# Suggested solution to written paper III (ECON 4135)

Auction theory predicts that the price of the object to be auctioned increases with the number of bidders in the auction. More bidders means more competition for the object and thus higher prices. Using data from the Dutch flower auction this exercise aims at investigating the *causal* effect of number of bidders on the price of Begonia. The auctioneers randomly place the flowers in the order in which they will be auctioned. The auction starts at 06h30 sharp and continues until all flowers have been auctioned. All the transactions we observe in the data happen during the first 10 minutes of the auction.

## Problem 1

The price of the object that is being auctioned at transaction $i$ is explained by the number of bidders present at transaction $i$, $i = 1, ..., N$ , thus our parameter of interest is $\beta_1$ in Equation 1. We also control for the type of the Begonia as well as weekdays.

$$price_i = \beta_0 + \beta_1 bidders_i + \beta_2 type + \beta_3 tues + \beta_4 wednes + \beta_5 thurs + u_i \quad (1)$$

$u$ is an error term. We start out with estimating this model by OLS. Type the following in stata (SEE PROBLEM 1 IN THE TEXT FILE CALLED AUCTIONLOG):

- *unab controls: type2 tues wed thurs*
- *reg price bidders 'controls'*

From the stata window we see that an increase in the number of bidders is associated with an increase in the price ($\beta_1^{ols} = 4.173$). Moreover, type 2 is more expensive than type1. The price is higher on Tuesdays than on Mondays, whereas it is lower on both Wednesdays and Thursday than on Mondays. Note also that total number of transactions are 79.

## Problem 2

The error term $u$ embodies all factors other than our observed explanatory variables (bidders, type and weekday) that determine the price of object $i$, such as the size. Controlling for type (colour) only captures that some colours may be more popular than other colours. It can still be the case that the size varies between begonias sharing the same colour. Suppose an object has a high level of $u$ as a result of its large (unobserved) size. This increases the price since $price_i = \beta_0 + \beta_1 bidders_i + ... + u_i$, but it may also lead to more bidders since larger begonias may attract more bidders than smaller begonias. Since our data contain no information on the size of the begonia this is captured by the error term $u$. Thus the error term may then be positively correlated with number of bidders.

What are the consequences of this correlation between $bidders$ and $u$? Number of bidders present at transaction $i$ has now two effects on the price of object $i$ that is being auctioned at transaction $i$. There is both a direct effect via $\beta_1 bidders_i$ and an indirect effect via $u$ affecting $bidders_i$ which in turn affect $price_i$. The goal of regression is to estimate only the first effect, yielding an estimate of $\beta_1$. The OLS estimate from *problem 1* will instead combine these two effects. In our case this means that $\hat{\beta}_1^{OLS} > \beta_1$ since both effects are positive.

Using calculus we have $price_i = \beta_0 + \beta_i bidders_i + u_i(bidders_i)$ with total derivative $\frac{dprice_i}{dbidders_i} = \beta_1 + \frac{du_i}{dbidders_i}$ (for simplicity reasons we have now dropped all other control variables). The data give information on $\frac{dprice_i}{dbidders_i}$, so when applying OLS we obtain the total effect $\beta_1 + \frac{du_i}{dbidders_i}$ rather than $\beta_1$ alone. The OLS estimator for $\beta_1$ is therefore biased ($E[\hat{\beta}_1] \neq \beta_1 + \sum_{i=1}^{n} E[u_i]E[\frac{(bidders_i - \overline{bidders}_n)}{\sum_{i=1}^{n}(bidders_i - \overline{bidders}_n)^2}]$) and inconsistent ($\hat{\beta}_1 \overset{n\to\infty}{\longrightarrow} \beta_1 + \frac{E[(bidders_i - \overline{bidders}_n)u_i]}{var(bidders_i)} \neq \beta_1$) unless there is no association between $bidders$ and $u$. As already argued the correlation between $bidders$ and the error term is positive. Thus we expect $\beta_1$ to be upward biased.

## Problem 3

It is plausible to assume that $bidders$ is an endogenous explanatory variable, meaning that changes in $bidders$ are associated not only with changes in the $price$, but also with changes in $u$ ($corr(bidders, u) \neq 0$) as argued in **problem 2**). This implies that we are in need for a method to generate only exogenous variation in $bidders$ in order to estimate the causal effect of $bidders$ on $price$. An obvious way is through randomized experiments, but for most economics application such experiments are expensive of even infeasible.

