ONLINE APPENDIX

(For Online Publication Only)

This appendix presents the figures, data collection methods, regression results of extensions, and robustness checks that could not be included in the text due to space constraints.

- Table of contents -

C.3.1. Software intensity: Share of software patents

C.3.2. Software intensity: Share of citations directed to prior software patents

C.4. Separate regressions: U.S. and Non-U.S.

C.4.1. U.S. firms only

C.4.2. Non-U.S. firms only

D. Share of Software Intensive Firms: U.S. firms vs. Non-U.S. firms

D.1. Above-median

D.2. Above-75th percentile

D.3. Above-90th percentile

E. Software Intensity of Patent Portfolios of Firms: By Industry

E.1. Share of software patents

E.2. Share of citations directed at software patents

F. ICT Education and Immigration Flows

F.1. Description of Data

F.2. Supplementary Figures: Flow

F.2.1. ICT Graduates

F.2.2. ICT Graduates and ICT Immigrants

F.2.3. Offshoring to India

F.3. Supplementary Figures: Stock

F.3.1. Total

F.3.2. ICT Graduates

F.3.3. ICT Graduates and ICT Immigrants

F.3.4. Offshoring to India

A. Citation Function Regressions

A.1. Exclusion of Pharmaceutical Industry

In the main paper, we present citation function regression results by including all four industries (Table I in the text). It is worth showing the regression results driven by dropping pharmaceutical firms from the sample because the firms in pharmaceutical industry would take advantage of software in different ways from the other three industries (auto and auto parts, aerospace and defense, and medical devices). Software itself can be embedded in the products of three industries. But software can hardly be embedded in the main products (drugs) of pharmaceutical industry. Therefore, pharmaceutical firms mostly take advantage of software in the processes of developing and producing products. In other words, software in pharmaceutical industry is mainly related to *process innovation*. On the other hand, the other three industries can not only use software in those processes of developing and producing products, but also actually implant software into the products. Therefore, software in these three industries is related to both *process innovation and product innovation*.

For robustness checks, we present citation function estimation results by excluding patents granted to pharmaceutical firms from our sample patents (Table A.1). The estimation results reported in Table A.1 are very similar to Table I in the main text. In fact, Table A.1 shows somewhat stronger evidence on software-biased shift in the nature of technological change over time. Patents belonging to our sample firms in the three industries (auto and auto parts, aerospace and defense, and medical devices) are 36% more likely to cite software patents than non-software patents, controlling for the sizes of available software and non-software patent pools. In addition, the propensity of non-software patents generated by these firms to cite software prior art increased faster than we included non-software patents granted to pharmaceutical firms.

Table A.1. Citation Function Results: Exclusion of Pharmaceutical Industry

	Full Model				Citations from NSW to SW		Citations from SW to SW	
	Column 1		Column 2		Column 3		Column 4	
	Coefficient	Std.Error	Coefficient	Std.Error	Coefficient	Std.Error	Coefficient	Std.Error
Citing Grant Year								
1987	0.168	0.304	0.610*	0.272	-0.0728	0.238	0.736	0.451
1988	0.141	0.272	0.220	0.198	0.196	0.265	0.270	0.318
1989	0.255	0.274	0.390	0.205	0.146	0.237	0.479	0.337
1990	0.320	0.270	0.561**	0.216	0.284	0.247	0.702	0.363
1991	0.236	0.238	0.347*	0.176	0.143	0.210	0.452	0.293
1992	0.319	0.236	0.330*	0.162	0.181	0.203	0.423	0.269
1993	0.451	0.240	0.292*	0.147	0.309	0.210	0.370	0.243
1994	0.599*	0.243	0.327*	0.141	0.432*	0.213	0.390	0.230
1995	0.928***	0.267	0.567***	0.151	0.637**	0.224	0.645**	0.249
1996	1.091***	0.263	0.687***	0.149	0.601**	0.206	0.805**	0.251
1997	1.272***	0.259	0.845***	0.149	0.884***	0.220	0.997***	0.256
1998	1.457***	0.248	0.763***	0.130	0.925***	0.207	0.871***	0.222
1999	1.717***	0.243	0.766***	0.122	1.013***	0.201	0.807***	0.206
2000	2.525***	0.268	1.103***	0.128	1.899***	0.245	1.134***	0.219
2001	3.021***	0.262	1.454***	0.132	1.953***	0.236	1.538***	0.234
2002	3.673***	0.257	1.770***	0.136	2.593***	0.256	1.845***	0.246
2003	4.275***	0.247	2.038***	0.140	2.687***	0.258	2.114***	0.260
2004	5.648***	0.265	2.969***	0.161	3.686***	0.303	3.257***	0.312
2005	6.726	.	3.596	.	4.784	.	4.031	.
Cited Grant Year								
1986	-0.0603	0.0542	-0.0856*	0.0381	-0.156**	0.0601	-0.113	0.0584
1987	-0.0814	0.0541	-0.0536	0.0401	-0.215***	0.0566	-0.0805	0.0614
…	…	…	…	…	…	…	…	…
2003	-0.940***	0.0177	-0.950***	0.0118	-0.958***	0.0142	-0.963***	0.0145
2004	-0.960***	0.0210	-0.969***	0.0155	-0.972***	0.0189	-0.977***	0.0193
Citing patent: Firm industry								
Aerospace and Defense	0.255***	0.0326	0.118***	0.0215	0.0478	0.0355	0.0946**	0.0337
Medical Devices	1.208***	0.0487	0.417***	0.0267	0.580***	0.0467	0.194***	0.0395
Software Patent								
Citing from Software Patent			-0.159***	0.0372				
Cited Software Patent	0.359***	0.0251	-0.192***	0.0413				
Citing from Software Patent X Cited Software Patent			6.491***	0.137				
Obsolescence	0.305***	0.013	0.319***	0.009	0.329***	0.012	0.338***	0.014
Diffusion	4.53E-6***	1.17E-06	7.81E-6***	1.39E-06	7.19E-6***	1.72E-06	6.98E-5***	1.90E-05
Adj R-Squared	0.911		0.867		0.906		0.902	
Number of Obs	1260		2520		630		630	

