

Analysis of Seabright study on demand for Sky's pay TV services

Annex 7 to pay TV phase three document

Publication date:

26 June 2009

Comments on the study: "The effect of DTT availability on household's willingness to subscribe to Sky's pay TV services" by Paul Seabright

Andrew Chesher¹, April 28th 2009, revised June 22nd 2009²

1 Introduction

The Seabright study aims to understand how households' demand for Sky TV services is affected by availability of TV services via other platforms.

The study employs data on household choices aggregated to the postcode district level. The study provides two sets of estimates.

- 1. Ordinary least squares (OLS) estimates of regression functions. These suggest a statistically significant *negative* association between DTT availability and Sky subscription rates. The estimated effect is *larger* for subscriptions to Sky's *basic* package than for subscriptions to *pre-mium* packages.
- 2. Two stage least squares (2SLS) estimates of the coefficients in a structural model in which DTT availability is an endogenous variable and postcode district distance from the nearest DTT transmitter is an instrumental variable. For *basic* package subscriptions the 2SLS estimates of the effect of DTT availability are 13% *larger* in magnitude than the OLS estimates. The difference is very substantial (78%) for Sky *premium* subscriptions (2SLS: -0.059 OLS: -0.033).

The Seabright study concludes that the two stage least squares estimates are more reliable and it is these that produce the headline results of the study. The comments below follow upon re-analysis of the data.

2 Comments

2.1 Summary

1. A variety of alternative specifications were considered in a re-analysis of the Seabright study data.

¹Andrew Chesher is Professor of Economics at University College London and Director of the Centre for Microdata Methods and Practice at the IFS. The opinions expressed here are his alone.

 $^{^{2}}$ These comments were originally compiled on May 28th 2009. Some questions were raised with the authors of the report and discussed in a video-conference on May 27th 2009. Written replies to the questions were received on June 16th and these comments were then revised as indicated in the text.

- 2. The magnitudes of the estimated coefficients on DTT availability in equations for Sky *basic* package subscriptions are found to be generally robust to these changes in specification.
- 3. There is less robustness in the magnitude of the estimated coefficients on DTT availability in equations for Sky *premium* packages estimated using the 2SLS procedure.
 - (a) In a new analysis in which there is control for location in 115 groups of post code districts the 2SLS premium package estimate is closer to zero: -.040 than the value of -.059 reported in the study.
 - (b) In a new analysis in which the data are not weighted by number of delivery points the 2SLS premium package estimate is closer to zero: -.029 than the value of -.059 reported in the study.
 - (c) OLS estimates and estimates for basic packages are not very sensitive to weighting and inclusion of district controls.
- 4. Consideration of variation in the quality of DTT signal suggests that the proposed instrumental variable, *distance to DTT transmitter*, may have a direct role in the demand equation for Sky packages by capturing the effect of DTT signal quality on the effectiveness of DTT as a substitute for Sky services.
 - (a) A new analysis in which distance to DTT transmitter is not used as an instrumental variable but instead is allowed to attenuate the effect of DTT availability on Sky TV package demand gives an estimate of the coefficient on (good quality) DTT availability in the demand equation for Sky premium services of -.035 (compare with -.059).
- 5. It is possible that the cable access explanatory variable is endogenous. This could result in bias in any of the estimated coefficients and bias in estimates of the relative magnitudes of coefficients on DTT and cable availability which are given some prominence in the Seabright study report.
- 6. The estimated effects of DTT availability on demand for Sky premium services are relatively sensitive to model reformulation when compared with the estimated effects of DTT on demand for basic Sky TV service. Plausible model reformulations lead to estimates of the effect of DTT availability on the demand for Sky premium services of the order of one half of the value headlined in the study.³

³This point was added in the revision to the comments.

Details are given in the sections which follow.

2.2 Within district analysis

In the Seabright study data set there is a variable *district* which defines 115 postcode district *groups*. Group membership is determined by the leading non-numeric characters of the postcode district code. For example the St. Albans (AL) district has 10 postcode districts, $AL1,..., AL10.^4$

Within these geographic district groupings we can expect a degree of homogeneity in some of the factors that it was not possible to control for in the Seabright study including, importantly, factors related to entertainment and leisure alternatives to Sky and other TV services.

When indicator variables identifying the 115 districts are included in the model there is some control for these various factors, the resulting estimates being determined entirely by *within* district variation in the variables included in the analysis.

Table 1 show the estimates of the coefficient on DTT availability when district indicator variables are included in the fitted equations and, for comparison, the values given in Tables 5, 6, 7 and 8 of the Seabright report where the district indicator variables are not included.

The differences that arise on including district indicators are small except for the case of premium subscriptions estimated by the two stage least squares method. In that case the estimated coefficient on DTT availability is -0.40compared with -0.59 (district indicator variables excluded). The accuracy of estimation is little changed on including district indicators.

Table 2 shows the estimated coefficients (2SLS) on DTT availability obtained when district indicators are included and separate equations are estimated for different types of premium subscription. For sports only subscriptions and for sports and movie subscriptions the DTT coefficients are smaller once district indicators are included.

2.3 Weighting

The analysis reported on the Seabright study uses data weighted by number of delivery points per postcode district. An unweighted analysis was conducted. This gives similar results *except* in the equation for premium subscriptions when DTT availability is modelled as endogenous and estimates are obtained using the 2SLS procedure.

