**Appendix 2:** **Robustness check for the estimation of the user uptake of internet banking**

Considering that (branch density, measured by the natural log of the number of branches per square km) could vary significant within a country depending on the size of urban areas as opposed to rural areas, ideally we would like to use data on the number of branches per square km in urban and rural areas separately. However, such information is not available. To mitigate potential bias caused by the omission of such variables, we constructed an additional instrument for (branch density) measured by the natural log of the size of metropolitan land area () in the first stage regressions. We then repeated the analysis in the 2nd stage and reported the results in the following Table A3. The results are very consistent with what was shown in Table 2 in section 6.

*Table A3*: Estimation Results for the Diffusion of IB, without and without additional instrument for branch density

|  |  |  |
| --- | --- | --- |
|  | C measured by C5 | C measured by HHI |
| Dependent variable: user uptake of IB | Spec 2 with additional instrument for B (branch density) | Spec 2 without additional instrument for B (branch density) | Spec 2 with additional instrument for B (branch density) | Spec 2 without additional instrument for B (branch density) |
| *C (concentration)* | 0.041\*\*\*(0.0063) | 0.048\*\*\*(0.0079) | 8.452\*\*\*(1.3877) | 8.375\*\*\*(2.7339) |
| *R (regionalisation)* | 1.418\*\*\*(0.3753) | 1.723\*\*\*(0.3625) | 0.724\*\*(0.3204) | 0.887\*\*(0.3646) |
| *B (branch density)* | 0.064(0.0860) | 0.073(0.1001) | 0.237\*\*\*(0.0867) | 0.217\*\*(0.1023) |
| *E (Education,%)* | -0.024\*\*(0.0110) | -0.038\*\*\*(0.0113) | -0.030\*\*\*(0.0098) | -0.036\*\*\*(0.0117) |
| *I (Median income (log)* | 0.527\*\*\*(0.1733) | 0.528\*\*(0.1936) | 0.628\*\*\*(0.1722) | 0.571\*\*(0.2099) |
| *Gi (Gini income , 0-100)* | 0.105\*\*\*(0.0224) | 0.082\*\*\*(0.0255) | 0.090\*\*\*(0.0220) | 0.078\*\*\*(0.0246) |
| *U (Unemployment, %)* | 0.023\*\*\*(0.0059) | 0.025\*\*\*(0.0062) | 0.024\*\*\*(0.0057) | 0.025\*\*\*(0.0061) |
| *constant* | -68.299\*\*(28.0282) | -67.814\*\*(34.3305) | 3.970(28.4343) | -12.594(31.7151) |
| *t* | 0.354\*\*\*(0.0403) | 0.407\*\*\*(0.0530) | 0.288\*\*\*(0.0355) | 0.297\*\*\*(0.0468) |
| *C\*t* | -0.007\*\*(0.0004) | -0.001\*\*\*(0.0004) | -0.095(0.0883) | -0.209(0.1690) |
| *R\*t* | -0.059\*\*\*(0.0219) | -0.083\*\*\*(0.0211) | -0.033\*(0.0192) | -0.045\*\*(0.0220) |
| *B\*t* | -0.029\*\*\*(0.0072) | -0.027\*\*\*(0.0087) | -0.033\*\*\*(0.0071) | -0.031\*\*\*(0.0082) |
| *Gi \*t* | -0.010\*\*\*(0.0015) | -0.009\*\*\*(0.0017) | -0.009\*\*\*(0.0015) | -0.008\*\*\*(0.0016) |
|  | -0.053\*\*\*(0.0114) | -0.068\*\*\*(0.0141) | -11.115\*\*\*(2.9895) | -14.005\*\*\*(4.1566) |
|  | 0.002\*\*(0.0007) | 0.004\*\*\*(0.0009) | 0.258(0.2269) | 0.782\*\*\*(0.2567) |
|  | 0.351(0.2525) | -0.024(0.2844) | 0.189(0.2813) | -0.117(0.3028) |
|  | 0.033\* (0.0192) | 0.024(0.0212) | 0.035\* (0.0209) | 0.028 (0.0224) |
| *Adjusted R2* | 0.93 | 0.92 | 0.93 | 0.92 |
| *No. of Obs.* | 381 | 381 | 381 | 381 |

