Online Appendices to Appropriability Mechanisms, Innovation and Productivity: Evidence from the UK

Bronwyn H. Hall[1]

Vania Sena[2]

March 2015

Appendix A: Construction of the dataset

For this study we have constructed an *ad hoc* dataset by using the following five components available at the SecureLab, UK Data Service. These are all linked by the unique reporting unit number:

Business Structure Database (BSD): the dataset is derived from the Inter Departmental Business Register (IDBR) and provides longitudinal business demography information for the population of businesses in the UK. We use information on a company's industrial classification (SIC 92) as well as incorporation and market exit dates from the BSD to be able to define the age of the firm.[3]

Annual Respondents Database (ARD2): the ARD2 is constructed from the microdata collected in the Annual Business Inquiry (ABI) conducted by the ONS. The stratified survey sample is drawn from the IDBR.[4] The ARD covers both the production (including manufacturing) and the non-production sector (services). However the time series dimension varies across the two sectors: while for the production sector it is possible to have information available up to 1980 (and early 70s for some industries), the data for the services sector is available only after 1997. The information is assembled from the replies to the Census forms: as this is a mandatory requirement for UK-based business, the response rates to the ARD are rather high and this makes it highly representative of the underlying population. Each establishment has got a

[1] Corresponding author: University of California at Berkeley and University of Maastricht; NBER, IFS London, and NIESR. bhhall@berkeley.edu

[2] Essex Business School, University of Essex. vsena@essex.ac.uk

[3] The definition of market exit is problematic. It is not possible to identify whether a firm has ceased trading or if it has merely undergone a change in structure that leads to its original reference number becoming extinct.

[4] The stratification sample weights are as follows: businesses with (a) <10 employees 0.25, (b) 10-99 employees 0.5, (c) 100-249 employees all or ≥ 0.5 depending on industry, and (d) >250 employees all. Moreover, if a firm with <10 employees is sampled once, it is not sampled again for at least three years.

unique reference number that does not change over time and so allows us to build up a panel dataset. The ARD is a stratified random sample where sampling probabilities are higher for large establishments: indeed for establishments with more than 250 employees, the sampling probability is equal to one. The ARD contains all the basic information (namely the inputs and output variables) needed to estimate the production function. Output is measured by the deflated added value. Employment is measured by the total number of employees. As for capital, it is well known that the ARD does not contain information on capital stock. However, stock of capital has been constructed at the ONS by using the perpetual inventory method.

UK Community Innovation Survey (CIS) 3, 4, and 5: the CIS is a stratified sample of firms with more than 10 employees drawn from the IDBR. The CIS contains detailed information on firms' self-reported innovative activities. This covers firms' innovation activities over a three-year window targeting firms with more than ten employees. The CIS is a survey carried out by national statistical agencies in all 25 EU member states under the coordination of Eurostat. The sampling frame for the UK CIS was developed from the Interdepartmental Business Register (IDBR) with the survey being conducted by post. Firms are asked whether they have produced any innovation in the reference period (i.e. the three years before the survey starts) and if so, what type of innovation they have introduced. In turn innovation can be of three types: product innovation, process innovation and wider (or organisational) innovation. Unsurprisingly, firms can be simultaneously produce two type of innovations (or even three types) and this allows us to construct our dependent variable as the total number of innovations produced by a firm over the period 2005-07. This variable can then vary between 0 (as firms may not produce any innovation in the reference period and therefore are recorded as non-innovators) and 3 (if firms produced a product, a process and a wider innovation at the same time). The CIS provides information on what external sources of information a firm uses and whether it collaborates with other organisations to develop innovation. In addition, the Survey contains information on R&D expenditure, the proportion of the workforce with a degree in engineering or a science subject and whether or not the plant is part of a group. We use three surveys: CIS 3 which covers the period 1998-2000, CIS 4 which covers 2002-2004, and CIS 5 which covers 2004-2006. The sample frames differ for the three CIS waves both in terms of size and industry coverage. For CIS 3, the sample frame consists of 19,625 enterprises with responses from 8,172 enterprises (42% response rate); CIS 3 covers both production (manufacturing, mining, electricity, gas and water, construction) and services sectors whereas the retail sector has been excluded. CIS 4 has the largest sample size out of the three CIS waves with a sample frame of 28,355 enterprises and responses from 16,446 enterprises (58% response rate); it also includes the following sectors: sale, maintenance & repair of motor vehicles (SIC 50); Retail Trade (SIC 52); and Hotels & restaurants (SIC 55). CIS 5 was answered by 14,872 firms which correspond to a response rate of 53% (Robson and Haigh, 2008). It covers the same industries as CIS 4 with the addition of SIC 921 (motion picture and video activities) and 922 (radio and television activities).

