fit()
19 print(ivRegOID.sargan)
20 print(ivRegOID.wooldridge_overid)
```