A crude experiment of treatment approach is still possible using observational data, provided there exists an instrument $z$ that has the property that changes in $z$ are associated with changes in $bidders$, but do not lead to change in $price$ aside from the indirect route via $bidders$. The more direct path of $z$ being a regressor in the model for $price$ is ruled out (the so-called "exclusion restriction"). The exclusion restriction is however not testable.

More formally, a variable $z$ is called an instrument or instrumental variable for our endogenous regressor $bidders$ if (1) $z$ is uncorrelated with the error term $u$ and (2) $z$ is correlated with the regressor $bidders$. These two properties are necessary for consistency. There is also a third property which is a strengthening of the second to ensure good finite-sample performance of the IV estimator. The third property (3) is: $z$ should be strongly correlated, rather than weakly correlated with $bidders$.

## Problem 4

Type the following in stata (SEE PROBLEM 4 IN THE TEXT FILE CALLED AUCTIONLOG):

- *label var time "time (seconds after 06h30)"*
- *label var bidders "observations"*
- twoway (scatter bidders time, msymbol(oh i) c(i l) legend(pos(7) ring(0) col(1) ) ) (lfit bidders time), ytitle("# bidders") scheme(s1mono)

This gives us the following scatterplot shown in Figure 1. The figure indicates that number of bidders present at transaction $i$ is an increasing function of $time$. Thus $time$ is positively correlated with number of bidders.

Since time is measured in seconds after 06h30 we can imagine that this positive relationship between $bidders$ and $time$ arises because of some small unexpected delays on the way to the auction; such as a little bit more traffic than usual, waiting for traffic lights to turn green again, etc. Also recall that this auction takes place in Holland where it is common for both bikers, cars and pedestrians to wait for bridges to close when there is a lot of boat traffic on the canals.

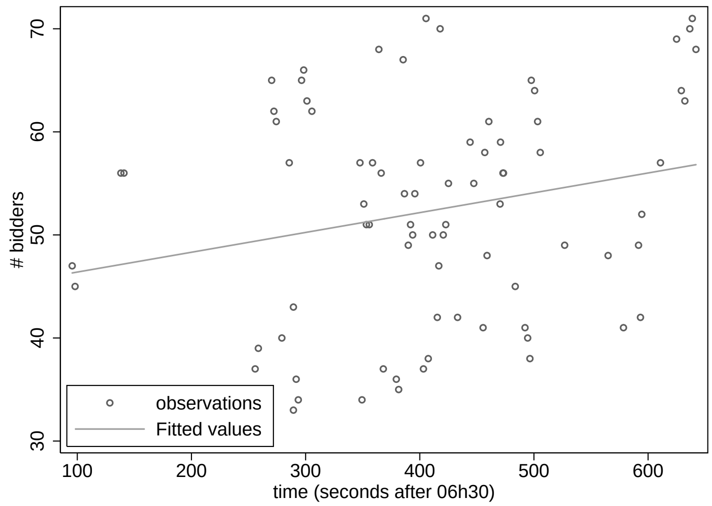


Figure 1: The relation between number of bidders and time

We are in need for an instrument for our endogenous regressor $bidders$. Could we instrument $bidders$ with $time$? As already outlined in the figure, $time$ satisfies condition 2. Additionally, it also most likely satisfies condition 1 ($time$ does not directly determine the transaction price of object $i$).

## Problem 5

Use OLS to estimate the first stage given by Equation

$$bidders_i = \pi_0 + \pi_1 time + \pi_2 type2 + \pi_3 tues + \pi_4 wed + \pi_5 thurs + v \quad (2)$$

*Type the following in stata* (SEE PROBLEM 5A IN THE TEXT FILE CALLED AUCTIONLOG):

- *reg bidders time 'controls'*

The point estimate of time equals 0.024 ($\hat{\pi}_1 = 0.0242$) and indicates that a one second increase in time is associated with 0.024 more bidders. t-test statistic equals 2.32.