 *Figure A1.* Predicted internet banking uptake: the impact of concentration given each level of regionalisation from Table A3



*Figure A2.* Predicted internet banking uptake: the impact of regionalisation given each level of concentration from Table A3

****

**Appendix 3: Robustness check for the estimation of endogenous market concentration**

1. Estimation using 2SLS:

Recall in section 5.2 we described the control function approach used to control for a potential bias arising from the endogeneity of *D* which is a binary variable. As indicated in Wooldridge (2015), this approach exploits the binary nature of endogenous variable but it is generally inconsistent if the probit model for the binary endogenous variable is mis-specified. This is in contrast with the robustness of the usual 2SLS estimator which does not use any distributional assumptions in the reduced form. If the probit model for the binary endogenous variable is correctly specified, then the control function procedure and 2SLS should give estimates that differ only due to sampling error. For this reason, we re-estimated equation (3) in section 3.2 using the 2SLS procedure and the results are shown in Table 4 (section 6) and Figure A3 below. The 2SLS results are very consistent with our results in Table 4 and Figure 5 in section 6 confirming the robustness of our results using the control function approach.

*Figure A3.* Predicted market concentration varying with regionalisation and the uptake of internet banking (using estimates from 2SLS in Table 4)



1. Estimation by adding additional instrument for B (branch density)

*Table A4:* Estimation results for national concentration in banking with and without additional instrument for branch density

|  |  |  |
| --- | --- | --- |
| *Dependent variable*:  | *With additional instrument for B* | *Without additional instrument for B* |
| *Estimated coefficients**CF* | *Estimated coefficients**2SLS* | *Estimated coefficients**CF* | *Estimated coefficients**2SLS* |
| *t* | 0.089\*\*\*(0.0098) | 0.083\*\*\*(0.0105) | 0.090\*\*\*(0.0082) | 0.081\*\*\*(0.0113) |
|  | 61.512\*\*\*(6.6646) | 68.607\*\*\*(7.0343) | 49.706\*\*\*(7.4198) | 58.581\*\*\*(7.6082) |
|  | -5.858\*\*\*(1.6776) | -5.025\*\*\*(1.5396) | -11.674\*\*\*(2.1856) | -8.038\*\*\*(1.7908) |
|  | 0.124\*\*(0.0653) | 0.136\*\*\*(0.0506) | 0.286\*\*\*(0.0670) | 0.259\*\*\*(0.0688) |
|  | 68.643\*\*\*(22.7321) | 58.656\*\*\*(20.7393) | 149.774\*\*\*(30.1833) | 98.528\*\*\*(24.1566) |
| *D* | -2.395\*\*\*(0.6092) | -1.739\*\*(0.7523) | -2.058\*\*\*(0.5838) | -1.368\*(0.8381) |
| *D\*t* | -0.092\*\*\*(0.0144) | -0.084\*\*\*(0.0138) | -0.090\*\*\*(0.0101) | -0.077\*\*\*(0.0144) |
|  | 36.481\*\*\*(6.7413) | 26.644\*\*\*(8.9247) | 31.362\*\*\*(6.4704) | 20.863\*\*\*(9.8031) |
|  | 11.053\*\*\*(2.2500) | 9.442\*\*\*(2.0043) | 13.119\*\*\*(2.5263) | 7.686\*\*\*(2.4002) |
|  | -0.134\*\*(0.0661) | -0.160\*\*\*(0.0527) | -0.159\*\*\*(0.0615) | -0.154\*\*(0.0700) |
|  | -146.062\*\*\*(32.4660) | -126.042\*\*\*(24.2483) | -179.961\*\*\*(36.7303) | -101.104\*\*\*(33.7451) |
| A (crisis) | 8.492\*\*\*(2.4667) | 8.346\*\*\*(2.5152) | 9.572\*\*\*(2.0891) | 8.355\*\*\*(2.7722) |
| Constant | 5.768(3.7189) | 5.385(3.9655) | 3.706(3.1020) | 3.884(4.0957) |
|  | -13.997(17.5735) |  | -8.587(16.3207) |  |
|  | -0.493\*\*\*(0.1548) |  | -0.981\*\*(0.0883) |  |
|  | 0.070(0.1351) |  | -0.128(0.0937) |  |
|  | -33.715(40.2708) |  | -199.214\*\*\*(35.9062) |  |
| Adjusted *R2* | 0.69 | 0.69 | 0.77 | 0.65 |
| No of Obs. | 430 | 430 | 430 | 430 |