The data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Regression specifications are estimated in STATA using the nonlinear least squares algorithm. The dependent variable is an empirical measure of the probability a citing patent with given attributes cites a cited patent with a particular set of attributes. All presented coefficients are relative to base categories, which are the following: citing patent grant year = 1986, cited patent grant year = 1985, citing firm industry = "Automobiles." The rest of the base categories are model specific. * p<0.10, ** p<0.05, *** p<0.01

B. Innovation (Patent) Production Function Regressions

B.1. Software intensity: Share of citations directed to prior software patents

Section B presents robustness checks of patent production function estimation results. In the interest of brevity, only Table II using the share of software patents as software intensity is reported in the main text. Table B.1 reports patent production function estimation results using *software intensity as the share of citations directed to prior software patents* - more precisely, it is the ratio of the number of backward citations to software patents made by the patents generated by the firm from 1981 to 2005 to the number of backward citations to all patents made by the patents generated by the firm from 1981 to 2005 (see section IV.C for detailed explanation on the two software intensity variables). Therefore, Table B.1 corresponds to Table II in the text. The results in Table B.1 are qualitatively similar to the results in Table II. In Table B.1, we do not observe statistically significant relative patent productivity gains by above-median software intensity firms in the periods of 1986-1990 and 1991-1995. However, we do observe positive and statistically significant coefficients in the last two periods (1996-2000 and 2001-2005). Besides, the key pattern – increasing R&D productivity gap over time in favor of more software intensive firms – remains the same. It is therefore clear that the main results of patent production function estimations are not sensitive to different software intensity variables.

Moreover, we constructed an alternative software intensity variable – the share of citations made from posterior patents to a firm's software patents – in order to control the quality of a firm's software patents. More precisely, it is the ratio of the number of forward citations to a firm's *software* patents made by the all USPTO patents granted from 1981 to 2010 to the number of forward citations to a firm's *all* patents make by the all USPTO patents granted from 1981 to 2010. The results using this new software intensity variable are qualitatively similar to the results in Table II and Table B.1. Those additional regression results are available from the authors upon request.

Table B.1. Software intensity: Share of citations directed to prior software patents

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	NB: RE	NB: FE	NB: RE	NB: FE	NB: RE	NB: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.085*****	**0.058*****	**0.190*****	**0.174*****	**0.178*****	**0.162*****
	(0.013)	(0.013)	(0.011)	(0.011)	(0.011)	(0.012)
Software Intensity Dummy	**-0.276****	**-0.252****	**-0.297****	**-0.308*****	**-0.318*****	**-0.308*****
	(0.112)	(0.114)	(0.116)	(0.117)	(0.113)	(0.115)
Software Intensity Dummy * 1986-1990	**0.103**	**0.106**	**0.097**	**0.100**	**0.112**	**0.115**
	(0.125)	(0.126)	(0.139)	(0.139)	(0.131)	(0.131)
Software Intensity Dummy * 1991-1995	**0.208***	**0.215***	**0.100**	**0.101**	**0.198**	**0.203**
	(0.119)	(0.120)	(0.132)	(0.132)	(0.125)	(0.125)
Software Intensity Dummy * 1996-2000	**0.466*****	**0.478*****	**0.340*****	**0.340*****	**0.494*****	**0.512*****
	(0.114)	(0.115)	(0.126)	(0.126)	(0.121)	(0.122)
Software Intensity Dummy * 2001-2005	**0.564*****	**0.579*****	**0.524*****	**0.537*****	**0.469*****	**0.481*****
	(0.118)	(0.119)	(0.129)	(0.130)	(0.128)	(0.128)
1986-1990	**0.009**	**0.028**	**-0.049**	**-0.037**	**-0.075**	**-0.066**
	(0.092)	(0.092)	(0.103)	(0.103)	(0.096)	(0.097)
1991-1995	**0.089**	**0.119**	**0.094**	**0.116**	**-0.040**	**-0.022**
	(0.089)	(0.089)	(0.098)	(0.098)	(0.093)	(0.093)
1996-2000	**0.354*****	**0.389*****	**0.468*****	**0.498*****	**-0.036**	**-0.016**
	(0.086)	(0.087)	(0.094)	(0.094)	(0.091)	(0.091)
2001-2005	**-0.257*****	**-0.212****	**-0.300*****	**-0.280*****	**-1.098*****	**-1.068*****
	(0.091)	(0.091)	(0.098)	(0.098)	(0.097)	(0.097)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	3854	3843	3854	3843	3854	3843