One problem that arises in combining these datasets is the identification of the relevant sampling unit. The ARD2 is apparently sampled at the reporting unit level (which may not coincide with the firm in the case of multi-unit firms), where it is possible that a reporting unit may belong to a larger enterprise, although most of the enterprises consist of a single reporting unit. In principle, the UK CIS is sampled at the enterprise level. Thus for multi-establishment

enterprises, there is some ambiguity about whether we have the full complement of data from the ARD2. Fortunately this problem will affect relatively few firms in our sample.

Table A1: Choosing the sample

	Observations	Firms
Total CIS observations	**68,112**	**46,638**
Not matched to ARD	20,005	
ARD-CIS match	**48,107**	
Drop missing industries, primary inds, inds 80-98	26,092	
Drop non-profits, government, missing legal status	519	
Unable to construct capital stock	5,040	
Potential CIS sample	**16,456**	**11,421**
Missing employment on CIS	1,049	
Large estimation sample	**15,407**	**10,844**
Missing capital, turnover, or materials	3,761	
Trim ratios for production function at 1%	796	
Estimation sample	**10,850**	**7,255**
CIS 6 and 7 sample	**3,706**	**3,068**
CIS 3,4,5 sample	**7,144**	**5,553**

Note: the ratios trimmed are those for sales/capital stock, sales/employment, sales/materials, capital stock/employment, and R&D and innovation spending intensity. Observations in the one per cent tails of the distribution were excluded.

Table A2: Sectoral breakdown 1998-2010

		Number of observations			*Share*	
SIC	*Description*	*Total*	*Large*	*SMEs*	*Large*	*SMEs*
23-25, ex						
244,245	Mfg of chemical, rubber, plastic, oil	445	267	178	60.0%	40.0%
30, 32	Mfg of computers & electronic inst	156	80	76	51.3%	48.7%
31	Mfg of elec equipment	192	92	100	47.9%	52.1%
28	Mfg of fabricated metal goods	318	117	201	36.8%	63.2%
15,16	Mfg of food, beverage, and tobacco	713	500	213	70.1%	29.9%
33	Mfg of medical & scientific inst	171	92	79	53.8%	46.2%
17-19, 36	Mfg of misc low-tech goods	502	237	265	47.2%	52.8%
34	Mfg of motor vehicales	294	176	118	59.9%	40.1%
29	Mfg of non-elec machinery	413	239	174	57.9%	42.1%
35	Mfg of other transport equipment	188	98	90	52.1%	47.9%
244	Mfg of pharmaceuticals	57	47	10	82.5%	17.5%
26, 27	Mfg of primary metals	301	153	148	50.8%	49.2%
22	Mfg of printed goods	285	202	83	70.9%	29.1%
245	Mfg of soap & toiletries	61	46	15	75.4%	24.6%
20, 21	Mfg of wood & furniture	243	97	146	39.9%	60.1%
	Total manufacturing	**4339**	**2443**	**1896**	**56.3%**	**43.7%**
45	Construction	803	413	390	51.4%	48.6%
64	Post, telephone, and telegraph	148	85	63	57.4%	42.6%
37,40,41,90	Utility services	134	68	66	50.7%	49.3%
50-52	Wholesale & retail trade	2077	1372	705	66.1%	33.9%
	Total utilites & trade	**3162**	**1938**	**1224**	**61.3%**	**38.7%**
72	Computer services	206	125	81	60.7%	39.3%
65-70	Financial, insurance, real estate	286	181	105	63.3%	36.7%
55	Hotel & restaurant services	499	401	98	80.4%	19.6%
71	Leasing services	163	72	91	44.2%	55.8%
52, 74	Other business services	1394	932	462	66.9%	33.1%
73	R&D services	76	49	27	64.5%	35.5%
60-63	Transportation services	725	455	270	62.8%	37.2%
	Total services	**3349**	**2215**	**1134**	**66.1%**	**33.9%**
	Total	**10850**	**6596**	**4254**	**60.8%**	**39.2%**

Table A3: Estimation sample - CIS 3, 4, 5 matched to BSD (2000-2006)