*Figure A4*. Predicted market concentration varying with different levels of regionalisation and maturity of internet banking with additional instrument for branch density: control function approach



 *Figure A5*. Predicted market concentration varying with different levels of regionalisation and maturity of internet banking with additional instrument for branch density: 2SLS approach



1. Sensitivity analysis by excluding Luxembourg from our sample

We have already noted that Luxembourg might be an outlier in the concentration equation given its small population size and large total bank assets as measured by the ECB. If we exclude Luxembourg from the sample for estimating concentration, the results are reported in Table A5 and Figure A6 and A7. We have conducted F tests which confirm that the two specifications (with or without including Luxembourg as a potential outlier) in the table below are statistically different. However, the qualitative pattern described in section 6 has not changed.

*Table A5:* Estimation results for national concentration in banking: excluding Luxembourg

|  |  |  |
| --- | --- | --- |
|  | *With Luxembourg* | *Without Luxembourg* |
| *Dependent variable*:  | *Estimated coefficients**CF* | *Estimated coefficients**2SLS* | *Estimated coefficients**CF* | *Estimated coefficients**2SLS* |
| *t* | 0.090\*\*\* (0.0082) | 0.081\*\*\* (0.0113) | 0.096\*\*\* (0.0078) | 0.113\*\*(0.0108) |
|  | 49.706\*\*\*(7.4198) | 58.581\*\*\*(7.6082) | 65.785\*\*\*(8.1423) | 70.333\*\*\*(7.2065) |
|  | -11.674\*\*\*(2.1856) | -8.038\*\*\*(1.7908) | -14.052\*\*\*(1.7073) | -14.066\*\*\*(1.4633) |
|  | 0.286\*\*\* (0.0670) | 0.259\*\*\* (0.0688) | 0.344\*\*\* (0.0712) | 0.368\*\*\* (0.0617) |
|  | 149.774\*\*\*(30.1833) | 98.528\*\*\*(24.1566) | 174.710\*\*\*(23.9657) | 168.517\*\*\*(19.1204) |
| *D* | -2.058\*\*\*(0.5838) | -1.368\*(0.8381) | -1.784\*\*\*(0.5335) | -2.007\*\*\*(0.7054) |
| *D\*t* | -0.090\*\*\*(0.0101) | -0.077\*\*\*(0.0144) | -0.084\*\*\*(0.0102) | -0.089\*\*\*(0.0134) |
|  | 31.362\*\*\*(6.4704) | 20.863\*\*\*(9.8031) | 29.377\*\*\*(6.6548) | 33.711\*\*\*(8.0315) |
|  | 13.119\*\*\*(2.5263) | 7.686\*\*\*(2.4002) | 12.272\*\*\*(2.0332) | 8.422\*\*\*(1.9204) |
|  | -0.159\*\*\*(0.0615) | -0.154\*\*(0.0700) | -0.078(0.0664) | -0.004(0.0635) |
|  | -179.961\*\*\*(36.7303) | -101.104\*\*\*(33.7451) | -164.155\*\*\*(29.5933) | -105.115\*\*\*(26.9514) |
| A (crisis) | 9.572\*\*\* (2.0891) | 8.355\*\*\*(2.7722) | 8.639\*\*\* (2.4853) | 6.994\*\*\*(2.8100) |
| Constant | 3.706(3.1020) | 3.884(4.0957) | -26.816\*\*\*(3.8007) | -45.120\*\*\*(4.6722) |
|  | -8.587(16.3207) |  | -64.828\*\*\*(15.5927) |  |
|  | -0.981\*\*\*(0.0883) |  | -0.897\*\*\*(0.0842) |  |
|  | -0.128(0.0937) |  | 0.099(0.1058) |  |
|  | -199.214\*\*\*\*(35.9062) |  | -152.253\*\*\*(41.0728) |  |
| Adjusted *R2* | 0.77 | 0.65 | 0.79 | 0.72 |
| No of Obs. | 430 | 430 | 413 | 413 |

