Appendix B: test for endogeneity of website provision to magazine demand

In this Appendix I test for endogeneity of website provision to magazine demand using a framework I derive in Kaiser (2002). This method is based on Heckman–type correction terms (Heckman 1979). I test for two different forms of website endogeneity: (i) endogeneity of *having* a website and (ii) endogeneity of *launching* a website at time t.

Model framework

Let $Website_{jt}$ be a dummy variable that is coded 1 if magazine j launches a website at time t (read analogously henceforth: runs a website at time t). Magazine j is assumed to run a website at time t if a latent variable $Website^*_{jt}$ is larger than zero:

$$Website_{jt} \;=\; \begin{cases} 1 & \text{if} \quad Website^*_{jt} \;=\; \boldsymbol{Z_{jt}}\,\boldsymbol{d} \;+\; v_{jt} \;>\; 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $\boldsymbol{d}$ is a vector of parameters (relating the vector of explanatory variables $\boldsymbol{Z_{jt}}$ to $Website^*_{jt}$).

The natural logarithm of relative magazine market shares is now defined by:

$$\begin{aligned} ln(s_{jt}/s_{0t}) \;&=\; \boldsymbol{x_{jt}\beta} + \alpha_{jt}p_{jt} + \sigma ln(\bar{s}_{j|gt}) + \boldsymbol{\kappa w_{jt}} + \boldsymbol{\delta a_{\neq jt}} + \tau_t + \; cWebsite_{jt} + \xi_{jt} \quad (2) \\ &=\; \boldsymbol{\phi w_{jt}} + \; cWebsite_{jt} + \xi_{jt}. \quad (3) \end{aligned}$$

The disturbance terms v_{jt} and ξ_{jt} are bivariate i.i.d. normal distributed with mean zero and variance–covariance $\boldsymbol{\Sigma}$. Note that

$$E[\xi_{jt} \mid \; -(v_{jt} + \boldsymbol{Z_{jt}d}) > 0] \;=\; -\;\rho\;\sigma_\xi\;\frac{\phi(-\frac{\boldsymbol{Z_{jt}d}}{\sigma_v})}{\Phi(\frac{-\boldsymbol{Z_{jt}d}}{\sigma_v})} \;=\; -\,\rho\;\sigma_\xi\;\lambda_{jt}, \tag{4}$$

where σ_ξ and σ_v are the standard errors of the disturbance terms ξ_{jt} and v_{jt}, respectively, and that

$$E[\xi_{jt}| - (v_{jt} + \boldsymbol{Z_{jt}d}) < 0] \;=\; \rho\;\sigma_\xi\;\frac{\phi(\frac{\boldsymbol{Z_{jt}d}}{\sigma_v})}{\Phi(\frac{\boldsymbol{Z_{jt}d}}{\sigma_v})} \;=\; \rho\;\sigma_\xi\;\mu_{jt}. \tag{5}$$

The relative market share equation accounting for endogeneity of website launching is

$$ln(s_{jt}/s_{0t}) = \boldsymbol{\phi w_{jt}} + cWebsite_{jt} \;+\; \rho\sigma_\xi\mu_{jt}Website_{jt} \;-\; \rho\sigma_\xi\lambda_{jt}(1 - Website_{jt}) \;+\; \xi_{jt}. \tag{6}$$

Equation (6) can be estimated in a two–step procedure. First, estimate $\boldsymbol{d}/\sigma_v$ by a probit model and calculate $\hat{\lambda}_{jt}$ and $\hat{\mu}_{jt}$. Second, estimate equation (6) by OLS. This procedure leads to consistent parameter estimates. The related variance–covariance matrix is consistently estimated only if the Heckman–type correction

terms are insignificantly different from zero as they are in the present case. The full information maximum likelihood estimator is derived in Kaiser (2000).

First stage Probit estimation

The first stage reduced form specification corresponding to Equation (1) consists of variables that are assumed to influence the website decision but not magazine demand (these are the so–called 'exclusion restrictions') and variables that affect magazine demand, vector $\boldsymbol{w_{jt}}$. My choice of exclusion restrictions follows Kaiser (2003) where I study the timing of website launching. I argue that the share of readers that regularly uses the internet is related to having a website but not to magazine demand so the share of magazine j's readers that regularly uses the internet and its square are my exclusion restrictions. My specification of magazine demand is a restricted version of the demand model I estimate in the main text. This needs to be so since a richer specification of magazine demand leads to a richer specification in the website choice decision which in turn lead to an overfitting of the probit model. Hence, my set of variables in the magazine demand equation consists of cover prices, the natural logarithms of within group market share, the natural logarithms of content pages and its square, advertising share and its square, a linear and a quadratic time trend and the content concentration index. The cover prices and within group market shares used in the estimations come from 2SLS regression to account for their endogeneity. Instruments used in the 2SLS regressions are the same as in the main text.

Table B 1 presents first stage reduced form probit estimates for the probability to (i) maintain a website at time t and (ii) launch a website at time t. Note that the number of observations in second specification drops since all observations for all periods larger than t related to magazines that launch a website at t naturally have to be removed (to give an example: magazine B launches a website at $t = 2$, all observations related to magazine B and to periods $t > 2$ are discarded in the estimations).