	All observations		*R&D firms*		*Inn. spend firms*	
Variable	*Median*	*IQ range*	*Median*	*IQ range*	*Median*	*IQ range*
Observations	7,144		2,162		4,414	
Number of employees	305	627.5	353.5	763	315	607
Turnover*	25000	71327	35811	89629	27385	73283
Value added*	8951	21916	11729	25951	9748	22167
Capital*	5002	15407	7572	19522	5661	16315
Purchased goods & services*	12014	39102	18794	42940	13490	39168
Output-employee ratio*	85.87	93.85	94.82	87.81	89.52	93.68
Output-capital ratio	5.22	10.49	4.69	8.30	5.02	10.21
Output-materials ratio	1.80	2.25	1.75	1.87	1.76	2.19
Capital per employee*	17.17	31.97	20.72	34.95	18.48	32.92
R&D -turnover ratio	0.0000	0.0004	0.0025	0.0092	0.0000	0.0024
Innovation spend -turnover ratio	0.0019	0.0155	0.0142	0.0380	0.0104	0.0290
R&D per employee*	0.000	0.036	0.246	0.942	0.000	0.234
Innovation spend per employee*	0.158	1.421	1.333	3.728	0.915	2.663
Age in 2011 in years	28	18	29	17	28	18
Importance of formal IP in the 3-digit sector	0.35	0.33	0.44	0.28	0.38	0.35
Importance of informal IP in the 3-digit sector	0.40	0.36	0.59	0.31	0.48	0.36
Perception of market risk in the 3-digit sector	0.37	0.34	0.40	0.24	0.38	0.29
Perception of financial constraints in the 3-d sector	0.34	0.25	0.34	0.22	0.34	0.25
Importance of regulation & standards in the 3-digit sector	0.33	0.32	0.39	0.25	0.36	0.31
Importance of environmental, H&S regs. in the 3-digit sector	0.29	0.29	0.37	0.30	0.31	0.32

* Units are 1000s of GBP.

Table A4: Dummy variable means

	All observations	*R&D firms*	*Innov. spend firms*
Number of observations	7,144	2,162	4,414
		Share of firms	
formal IP of med or high importance	35.1%	62.3%	46.7%
informal IP of med or high importance	44.8%	77.3%	60.7%
foreign ownership	25.0%	29.4%	26.3%
exports	48.1%	70.6%	57.0%
market risk high	43.2%	55.6%	50.0%
financial constraints	39.5%	46.2%	44.5%
innovate to improve range	39.2%	64.8%	52.0%
innovate for new markets	40.8%	66.6%	54.1%
innovate for quality improvement	47.3%	73.7%	62.4%
innovate to increase flexibility	37.1%	57.6%	49.0%
innovate to increase capacity	34.2%	52.3%	45.3%
innovate to reduce unit cost	37.2%	59.9%	49.3%
innovate to meet regulations or standards	34.3%	50.7%	43.7%
innovate for environment or health&safety	31.0%	49.0%	40.3%
collaborates	19.2%	37.0%	26.4%
within group important info source	58.4%	91.4%	78.3%
suppliers important info source	52.7%	75.2%	70.1%
customers important info source	55.9%	83.4%	73.7%
competitors important info source	41.9%	62.0%	54.8%
universities important info source	10.3%	20.8%	14.1%
product imitator only	11.0%	19.6%	15.8%
product innovator	33.4%	63.0%	47.1%
new-to-market product innovator	22.4%	43.4%	31.3%
process imitator only	19.2%	32.1%	26.8%
process innovator	26.4%	46.8%	37.2%
new-to-market process innovator	7.2%	14.7%	10.4%

Table A5: Average composition of innovation expenditure

	All	*SME*	*Large*	*Manu-facturing*	*Services & other*
Acquisition of mach. & comp. hardware/software	45.1%	48.0%	43.0%	43.2%	47.0%
Internal R&D spending	18.6%	17.7%	19.2%	25.1%	12.0%
Marketing expense	13.5%	11.8%	14.9%	10.6%	16.5%
Training expense	9.5%	10.2%	8.9%	5.4%	13.4%
Design expense	6.4%	5.9%	6.8%	8.8%	4.2%
External R&D spending	3.7%	3.5%	3.9%	4.2%	3.2%
Acquisition of external knowledge	3.2%	2.9%	3.4%	2.6%	3.7%
Observations with nonzero spending	4,414	1,876	2,538	2,199	2,215
Share with nonzero spending	61.8%	57.1%	65.8%	71.1%	54.7%

The shares shown are for firms that have some form of innovation spending reported.