The software intensity is based on the share of citations directed to prior software patents. The patent-related data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Firm-level R&D data are collected from Compustat database, Edgar database, Amadeus database, the Kaisha Shiki Ho Survey database, R&D scoreboard, TS 2000 database (the Korea Listed Companies Association), and firm annual reports. * p<0.10, ** p<0.05, *** p<0.01

B.2. Software intensity to vary within firms over time

The next robustness checks concern the time-invariant software intensity dummy variables in the main paper. As noted in the main paper, we use time-invariant software intensity dummy variables instead of annual software intensity measures (see section IV.C). Section B.2 provides robustness checks of patent production function estimation results (Table II) by allowing software intensity to *vary within firms over time*. We do not introduce software intensity variables that vary every year in this section because of the same reason we discussed in the text. Instead, we allow firms to have different levels of software intensity for *each subperiod*. In other words, firms' subperiod-specific software intensity dummy variables switch between the above-median software intensity group and the below-median software intensity group. This approach allows us to observe R&D productivity gaps between the two groups not only across subperiods but also every year within each subperiod.

The tables in Section B.2 report patent production function estimation results using subperiod-variant software intensity and different dependent variables.[1] The dependent variables of Table B.2.1, B.2.2 and B.2.3 are the number of patents, the number of claims, and the number of forward citations, respectively. The key variables are interaction terms of software intensity dummy variables and year dummy variables. We find that the coefficients on our key variables (the interaction terms) are statistically less significant than the key coefficients in Table II in the main text when we use the number of patents as a dependent variable. However, the coefficients of our key variables of the tables in this section show a rising trend *only* for the last subperiod (1997-2005), which supports the results in Table II. Especially, Table B.2.2 and Table B.2.3 provide evidence that highly software intensive firms started to produce more patents per R&D dollar than less software intensive firms starting the late 1990s, without sacrificing the quality of their patent portfolios.

[1] We construct the subperiods (1981-1988, 1989-1996, and 1997-2005) somewhat different from the subperiods in Table II in the text. The main reason behind this construction is the limited number of observations for each subperiod, especially early subperiods. The results using different subperiods (1981-1985, 1986-1990, 1991-1995, 1996-2000, and 2001-2005) are qualitatively similar to the results in Section B.2. These results are available from the authors by request.

Table B.2.1. Dependent variable: Number of patents

Dependent Variable	**Number of Patents**		
	NB: FE	NB: FE	NB: FE
	1981-1988	**1989-1996**	**1997-2005**
Log R&D	**0.259***** (0.049)	**0.367***** (0.034)	**0.040*** (0.022)
Software Intensity Dummy_8188	**-0.062** (0.519)		
Software Intensity Dummy_8996		**-0.762***** (0.212)	
Software Intensity Dummy_9705			**0.145** (0.139)
Software Intensity Dummy * 2nd Year	**-0.257** (0.277)	**0.171** (0.149)	**0.108** (0.114)
Software Intensity Dummy * 3rd Year	**-0.338** (0.268)	**0.125** (0.147)	**0.211*** (0.110)
Software Intensity Dummy * 4th Year	**0.120** (0.267)	**0.208** (0.151)	**0.024** (0.105)
Software Intensity Dummy * 5th Year	**-0.300** (0.241)	**0.070** (0.142)	**-0.012** (0.104)
Software Intensity Dummy * 6th Year	**-0.451*** (0.241)	**0.031** (0.137)	**0.004** (0.108)
Software Intensity Dummy * 7th Year	**-0.313** (0.240)	**0.065** (0.134)	**0.089** (0.120)
Software Intensity Dummy * 8th Year	**-0.508**** (0.237)	**0.059** (0.134)	**0.100** (0.148)
Software Intensity Dummy * 9th Year			**0.189** (0.203)
Year Dummies	**Yes**	**Yes**	**Yes**
Industry Dummies	**Yes**	**Yes**	**Yes**
Number of Obs	615	1214	1914

Fixed effect regression results are reported. Random effect regression results are qualitatively the same. Those results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