In order to say something about whether $time$ a weak or a strong instrument we continue with calculating the F-test statistic for joint significance. The easiest way to do that is to type *"test $time$" in stata* (SEE PROBLEM 5B IN THE TEXT FILE CALLED AUCTIONLOG). This gives us a F-test statistic which equals 5.36. Staiger and Stock (Econometrica, 1997) use as a rule of thumb that the F-test statistic for joint significance of the instruments in the first stage should exceed 10. This suggest that $time$ is a weak instrument involving that $time$ does not explain much variation in the endogenous regressor $bidders$ ($cov(time, bidders)$is small).

YOU GUYS SHOULD ALSO CALCULATE THE F-TEST STATISTIC "BY HAND" JUST TO CHECK THAT YOU GET THE SAME RESULT.

## Problem 6

We now substitute $bidders$ in Equation (1) with the expression for $bidders$ in the first stage given in Equation (2) to obtain the following expression for $price$ given by Equation (3).

$$price_i = \gamma+\beta_1\pi_1 time+(\beta_2+\beta_1\pi_2)type2+(\beta_3+\beta_1\pi_3)tue+(\beta_4+\beta_1\pi_4)wed+(\beta_5+\beta_1\pi_5)thu+e \tag{3}$$

where $\gamma = (\beta_0 + \beta_1\pi_0)$ and $e = u + \beta_1 v$

The term $\beta_1\pi_1$ capture both the effect of $bidders$ on $price$ ($\beta_1$) and the the effect of $time$ on number of $bidders$ ($\pi_1$). In what follows we will show how to separate $\beta_1$ from $\pi_1$.

## Problem 7

Problem 7 involves calculating the Two Stage Least Squares (TSLS) estimator for $\beta_1$. We start out with estimating Equation (3) with OLS. *Type the following in stata (*SEE PROBLEM 7A IN THE TEXT FILE CALLED AUCTIONLOG):

- *reg price time 'controls'*

From the stata window we see that $\hat{\beta}_1\hat{\pi}_1 = 0.0584$ or $\hat{\beta}_1 = \frac{0.0584}{\hat{\pi}_1}$ . Recall from first stage in **problem 5** that $\hat{\pi}_1 = 0.0242$. The second stage estimator for $\hat{\beta}_1$ is then given by $\hat{\beta}_{1,TSLS} = \frac{0.0584}{0.0242} = 2.41$

Instead of calculating $\hat{\beta}_{1,TSLS}$ "by hand" we can also make use of the **ivreg** command in stata which does the calculations for us. *Type the following in stata* (SEE PROBLEM 7B IN THE TEXT FILE CALLED AUCTIONLOG):

- *ivreg price (bidders=time) 'controls'*

From the stata window we see that the point estimate for $bidders$ equals 2.41. Note however that this effect lacks precision (i.e. the effect is not significant anymore). The corresponding standard error equals 3.569. Be aware of that loss in precision increases with weaker instruments (hint: look at the formula for the variance of the TSLS estimator).

The TSLS estimator gets its name from the result that it can be obtained by two consecutive OLS regressions. In our case: OLS regression of $bidders$ on $time$ to get $\hat{bidders}$ (Equation (2)) followed by OLS regression of $price$ on $\hat{bidders}$ (Equation (3)) which gives $\hat{\beta}_{1,TSLS}$.

## Problem 8

When ruling out spurious variation (i.e. the indirect effect via $u$) the point estimate substantially decreases. This indicate that the $\beta_1^{ols} = 4.173$ is clearly overestimated as suggested by the reasoning in **problem 2**. However, recall that the instrumental variable estimator is also not unbiased. In case of weak instruments, the bias can be severe even in relatively large samples. As already detected $time$ is most likely a weak instrument.

The last question asks us to say something about the possible bias of the TSLS estimator relative to the bias in the OLS estimator using the formula in Appendix 12.5 in Stock&Watson.

Our obtained F-test statistic form **problem 5** equals 5.36. This implies that the bias of the TSLS estimator relative to the bias in the OLS estimator equals $\frac{1}{F-1} = \frac{1}{5.36-1} = \frac{1}{4.36}$. This suggests that we only partially correct for the bias in $\beta_1$ caused by endogeneity in $bidders$ when instrumenting $bidders$ with $time$ measured in seconds after 06h30. However, this should be interpreted with care as the formula refers to the case with may instruments.

## Useful to know

If you think that Word and Scientific Workplace is crap, check out www.lyx.org (it's free).

**ftp://ftp.lyx.org/pub/lyx/bin/1.5.2/LyX-1.5.2-1-Installer.exe**