⁴All London postcodes are gathered in one group.

Weighting is expected to improve the efficiency of estimation but *not* to lead to substantial changes in estimated coefficients. The change on removing weighting suggests some fragility in the 2SLS estimates of the premium package model. The estimated coefficients with and without weighting are shown in Table $3.^5$ The 2SLS coefficient for premium packages is smaller without weighting than with weighting.

2.4 Endogeneity and DTT signal quality

In the Seabright study access to DTT is treated as an *endogenous* variable. The endogeneity of the DTT access variable can be debated, but the debate cannot be resolved using data alone. In order to know if it is endogenous one has to maintain the existence of a valid instrumental variable and the validity of the chosen variable cannot be tested.

In the Seabright study the variable *distance to DTT transmitter* (abbreviated to DTRANS here) is maintained to be a valid instrumental variable. It will not be valid if it has a direct role in the determination of demand for Sky packages. It is worth considering if this could be the case.

DTRANS may have a direct role if it is associated with the *quality* of the DTT signal that households receive. Where signal quality is poor, which may be the case for households far from a transmitter, DTT may be a less effective substitute for Sky TV services. In this case the size of the DTT coefficient effectively depends on distance to transmitter.

This is investigated by including an *interaction* term DTT*DTRANS in the model. The argument just advanced leads one to expect a *positive* coefficient on this interaction term. Table 4 shows OLS⁶ estimates of coefficients on DTT, negative as in the Seabright model, and DTRANS*DTT, positive as suggested above.

The magnitudes of the coefficients on DTT *alone* show the effect of DTT availability on demand for Sky packages when DTRANS is close to zero which, under the interpretation advanced here, corresponds to good quality DTT.

For Premium packages the estimate is considerably *smaller* (-.045 excluding district, -.035 including district) then the 2SLS estimate, -.059, given in the Seabright study.

It has been argued that if there is variation in the quality of DTT availability not captured by the binary DTT availability then DTRANS could still be a suitable instrumental variable because it will be correlated with

⁵Conventional and heteroskedasticity robust standard errors are reported. None of the tables here report standard errors corrected for spatial correlations.

⁶2SLS estimates cannot be calculated now DTRANS appears in the equation.

the quality sensitive measure of DTT availability, say QDTT, that actually affects demand for Sky TV services. However for DTRANS to be a valid instrumental variable it would *also* have to be *uncorrelated* with the *difference* between the observed binary DTT variable and the quality sensitive measure QDTT. Since DTT is binary it is difficult to find a model in which that can happen while maintaining a correlation between DTRANS and QDTT.⁷

2.5 Endogeneity and cable TV access

The availability of cable may be endogenous because cable infrastructure may have been installed more comprehensively in areas where demand for pay TV services is thought to be relatively high.

If cable availability is endogenous then all the estimates reported in the Seabright study are potentially biased. Some care should be taken in comparing the sizes of estimated coefficients on DTT and cable availability.

Without an instrumental variable for cable availability it cannot be known whether this is a serious problem or not. None was proposed in the Seabright study where the possibility of endogeneity of the cable availability variable was not considered.

There is no obvious instrumental variable available in the data set employed in the study. Population density probably affects the economic decision to invest in cable infrastructure and might be considered as a candidate instrumental variable. However it is likely to be associated *directly* with demand for Sky TV services because alternatives to Pay TV are more likely to be available in higher population density areas.

It is not possible to determine the direction of any bias resulting from the endogeneity of cable availability.⁸

⁷This paragraph was added in the revision of these comments.

⁸This paragraph was added in the revision of these comments.

Basic or	OLS or	As reported		Including district	
Premium	2SLS	coeff	std error	coeff	std error
Basic	OLS	072	.0019	067	.0018
Basic	2SLS	082	.0056	089	.0048
Premium	OLS	033	.0027	030	.0026
Premium	2SLS	059	.0079	040	.0068

Table 1: 2008: Estimated coefficients on DTT availability allowing for district indicator variables

Table 2: Estimated 2SLS coefficients on DTT availability for elements ofpremium subscriptions allowing for district dummy variables

Premium element	Year	No district indicators		Including district indicators		
		coefficient	std error	coefficient	std error	
Sports only	2008	026	.0024	017	.0019	
Movies only	2008	011	.0013	011	.0012	
Sports and movies	2008	018	.0055	009	.0047	

Table 3: 2008: Estimated coefficients on DTT availability and estimated standard errors with and without weighting. s.e. denotes estimated standard error

Package	Method	Unweighted			Weighted		
		coeff	s.e.	robust s.e.	coeff	s.e.	robust s.e.
Basic	OLS	086	.0020	.0027	072	.0019	.0025
Basic	2SLS	095	.0048	.0067	082	.0056	.0067
Premium	OLS	029	.0027	.0030	033	.0027	.0028
Premium	2SLS	029	.0065	.0078	059	.0079	.0084

Table 4: 2008: Estimated coefficients on DTT availability and its interaction with distance to DTT transmitter

Package	District	DT	Т	DTRANS*DTT		
	indicators?	coefficient	std error	coefficient	std error	
Basic	No	074	.0020	.023	.0088	
Premium	No	045	.0051	.097	.0272	
Basic	Yes	074	.0031	.074	.0173	
Premium	Yes	035	.0029	.058	.0145	