Table B.2.2. Dependent variable: Number of claims

Dependent Variable	**Number of Claims**		
	NB: FE	NB: FE	NB: FE
	1981-1988	**1989-1996**	**1997-2005**
Log R&D	**0.406***** (0.036)	**0.294***** (0.022)	**0.249***** (0.017)
Software Intensity Dummy_8188	**1.680***** (0.364)		
Software Intensity Dummy_8996		**0.714***** (0.190)	
Software Intensity Dummy_9705			**-0.156** (0.129)
Software Intensity Dummy * 2nd Year	**-0.349** (0.428)	**0.216** (0.237)	**0.205** (0.151)
Software Intensity Dummy * 3rd Year	**-0.625** (0.405)	**0.151** (0.229)	**0.433***** (0.145)
Software Intensity Dummy * 4th Year	**-0.408** (0.394)	**0.072** (0.231)	**0.152** (0.139)
Software Intensity Dummy * 5th Year	**-0.401** (0.375)	**-0.263** (0.215)	**0.169** (0.138)
Software Intensity Dummy * 6th Year	**-0.746**** (0.367)	**-0.181** (0.210)	**0.280**** (0.142)
Software Intensity Dummy * 7th Year	**-0.544** (0.365)	**-0.219** (0.206)	**0.397**** (0.156)
Software Intensity Dummy * 8th Year	**-0.699*** (0.359)	**-0.347*** (0.204)	**0.430**** (0.180)
Software Intensity Dummy * 9th Year			**0.514**** (0.222)
Year Dummies	**Yes**	**Yes**	**Yes**
Industry Dummies	**Yes**	**Yes**	**Yes**
Number of Obs	615	1214	1914

Fixed effect regression results are reported. Random effect regression results are qualitatively the same. Those results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

Table B.2.3. Dependent variable: Number of citations

Dependent Variable	**Number of Citations**		
	NB: FE	NB: FE	NB: FE
	1981-1988	**1989-1996**	**1997-2005**
Log R&D	**0.449***** (0.036)	**0.296***** (0.023)	**0.218***** (0.019)
Software Intensity Dummy_8188	**1.625***** (0.362)		
Software Intensity Dummy_8996		**0.820***** (0.192)	
Software Intensity Dummy_9705			**-0.113** (0.127)
Software Intensity Dummy * 2nd Year	**-0.376** (0.430)	**0.114** (0.235)	**0.267**** (0.132)
Software Intensity Dummy * 3rd Year	**-0.408** (0.413)	**0.033** (0.228)	**0.360***** (0.130)
Software Intensity Dummy * 4th Year	**-0.201** (0.400)	**0.009** (0.232)	**0.148** (0.128)
Software Intensity Dummy * 5th Year	**-0.304** (0.378)	**-0.320** (0.214)	**0.133** (0.133)
Software Intensity Dummy * 6th Year	**-0.418** (0.373)	**-0.251** (0.211)	**0.159** (0.142)
Software Intensity Dummy * 7th Year	**-0.329** (0.370)	**-0.382*** (0.207)	**0.227** (0.162)
Software Intensity Dummy * 8th Year	**-0.538** (0.365)	**-0.364*** (0.210)	**0.322*** (0.189)
Software Intensity Dummy * 9th Year			**0.324** (0.236)
Year Dummies	**Yes**	**Yes**	**Yes**
Industry Dummies	**Yes**	**Yes**	**Yes**
Number of Obs	615	1206	1914

Fixed effect regression results are reported. Random effect regression results are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

B.3. Poisson model

In the main paper, we present patent production function regression results using a *negative binomial* estimation method (Table II in the text). For robustness checks, patent production function regression results using a Poisson model are presented in Section B.3. The regression results in Table B.3.1 and B.3.2 use the share of software patents and the share of citations directed to prior software patents in order to construct software intensity, respectively. The results in Table B.3.1 are analogous to the results in Table II. Similarly, the results in Table B.3.2 are similar to the results in Table B.1. There results suggest that our main results of patent production function regression are not sensitive to the choice of count data regression models.

Table B.3.1. Poisson Model (Software intensity: Share of software patents)

Dependent Variable	Number of Patents		Number of Claims		Number of Citations	
	Poisson: RE	Poisson: FE	Poisson: RE	Poisson: FE	Poisson: RE	Poisson: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.240*****	**0.234*****	**0.189*****	**0.189*****	**0.337*****	**0.336*****
	(0.005)	(0.005)	(0.001)	(0.001)	(0.002)	(0.002)
Software Intensity Dummy	**0.246**		**0.348****		**0.291***	
	(0.155)		(0.157)		(0.161)	
Software Intensity Dummy * 1986-1990	**0.311*****	**0.312*****	**0.330*****	**0.330*****	**0.301*****	**0.301*****
	(0.024)	(0.024)	(0.007)	(0.007)	(0.007)	(0.007)
Software Intensity Dummy * 1991-1995	**0.412*****	**0.412*****	**0.426*****	**0.426*****	**0.562*****	**0.562*****
	(0.024)	(0.024)	(0.007)	(0.007)	(0.006)	(0.006)
Software Intensity Dummy * 1996-2000	**0.419*****	**0.419*****	**0.471*****	**0.471*****	**0.589*****	**0.589*****
	(0.022)	(0.022)	(0.007)	(0.007)	(0.006)	(0.006)
Software Intensity Dummy * 2001-2005	**0.505*****	**0.505*****	**0.638*****	**0.638*****	**0.760*****	**0.760*****
	(0.023)	(0.023)	(0.007)	(0.007)	(0.008)	(0.008)
1986-1990	**-0.163*****	**-0.162*****	**-0.003**	**-0.003**	**-0.092*****	**-0.091*****
	(0.022)	(0.022)	(0.007)	(0.007)	(0.006)	(0.006)
1991-1995	**-0.129*****	**-0.126*****	**0.140*****	**0.140*****	**-0.207*****	**-0.207*****
	(0.021)	(0.021)	(0.006)	(0.006)	(0.006)	(0.006)
1996-2000	**0.118*****	**0.123*****	**0.541*****	**0.542*****	**-0.403*****	**-0.403*****
	(0.021)	(0.021)	(0.006)	(0.006)	(0.006)	(0.006)
2001-2005	**-0.311*****	**-0.304*****	**0.140*****	**0.141*****	**-1.903*****	**-1.902*****
	(0.022)	(0.022)	(0.006)	(0.006)	(0.008)	(0.008)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	3884	3873	3884	3873	3884	3873