*Notes:* D is constructed using the threshold of IB at 25%.

*Figure A6.* Predicted market concentration varying with different levels of regionalisation and maturity of internet banking without Luxembourg: Control Function approach from Table A5



*Figure A7.* Predicted market concentration varying with different levels of regionalisation and maturity of internet banking without Luxembourg : 2SLS approach from Table A5



1. Sensitivity analysis by using a different threshold value of *IB* for *D*:

We explored different threshold values for construction of the dummy variable *IB* used in the concentration estimations (equation (3)). Table A6 and Figures A8-11 report the results using a different cut-off point: D = 1 if IB > 30% and 0 otherwise. It is clear that the results are similar those reported in section 6 (Table 4 and Figure 5) , Table A5 and Figures A6 and A7 except that the control function for *D* becomes significant.

*Table A6*: Estimation results for national concentration in banking using a different threshold value of *IB* (30%) for *D*

|  |  |  |
| --- | --- | --- |
|  | *With Luxembourg* | *Without Luxembourg* |
| *Dependent variable*:  | *Estimated coefficients**CF* | *Estimated coefficients**2SLS* | *Estimated coefficients**CF* | *Estimated coefficients**2SLS* |
| *t* | 0.088\*\*\*(0.0074) | 0.079\*\*\*(0.0097) | 0.096\*\*\*(0.0065) | 0.111\*\*\*(0.0088) |
|  | 47.640\*\*\*(8.4732) | 59.993\*\*\*(7.8319) | 63.717\*\*\*(8.7066) | 68.997\*\*\*(7.3570) |
|  | -11.905\*\*\*(1.8549) | -7.536\*\*\*(1.8672) | -14.337\*\*\*(1.6527) | -14.918\*\*\*(1.5900) |
|  | 0.334\*\*\*(0.0638) | 0.310\*\*\*(0.0752) | 0.392\*\*\*(0.0732) | 0.411\*\*\*(0.0651) |
|  | 151.265\*\*\*(24.9843) | 90.786\*\*\*(24.9911) | 177.250\*\*\*(21.8143) | 178.498\*\*\*(20.7751) |
| *D* | -2.274\*\*\*(0.6167) | -1.433\*(0.8823) | -2.020\*\*\*(0.5393) | -2.395\*\*\*(0.7185) |
| *D\*t* | -0.088\*\*\*(0.0089) | -0.070\*\*\*(0.0138) | -0.082\*\*\*(0.0105) | -0.081\*\*\*(0.0133) |
|  | 31.313\*\*\*(7.3442) | 17.401\*\*\*(10.0909) | 29.311\*\*\*(6.2517) | 35.045\*\*\*(7.4821) |
|  | 13.981\*\*\*(1.9327) | 7.340\*\*\*(2.8618) | 13.119\*\*\*(1.8877) | 9.374\*\*\*(2.2808) |
|  | -0.198\*\*\*(0.0725) | -0.206\*\*\*(0.0797) | -0.119\*(0.0684) | -0.019(0.0691) |
|  | -191.799\*\*\*(28.7854) | -96.623\*\*\*(41.1404) | -175.481\*\*\*(27.2524) | -117.358\*\*\*(32.8126) |
| A (crisis) | 9.956\*\*\*\*(2.4685) | 8.151\*\*\*(2.6364) | 8.795\*\*\*(2.0525) | 6.637\*\*\*(2.3683) |
| Constant | 3.416(3.6059) | 3.790(4.1917) | -27.014\*\*\*(4.3024) | -48.342\*\*\*(4.8895) |
|  | 0.323(17.4236) |  | -54.732\*\*\*(17.1667) |  |
|  | -0.997\*\*\*(0.0809) |  | -0.915\*\*\*(0.0952) |  |
|  | -0.266\*\*\*(0.0901) |  | 0.211\*\*(0.0991) |  |
|  | -211.504\*\*\*(39.3120) |  | -162.920\*\*\*(31.5484) |  |
| Adjusted *R2* | 0.78 | 0.65 | 0.80 | 0.72 |
| No of Obs. | 430 | 430 | 413 | 413 |