The tests for joint significance of the parameters indicate a good for of the probit models. The high significance of the exclusion restrictions (they are also jointly highly significant) indicate that they are highly correlated with the website choice. They are well chosen exclusion restrictions if they are also uncorrelated with the ξ's (which remains to be shown).

Magazine demand estimation

Table B 2 presents second stage structural form OLS estimates for relative magazine demand. The corresponding variance–covariance matrix corrects for heteroscedasticity and autocorrelation. The correction terms are neither jointly nor separately significant at the usual significance levels in any equation. There hence is no evidence for endogeneity of either website launching or website maintaining.

Check for orthogonality of the exclusion restrictions

Lastly, I check for validity of my exclusion restrictions. In order for them to be valid they (i) have to have a significance effect on the website decisions (which they clearly have) and (ii) have to be uncorrelated with the error terms from

the second stage level equation. The latter can easily be tested by running and OLS regression of the exclusion restrictions on the error terms from the magazine demand equation. Estimation results are shown in Table B 3. Orthogonality of the exclusion restrictions cannot be rejected at the usual significance levels as indicated by the test for joint significance.

Table 1: First stage reduced form probit estimates

	Probability of website maintaining		Probability of website launching	
	Coeff.	p–value	Coeff.	p–value
Cover price	-0.0052942	0.978	0.0037897	0.991
log(within group market share)	0.0638981	0.824	-0.05586	0.914
ln(ed. pages)	-12.9178	0.012	-1.370576	0.887
ln(ed. pages)^2	1.184487	0.005	0.1989357	0.801
Share ad. pages	35.00019	0.000	40.01963	0.006
Share ad. pages^2	-30.18605	0.006	-48.87943	0.008
Time	-0.0284638	0.188	-0.0400972	0.206
Time^2	-0.0001315	0.000	-0.0000742	0.162
Content conc.	-11.22622	0.005	-3.810638	0.545
Share readers online	0.2415742	0.000	0.2446624	0.001
Share readers online^2	-0.0029823	0.000	-0.0052454	0.002
Constant	26.43297	0.098	-8.661765	0.765
Miscellaneous				
	Test-stat.	**p–value**	**Test-stat.**	**p–value**
Joint significance	128.7378	0.000	37.3832	0.0001
Pseudo R^2	0.6175		0.3000	
# of obs.	860		705	

Table 2: OLS structural form second stage magazine demand estimates

	Endogeneity of website maintaining		Endogeneity of website launching	
	Coeff.	***p*–value**	**Coeff.**	***p*–value**
Cover price	-0.4918	0.000	-0.5077	0.000
log(within group market share)	0.4750	0.000	0.5018	0.000
ln(ed. pages)	2.6231	0.131	2.6601	0.186
ln(ed. pages)^2	-0.1728	0.221	-0.1739	0.290
Share ad. pages	-0.6202	0.313	0.0503	0.945
Share ad. pages^2	2.8256	0.003	1.5227	0.194
Time	-0.0180	0.000	-0.0198	0.000
Time^2	0.0000	0.000	0.0000	0.006
Content conc.	-1.2079	0.017	-1.1638	0.024
$Website$	-0.0006	0.994	0.1696	0.782
$\rho\sigma_\xi Website$	0.0869	0.165	0.3390	0.180
$\rho\sigma_\xi(1 - Website)$	-0.0573	0.396	-0.0563	0.867
Constant	-11.2953	0.034	-11.4249	0.063
Miscellaneous				
Joint significance	126.5732	0.000	91.2298	0.000
Joint significance correction terms	1.8662	0.156	0.9211	0.399
Adj. R^2	0.5459		0.5327	
# of obs.	860		705	

Table 3: OLS regression of exclusion restriction on market demand equation residual

	Endogeneity of website maintaining		Endogeneity of website launching	
	Coeff.	***p*–value**	**Coeff.**	***p*–value**
Share readers online	-0.0044	0.395	-0.0114	0.198
Share readers online^2	0.0001	0.413	0.0003	0.292
Constant	0.0199	0.531	0.0427	0.278
Miscellaneous				
Joint significance	0.3638	0.696	0.9324	0.393
Adj. R^2	0.0008		0.0027	
# of obs.	860		705	

Additional references

Heckman, J.J. (1979). Sample selection bias as a specification error. Econometrica 47, 153–161.

Kaiser, U. (2003). Why do magazines go online?, University of Southern Denmark at Odense mimeo, download: http://www.ulrichkaiser.com/papers/when.html.

Kaiser, U. (2002). An empirical test of models explaining research expenditures and research cooperation. International Journal of Industrial Organization 20(6), 747–774.

Kaiser, U. (2000). Research cooperation and research expenditures with endogenous absorptive capacity: theory and microeconometric evidence for German services. ZEW discussion paper 00–25; download: ftp://ftp.zew.de/pub/zew-docs/dp/dp0025.pdf.