The software intensity is based on the *share of software patents*. The patent-related data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Firm-level R&D data are collected from Compustat database, Edgar database, Amadeus database, the Kaisha Shiki Ho Survey database, R&D scoreboard, TS 2000 database (the Korea Listed Companies Association), and firm annual reports. * p<0.10, ** p<0.05, *** p<0.01

Table B.3.2. Poisson Model (Software intensity: Share of citations directed to prior software patents)

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	Poisson: RE	Poisson: FE	Poisson: RE	Poisson: FE	Poisson: RE	Poisson: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.224*****	**0.218*****	**0.179*****	**0.178*****	**0.328*****	**0.327*****
	(0.005)	(0.005)	(0.001)	(0.001)	(0.002)	(0.002)
Software Intensity Dummy	**0.267***		**0.405****		**0.283***	
	(0.156)		(0.158)		(0.162)	
Software Intensity Dummy * 1986-1990	**0.173*****	**0.174*****	**0.163*****	**0.163*****	**0.150*****	**0.150*****
	(0.022)	(0.022)	(0.007)	(0.007)	(0.006)	(0.006)
Software Intensity Dummy * 1991-1995	**0.235*****	**0.237*****	**0.189*****	**0.189*****	**0.394*****	**0.394*****
	(0.021)	(0.021)	(0.007)	(0.007)	(0.006)	(0.006)
Software Intensity Dummy * 1996-2000	**0.367*****	**0.369*****	**0.347*****	**0.347*****	**0.514*****	**0.514*****
	(0.021)	(0.020)	(0.006)	(0.006)	(0.006)	(0.006)
Software Intensity Dummy * 2001-2005	**0.542*****	**0.545*****	**0.581*****	**0.582*****	**0.772*****	**0.772*****
	(0.021)	(0.021)	(0.006)	(0.006)	(0.008)	(0.008)
1986-1990	**-0.030**	**-0.029**	**0.148*****	**0.148*****	**0.049*****	**0.049*****
	(0.019)	(0.019)	(0.006)	(0.006)	(0.005)	(0.005)
1991-1995	**0.042****	**0.044****	**0.350*****	**0.350*****	**-0.042*****	**-0.041*****
	(0.018)	(0.018)	(0.006)	(0.006)	(0.005)	(0.005)
1996-2000	**0.198*****	**0.202*****	**0.669*****	**0.670*****	**-0.305*****	**-0.305*****
	(0.018)	(0.018)	(0.005)	(0.005)	(0.005)	(0.005)
2001-2005	**-0.292*****	**-0.287*****	**0.224*****	**0.224*****	**-1.867*****	**-1.867*****
	(0.019)	(0.019)	(0.006)	(0.006)	(0.007)	(0.007)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	3884	3873	3884	3873	3884	3873

The software intensity is based on the *share of citations directed to prior software patents*. The patent-related data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Firm-level R&D data are collected from Compustat database, Edgar database, Amadeus database, the Kaisha Shiki Ho Survey database, R&D scoreboard, TS 2000 database (the Korea Listed Companies Association), and firm annual reports. * p<0.10, ** p<0.05, *** p<0.01

B.4. Exclusion of Pharmaceutical Industry

In this section, we report innovation (patent) regression function estimation results by dropping the firms in pharmaceutical industry of our sample. The rationale behind these robustness checks is described in section A.1. Two different measures of software intensity are used. Table B.4.1 reports the regression results using the share of software patents as software intensity. The regression results based on the other software intensity variable (share of citations directed to prior software patents) are found in Table B.4.2. The regression results in both tables are qualitatively the same as the results in Table II in the main paper.

Table B.4.1. Exclusion of Pharmaceutical Industry (Software intensity: Share of software patents)