Appendix B: Additional estimates

In this appendix we present some additional estimates of our model that use innovation spending in place of R&D spending as an explanatory variable.

Tables B1 and B2 show relatively few differences from Tables 2 and 3 (using R&D spending). That is, using innovation spending instead of R&D as a predictor of the preference for formal and informal IP and innovation makes little difference to the coefficient estimates. The largest differences statistically are the increase in the within group information source coefficients and the decrease in the suppliers information source coefficient. This may reflect the changes in these coefficients in the innovation spending equations, and raises some concern about the interpretation of these coefficients. That is, stronger coefficients in the innovation spending model seem to be reflected in strengthened coefficients of the opposite sign in the IP-innovation probability model. Recall that the latter model includes the value of R&D or innovation spending that is predicted based partly on these coefficients.

Table B1. Multivariate Probit estimates of IP choice and product innovation

7,144 observations on 5,684 firms; Log likelihood = -9,005.1

	Formal IP methods			*Informal IP methods*			*Product Innovator or imitator*		
	Coeff.	*Std. err*		*Coeff.*	*Std. err*		*Coeff.*	*Std. err*	
Log (predicted IS per employee)	1.026	0.065	***	0.945	0.063	***	0.453	0.066	***
Log (n of employees)	0.399	0.020	***	0.329	0.019	***	0.165	0.020	***
Log (firm age in 2011)	0.028	0.053		0.042	0.053		-0.092	0.056	
D (collaborates)	-0.182	0.054	***	-0.093	0.056		0.395	0.055	***
Firm perception of market risk	0.331	0.043	***	0.373	0.044	***	0.175	0.044	***
Firm perception of fin. Constraints	0.092	0.042	*	0.270	0.044	***	0.007	0.044	
Firm - impt. of reg & standards	0.151	0.050	**	0.131	0.052	*	-0.114	0.053	*
Firm - impt. of env, H&S regs	0.030	0.051		0.142	0.054	**	-0.031	0.054	
D (innov to improve range)							0.704	0.051	***
D (innov for new markets)							0.234	0.054	***
D (innov for quality improvement)							0.266	0.058	***
D (within group impt info source)	0.204	0.055	***	0.372	0.054		0.441	0.058	***
D (suppliers important info source)	-0.288	0.048	***	-0.067	0.048		-0.108	0.051	*
D (customers impt info source)	0.054	0.052		0.250	0.051	***	0.191	0.056	***
D (competitors impt info source)	0.048	0.046		0.007	0.046		-0.173	0.049	***
D (universities impt info source)	0.167	0.062	**	0.086	0.070		-0.063	0.065	
D (imitator)	-0.282	0.059	***	-0.275	0.064	***			
Year dummies (2)	65.4 (0.000)***			80.1 (0.000)***			1.5 (0.464)		
Two-digit sector dummies (25)	298.2 (0.000)***			105.4 (0.000)***			52.9 (0.000)***		
Wald test for model (d.f.)				5,269.2 (125)***					
Corr (formal IP, informal IP)	0.553	0.019	***						
Corr (formal IP, innovation)	0.202	0.026	***						
Corr (informal IP, innovation)	0.237	0.026	***						

Note: The method of estimation is maximum likelihood on a trivariate probit model. Standard Errors are clustered around the enterprise