*Figure A8.* Predicted market concentration varying with different levels of regionalisation and maturity of internet banking with Luxembourg: Control Function approach from Table A6



*Figure A9.* Predicted market concentration varying with different levels of regionalisation and maturity of internet banking with Luxembourg: 2SLS approach from Table A6



*Figure A10.* Predicted market concentration varying with different levels of regionalisation and maturity of internet banking without Luxembourg: Control Function approach from Table A6



*Figure A11*. Predicted market concentration varying with different levels of regionalisation and maturity of internet banking without Luxembourg: 2SLS approach from Table A6





**Appendix 4: Sample Size Discussion**

A panel of national data over a period relevant to the diffusion of an innovative product must inevitably be of limited sample size compared with, say, consumer level data. However, consumer (or firm) level data, even if was hypothetically available, would not be able to address our research questions. Our data sources and the nature of a diffusion process (which is fully completed by many countries within our sample period) mean that the size of the panel cannot be extended usefully in either the international or temporal dimensions. The question remains of whether our sample is inappropriately small. Calculating power for observed data provides no information beyond what the ordinary p-value (i.e. the probability of type I error) already provides (Hoenig and Heisey, 2001; Schulz and Grimes, 2005).

Nevertheless, to gain a greater feel for the possible limitations of our sample size, we have calculated the sample sizes necessary for 90% power in the following hypothetical cases:

To detect the size of the effect we have estimated based on our observed sample, what would be the sample size required to achieve the power of 90% at 1% significance level? More specifically, we calculated (using Stata) sample size required to detect the difference (at 1% significance level) in R-sq with and without our main covariates with the power of 90% for linear regressions. Our calculations suggest that to detect the differences in R-sq that are estimated in the current study, with a power of 90%, the sample size required for model 1 is 60 and for model 2 is 39.

We then calculated the minimum detectable effect given the number of main/control covariates and our sample size. Set power at 90% and significance level at 1% for linear regressions, the calculation shows that an average sample of our size could detect a minimum effect of the size indicated by r-sq difference of 0.009 for model 1 and 0.042 for model 2.

*Table A7:* Sample size analysis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Our sample size  | Number of main covariates  | Number of full set of covariates  | R-sq with main covariates | R-sq without main covariates | Detected effect of main covariates indicate by the difference in R-sq | Required sample size to detect the size of the effect of main covariates estimated by our sample (power at 90% and significance at 1%) | Minimum detectable difference in R-sq given our sample size (power at 90% and significance at 1%) |
| Model 1 | 381 | 9 | 23 | 0.9195 | 0.8638 | 0.056 | 60 | 0.009 |
| Model 2 | 430 | 10 | 22 | 0.7710 | 0.3381 | 0.433 | 39 | 0.042 |

The above calculations suggest our significance tests are unlikely to be underpowered. However, these calculations don’t completely answer the question “would the average hypothetical dataset of this size contain enough evidence to reject *H0*?” because a different sample may detect the effect of main covariates at a different magnitude. Given the limited data sample we have, our findings should be interpreted with the appropriate caution.