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	NB: RE	NB: FE	NB: RE	NB: FE	NB: RE	NB: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.076*****	**0.048*****	**0.178*****	**0.164*****	**0.170*****	**0.156*****
	(0.016)	(0.016)	(0.013)	(0.013)	(0.013)	(0.014)
Software Intensity Dummy	**-0.096**	**-0.031**	**-0.241***	**-0.231***	**-0.200**	**-0.167**
	(0.129)	(0.131)	(0.134)	(0.135)	(0.131)	(0.132)
Software Intensity Dummy * 1986-1990	**0.367****	**0.383*****	**0.307***	**0.339****	**0.314****	**0.333****
	(0.145)	(0.146)	(0.162)	(0.163)	(0.153)	(0.154)
Software Intensity Dummy * 1991-1995	**0.484*****	**0.513*****	**0.342****	**0.381****	**0.398*****	**0.423*****
	(0.139)	(0.140)	(0.154)	(0.155)	(0.147)	(0.147)
Software Intensity Dummy * 1996-2000	**0.484*****	**0.522*****	**0.332****	**0.395*****	**0.367*****	**0.419*****
	(0.132)	(0.133)	(0.147)	(0.147)	(0.141)	(0.142)
Software Intensity Dummy * 2001-2005	**0.549*****	**0.587*****	**0.424*****	**0.488*****	**0.323****	**0.357****
	(0.137)	(0.137)	(0.150)	(0.151)	(0.148)	(0.149)
1986-1990	**-0.178**	**-0.166**	**-0.170**	**-0.179**	**-0.205***	**-0.207***
	(0.115)	(0.116)	(0.128)	(0.129)	(0.122)	(0.122)
1991-1995	**-0.171**	**-0.153**	**-0.083**	**-0.090**	**-0.217***	**-0.217***
	(0.110)	(0.111)	(0.121)	(0.121)	(0.116)	(0.116)
1996-2000	**0.227****	**0.245****	**0.389*****	**0.371*****	**0.010**	**0.003**
	(0.105)	(0.106)	(0.115)	(0.116)	(0.111)	(0.112)
2001-2005	**-0.245****	**-0.218****	**-0.200***	**-0.213***	**-0.946*****	**-0.936*****
	(0.109)	(0.111)	(0.118)	(0.119)	(0.117)	(0.118)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	2733	2733	2733	2733	2733	2733

The software intensity is based on the *share of software patents*. * p<0.10, ** p<0.05, *** p<0.01

Table B.4.2. Exclusion of Pharmaceutical Industry (Software intensity: Share of citations directed to prior software patents)

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	NB: RE	NB: FE	NB: RE	NB: FE	NB: RE	NB: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.087*** **	**0.059*** **	**0.187*** **	**0.173*** **	**0.178*** **	**0.164*** **
	(0.016)	(0.016)	(0.013)	(0.013)	(0.014)	(0.014)
Software Intensity Dummy	**-0.023**	**0.039**	**-0.133**	**-0.116**	**-0.134**	**-0.105**
	(0.127)	(0.129)	(0.132)	(0.133)	(0.129)	(0.130)
Software Intensity Dummy * 1986-1990	**0.093**	**0.100**	**-0.021**	**-0.012**	**-0.027**	**-0.019**
	(0.142)	(0.143)	(0.159)	(0.159)	(0.150)	(0.150)
Software Intensity Dummy * 1991-1995	**0.108**	**0.123**	**-0.081**	**-0.069**	**-0.008**	**0.008**
	(0.137)	(0.138)	(0.152)	(0.152)	(0.144)	(0.144)
Software Intensity Dummy * 1996-2000	**0.324** **	**0.348*** **	**0.122**	**0.153**	**0.257***	**0.302** **
	(0.130)	(0.131)	(0.144)	(0.145)	(0.139)	(0.139)
Software Intensity Dummy * 2001-2005	**0.505*** **	**0.526*** **	**0.380** **	**0.410*** **	**0.313** **	**0.347** **
	(0.135)	(0.136)	(0.148)	(0.149)	(0.146)	(0.146)
1986-1990	**-0.004**	**0.014**	**0.030**	**0.042**	**0.002**	**0.008**
	(0.106)	(0.107)	(0.118)	(0.118)	(0.110)	(0.111)
1991-1995	**0.050**	**0.075**	**0.156**	**0.172**	**0.011**	**0.017**
	(0.102)	(0.103)	(0.112)	(0.112)	(0.106)	(0.107)
1996-2000	**0.334*** **	**0.361*** **	**0.519*** **	**0.528*** **	**0.087**	**0.087**
	(0.099)	(0.100)	(0.108)	(0.108)	(0.104)	(0.104)
2001-2005	**-0.200***	**-0.162**	**-0.158**	**-0.143**	**-0.925*** **	**-0.911*** **
	(0.104)	(0.105)	(0.112)	(0.112)	(0.110)	(0.111)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	2718	2718	2718	2718	2718	2718

The software intensity is based on the *share of citations directed to prior software patents*. * p<0.10, ** p<0.05, *** p<0.01

B.5. Separate regressions: U.S. and Non-U.S.

Table II in the main paper includes all firms in our sample, both U.S. firms and non-U.S. firms. An alternative approach would be to conduct separate regressions using two subsamples; U.S. firms and non-U.S. firms. Section B.5 reports patent production function estimation results using these two subsamples. The rationale behind these separate regressions is threefold. First, we observe that U.S. firms are disproportionally more software intensive than non-U.S. firms (see online appendix section D). Second, the United States has a substantial human resource advantage in the domain of software engineers, and it widens greatly over time (see online appendix section F). This means that U.S. firms can access skilled software engineers relatively easily in the domestic labor market than non-U.S. firms. Third, based on the first two observations, the relationship between software intensity and R&D productivity could be different between U.S. firms and non-U.S. firms. The regression results using U.S. firms (Table B.5.1) are very similar to the results in Table II in the text. Once again, all key patterns – the coefficients on our key variables (interaction terms) are statistically significant in the last two subperiods and are generally increasing over time – remain the same in Table B.5.2 (using non-U.S. firms). This suggests that our regression results are unlikely driven by relatively more software intensive U.S. firms only.

