Endogeneity Testing

Tomas Morley

13 maart 2017

Wooldridge (2010) proposes a test for endogeneity following the control function approach. Essentially, the procedure involves regressing the endogenous variable against a suitable instrument in a first stage reduced model, and then incorporating the resiudals from the resuced model in the estimation of the full structural model. The presence of endogeneity is confirmed by the significance of the control function (reduced form residual), as determined by a t-test under the null hypothesis of no exogeneity.

In our case, we explore the relationship between maize yield and nitrogen application at the field level. The possibility that nitrogen is endogenous has been explored by .... A suitable instrument for the endogenous variable is the relative price of nitrogen to maize. Both prices are assumed to be exogenous to the biophysical relationship connection maize yields to nitrogen applications. Both prices are given by the market and and while they may affect the decision as to whether or not to produce maize, or how much to produce, they are unlikely to be related to the maize yield response function which is driven by biophysical constraints, crop practices and input availability.

A complication is that in our sample of smallholder maize farmers, many report using no fertilizer and this may influence the choice of reduced form first stage model. Two possible options are a reduced form OLS model or a tobit model. The results for each and the average partial effects (APEs) are displayed in Table 1.

Table 1: First stage results

	OLS	SE	Tobit	SE	APE	BSE
(Intercept)	3.635**	0.923	8.339**	2.518	3.421	1.018
relprice	-0.062**	0.005	-0.18**	0.015	-0.074	0.007
sex	0	0.085	0.102	0.218	0.042	0.087
age	0.036**	0.013	0.1**	0.035	0.041	0.015
agesq	0**	0	-0.001**	0	0	0
logslope	-0.211**	0.045	-0.412**	0.116	-0.169	0.051
elevationsqt	0.002	0.01	-0.045	0.029	-0.019	0.012
crop_count2	0.266**	0.065	0.779**	0.17	0.32	0.069
impr	2.145**	0.081	3.757**	0.191	1.542	0.072
ed_any	-0.105	0.073	-0.109	0.188	-0.045	0.077
SOC2	-0.002	0.007	0	0.019	0	0.009
phdum55_2_70	0.125	0.071	-0.116	0.185	-0.048	0.087
dist_market	-0.005**	0.001	-0.014**	0.002	-0.006	0.001
GGD	-0.249**	0.063	-0.769**	0.172	-0.316	0.067
cost2large_town	-0.003**	0.001	-0.012**	0.003	-0.005	0.001
R-squared	0.466	NA	0.463	NA	NA	NA

The results are broadly similar for those coefficients that are significantly different from zero. Comparing the R-squared from the OLS model to the pseudo R-squared from the Tobit model indicate that there is very little difference between the two in terms of fit. However, it should be noted that the OLS model explicitley maximimizes the R-squared, whereas the tobit model is fit by ML. Nonethless, we find little reason to prefer the one model over the other.

Table 2: Second stage results

	TL-LM	SE	BSE	TL-Tob	SE	BSE
(Intercept)	3.756**	0.318	0.323	3.426**	0.293	0.304
logN	0.126*	0.054	0.055	0.201**	0.043	0.044
loglab	0.282**	0.022	0.027	0.281**	0.022	0.027
logNloglab	-0.028**	0.009	0.009	-0.027**	0.009	0.009
logarea	-0.207**	0.023	0.024	-0.206**	0.023	0.025
crop_count2	0.569**	0.046	0.046	0.556**	0.046	0.045
logslope	-0.261**	0.032	0.032	-0.242**	0.031	0.031
impr	0.633**	0.11	0.109	0.435**	0.098	0.098
v	0.087*	0.042	0.043	-0.004	0.022	0.022
SOC2	0.024**	0.005	0.005	0.027**	0.005	0.005
GGD	0.126**	0.032	0.032	0.154**	0.03	0.03
phdum55_2_70	0.121*	0.047	0.049	0.123**	0.047	0.047
dumoxen	0.115*	0.047	0.047	0.113*	0.047	0.047

The full structural model was then estimated by ordinary least squares incorporatoing the control function. The results are reported in table 2. We see that the reported results in table 2 are very close with the exception of the coefficients on the `logN` term and the residual term. Wooldridge (2010) described a method for constructing a generalized residual from a probit model. Here we use a generalized residual from a tobit model constructed as. There are reasons to suspect that the tobit coefficient in the is controlling for the fact that a large number of plots have no nitrogen applied. In this case the response to nitrogen, once this has been accounted for, is higher than estimated in the ols reduced form where. In short the endogeneity that we observe may in fact be more related to a selection issue or data censoring/truncation

$$\hat{v}_{tob} = d_1 \frac{-\phi(-x_i'\beta)}{\Phi(-x_i'\beta)} + d_2(\theta y_i - x_i'\beta)$$