Table B2. Multivariate Probit estimates of IP choice and process innovation

7,144 observations on 5,684 firms; Log likelihood = -8,994.8

	Formal IP methods			*Informal IP methods*			*Process Innovator or imitator*		
	Coeff.	*Std. err*		*Coeff.*	*Std. err*		*Coeff.*	*Std. err*	
Log (predicted IS per employee)	1.025	0.065	***	0.936	0.064	***	0.139	0.066	*
Log (n of employees)	0.400	0.020	***	0.327	0.020	***	0.098	0.021	***
Log (firm age in 2011)	0.036	0.054		0.045	0.053		0.002	0.056	
D (collaborates)	-0.172	0.054	**	-0.110	0.057		0.567	0.054	***
Firm perception of market risk	0.336	0.043	***	0.371	0.044	***	0.120	0.044	**
Firm perception of fin. Constraints	0.090	0.042	*	0.267	0.044	***	0.012	0.043	
Firm - impt. of reg & standards	0.153	0.050	**	0.133	0.052	*	-0.183	0.054	***
Firm - impt. of env, H&S regs	0.042	0.051		0.133	0.054	*	0.159	0.056	**
D (innov to increase flexibility)							0.482	0.056	***
D (innov to increase capacity)							0.410	0.053	***
D (innov to reduce unit cost)							0.176	0.054	***
D (within group impt info source)	0.213	0.056	***	0.358	0.054	***	0.516	0.062	***
D (suppliers important info source)	-0.271	0.048	***	-0.070	0.048		0.247	0.052	***
D (customers impt info source)	0.041	0.052		0.237	0.051	***	0.049	0.057	
D (competitors impt info source)	0.041	0.046		0.008	0.046		-0.127	0.049	**
D (universities impt info source)	0.169	0.063	**	0.110	0.071		-0.104	0.063	
D (imitator)	-0.197	0.061	***	0.042	0.065				
Year dummies (2)	65.4 (0.000)***			80.1 (0.000)***			21.1 (0.000)***		
Two-digit sector dummies (25)	298.2 (0.000)***			105.4 (0.000)***			45.2 (0.000)***		
Wald test for model (d.f.)				5,042.8 (125)***					
Corr (formal IP, informal IP)	0.556	0.018	***						
Corr (formal IP, innovation)	0.105	0.032	***						
Corr (informal IP, innovation)	0.118	0.033	***						

Note: The method of estimation is maximum likelihood on a trivariate probit model. Standard Errors are robust to heteroskedasticity, and clustered on enterprise.

Table B3 shows the production function estimates from the innovation spending model, which are essentially identical to those for the R&D spending model. The conclusion is that instrumenting either R&D or innovation spending and innovation itself by firm characteristics produces fitted values that are essentially the same in their relationship to productivity. Experiments using the components of innovation spending separately produced similar results. It appears that in the absence of better and more specific instruments, it may be difficult to see the differential impact of the different types of spending.

Table B3. OLS Estimates of the production function - innovation spending model

Dependent variable	*Log (turnover)*											
Type of innovation	*Product innovation*			*Process innovation*			*New-to-market product innovation*			*New-to-market process innovation*		
	Coeff.	*Std. err*		*Coeff.*	*Std. err*		*Coeff.*	*Std. err*		*Coeff.*	*Std. err*	
Log (n of employees)	0.664	0.011	***	0.664	0.011	***	0.663	0.011	***	0.664	0.011	***
Log (capital)	0.096	0.007	***	0.097	0.007	***	0.096	0.007	***	0.096	0.007	***
Log (materials)	0.276	0.010	***	0.277	0.010	***	0.276	0.010	***	0.277	0.010	***
Scale coefficient#	1.036	0.006	***	1.038	0.006	***	1.035	0.006	***	1.037	0.006	***
Predicted prob of innovation	0.003	0.051		-0.107	0.056		0.048	0.069		-0.282	0.180	
D (formal IP important)*Pred P of innov	-0.009	0.037		-0.013	0.034		0.019	0.032		0.012	0.027	
D (informal IP important)*Pred P of innov	0.028	0.034		0.035	0.032		0.026	0.029		0.030	0.025	
D (formal IP important)	0.126	0.066		0.194	0.076	*	0.088	0.075		0.433	0.186	*
D (informal IP important)	-0.008	0.070		0.020	0.080		-0.022	0.084		0.093	0.216	
Prob innov and formal IP	0.120	0.056	**	0.074	0.068		0.155	0.076	**	0.163	0.216	
Prob innov and informal IP	0.023	0.041		-0.052	0.051		0.052	0.054		-0.159	0.156	
Prob innov and both	0.140	0.031	***	0.129	0.031	***	0.159	0.038	***	0.286	0.105	***
F-test for 4 IP variables	3.7 (0.006)***			6.7 (0.009)***			2.7 (0.027)**			5.9 (0.000)***		
F-test for 2 survey dummies	36.0 (0.000)***			34.2 (0.000)***			35.5 (0.000)***			33.7 (0.000)***		
F-test for 25 industry dummies	22.1 (0.000)***			22.3 (0.000)***			21.7 (0.000)***			22.1 (0.000)***		
F-test for model (df=35)	1361.4 (0.000)***			1357.6 (0.000)***			1357.7 (0.000)***			1357.2 (0.000)***		
R-squared	0.902			0.902			0.902			0.902		
SSR	2,572.7			2,571.6			2,573.4			2,572.2		
Standard error	0.602			0.601			0.602			0.602		

Standard errors robust to heteroskedasticity, clustered on firm.

Shaded coefficients are derived from the estimated coefficients.

7,144 observations on 5,684 firms.

Test is for the scale coefficient equal to unity