Table B.5.1. U.S. firms only

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	NB: RE	NB: FE	NB: RE	NB: FE	NB: RE	NB: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**0.236*****	**0.203*****	**0.366*****	**0.342*****	**0.385*****	**0.359*****
	(0.018)	(0.019)	(0.015)	(0.016)	(0.017)	(0.017)
Software Intensity Dummy	**-0.257****	**-0.231***	**-0.402*****	**-0.409*****	**-0.389*****	**-0.398*****
	(0.121)	(0.125)	(0.124)	(0.126)	(0.117)	(0.120)
Software Intensity Dummy * 1986-1990	**0.202**	**0.188**	**0.283***	**0.266***	**0.390*****	**0.382*****
	(0.130)	(0.132)	(0.149)	(0.150)	(0.128)	(0.129)
Software Intensity Dummy * 1991-1995	**0.476*****	**0.470*****	**0.514*****	**0.498*****	**0.705*****	**0.704*****
	(0.123)	(0.125)	(0.141)	(0.142)	(0.123)	(0.124)
Software Intensity Dummy * 1996-2000	**0.644*****	**0.639*****	**0.742*****	**0.727*****	**0.778*****	**0.764*****
	(0.120)	(0.122)	(0.136)	(0.137)	(0.124)	(0.126)
Software Intensity Dummy * 2001-2005	**0.573*****	**0.565*****	**0.812*****	**0.797*****	**0.507*****	**0.499*****
	(0.132)	(0.134)	(0.146)	(0.148)	(0.143)	(0.145)
1986-1990	**-0.045**	**-0.025**	**-0.037**	**-0.025**	**-0.162**	**-0.157**
	(0.105)	(0.106)	(0.120)	(0.120)	(0.104)	(0.105)
1991-1995	**-0.063**	**-0.040**	**-0.006**	**0.008**	**-0.261****	**-0.256****
	(0.100)	(0.102)	(0.113)	(0.114)	(0.102)	(0.102)
1996-2000	**-0.054**	**-0.008**	**-0.016**	**0.008**	**-0.593*****	**-0.571*****
	(0.100)	(0.101)	(0.111)	(0.111)	(0.104)	(0.104)
2001-2005	**-0.778*****	**-0.721*****	**-0.896*****	**-0.878*****	**-1.837*****	**-1.810*****
	(0.111)	(0.112)	(0.120)	(0.121)	(0.119)	(0.121)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	2101	2101	2101	2101	2101	2101

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

Table B.5.2. Non-U.S. firms only

Dependent Variable	**Number of Patents**		**Number of Claims**		**Number of Citations**	
	NB: RE	NB: FE	NB: RE	NB: FE	NB: RE	NB: FE
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log R&D	**-0.010**	**-0.041****	**0.076*****	**0.047*****	**0.052*****	**0.024**
	(0.020)	(0.020)	(0.017)	(0.018)	(0.018)	(0.018)
Software Intensity Dummy	**-0.476***	**-0.434**	**-0.630****	**-0.624****	**-0.546***	**-0.527***
	(0.277)	(0.278)	(0.284)	(0.284)	(0.280)	(0.281)
Software Intensity Dummy * 1986-1990	**0.661****	**0.688****	**0.930*****	**0.976*****	**0.915*****	**0.953*****
	(0.296)	(0.297)	(0.311)	(0.311)	(0.304)	(0.305)
Software Intensity Dummy * 1991-1995	**0.564***	**0.597****	**0.704****	**0.741****	**0.659****	**0.688****
	(0.289)	(0.289)	(0.301)	(0.301)	(0.295)	(0.296)
Software Intensity Dummy * 1996-2000	**0.666****	**0.688****	**0.784*****	**0.801*****	**0.751*****	**0.781*****
	(0.282)	(0.282)	(0.294)	(0.294)	(0.289)	(0.290)
Software Intensity Dummy * 2001-2005	**0.716****	**0.741*****	**0.863*****	**0.891*****	**0.777*****	**0.801*****
	(0.284)	(0.284)	(0.295)	(0.296)	(0.292)	(0.293)
1986-1990	**0.020**	**0.043**	**-0.287**	**-0.264**	**-0.254**	**-0.228**
	(0.220)	(0.220)	(0.232)	(0.232)	(0.227)	(0.228)
1991-1995	**0.260**	**0.298**	**0.040**	**0.086**	**0.035**	**0.083**
	(0.215)	(0.215)	(0.225)	(0.225)	(0.221)	(0.221)
1996-2000	**0.803*****	**0.851*****	**0.717*****	**0.789*****	**0.407***	**0.476****
	(0.211)	(0.211)	(0.220)	(0.220)	(0.217)	(0.217)
2001-2005	**0.371***	**0.434****	**0.133**	**0.193**	**-0.505****	**-0.425***
	(0.214)	(0.214)	(0.222)	(0.223)	(0.220)	(0.221)
Industry Dummies	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Number of Obs	1787	1776	1787	1776	1787	1776

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

C. Tobin's Q Regressions

C.1. Software intensity: Share of citations directed to prior software patents

Given space constraints, we report Table III using the share of software patents as software intensity in the main text (see section IV.C for detailed explanation on the two software intensity variables). Hence, we would like to report robustness checks of Tobin's Q regression results in this section. Table C.1 presents Tobin's Q regression results by estimating equation (13) using *the share of citations directed to prior software patents as software intensity*. Therefore, Table C.1 is parallel to Table III in the text. The results in Table C.1 are qualitatively similar to the results in Table III. As we expected, the coefficient on the key variable of the last two periods (1989-1996 and 1997-2005) is positive and statistically significant. In addition, the size of coefficient of the last period (1997-2005) is greater than the one of the middle period (1989-1996).

As we did in Section B.1, we constructed an additional software intensity variable – the share of citations made from posterior patents to a firm's software patents– in order to control the quality of software patents. The results using this alternative software intensity variable are qualitatively similar to the results in Table III and Table C.1. Those additional regression results are available from the authors by request.

Table C.1. Software intensity: Share of citations directed to prior software patents

lnQ	**Full Sample** NLS	**1981-1988** NLS	**1989-1996** NLS	**1997-2005** NLS
RD/Assets	**-0.033** (0.060)	**-0.252** (0.159)	**-0.144*** (0.079)	**0.407***** (0.107)
RD/Assets * Software Intensity	**0.800***** (0.160)	**0.203** (0.285)	**0.512***** (0.180)	**1.021***** (0.286)
Software Intensity	**-0.301***** (0.046)	**-0.489***** (0.092)	**-0.227***** (0.060)	**-0.217***** (0.065)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	2288	365	719	1204
Adj R-Squared	0.415	0.388	0.418	0.488

The software intensity is based on the *share of citations directed to prior software patents*. The patent-related data for regression estimations presented in this table are drawn from the CASSIS patent database maintained by the United States Patent and Trademark Office and from the NBER Patent Data Project database. Firm-level R&D data are collected from Compustat, Edgar, Amadeus, the Kaisha Shiki Ho Survey database, R&D scoreboard, TS 2000 database (the Korea Listed Companies Association), and firm annual reports. Other firm-level financial data (such as assets, long-term debt, short-term debt, the number of stocks and the price of stocks) are drawn from Compustat, the Development Bank of Japan (BDJ) database, and the TS 2000 (the Korea Listed Companies Association). * p<0.10, ** p<0.05, *** p<0.01

C.2. R&D stock: Depreciation rate

The next set of robustness checks concerns about the level of R&D stock depreciation rate. Knowing that different R&D stock depreciation rates can affect the size of actual R&D stock for each year, it is instructive to ask whether the Tobin's Q regression results are actually different when using different R&D stock depreciation rates. Section C.2 delivers robustness checks of Tobin's Q regression by using different levels of R&D stock depreciation rates. Table III in the main text uses 15% of R&D stock depreciation rate. Table C.2.1 and C.2.2 report Tobin's Q regression results using 10% and 30% as R&D stock depreciation rates, respectively. The results in both tables are qualitatively similar to the results in Table III – the estimated (private) return to R&D investment for highly software intensive firms started to become larger than that of less software intensive firms in the middle period (1989-1996) and the difference intensified in the most recent period (1997-2005). It is important to note that the choice of different levels of R&D stock depreciation rates (10%, 15% and 30%) does not alter the main results qualitatively.

The regression results using (additionally) different levels of depreciation rates such as 20% and 25% are analogous to the results in Table III, Table C.2.1 and Table C.2.2. Those additional regression results are available from the authors upon request.

Table C.2.1. Depreciation rate: 10%

lnQ	**Full Sample**	**1981-1988**	**1989-1996**	**1997-2005**
	NLS	NLS	NLS	NLS
RD/Assets	**-0.101****	**-0.096**	**-0.077**	**0.115**
	(0.046)	(0.164)	(0.071)	(0.072)
RD/Assets *	**0.653*****	**0.147**	**0.274***	**1.091*****
Software Intensity	(0.120)	(0.268)	(0.152)	(0.229)
Software Intensity	**-0.272*****	**-0.582*****	**-0.197*****	**-0.214*****
	(0.044)	(0.088)	(0.062)	(0.065)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	2284	365	718	1201
Adj R-Squared	0.412	0.459	0.415	0.491

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

Table C.2.2. Depreciation rate: 30%

lnQ	**Full Sample**	**1981-1988**	**1989-1996**	**1997-2005**
	NLS	NLS	NLS	NLS
RD/Assets	**0.043**	**-0.047**	**0.103**	**0.595*****
	(0.094)	(0.245)	(0.116)	(0.170)
RD/Assets *	**1.359*****	**0.520**	**1.078*****	**2.913*****
Software Intensity	(0.245)	(0.476)	(0.368)	(0.551)
Software Intensity	**-0.291*****	**-0.651*****	**-0.309*****	**-0.280*****
	(0.044)	(0.093)	(0.067)	(0.065)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	2297	365	724	1208
Adj R-Squared	0.423	0.461	0.435	0.517

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

C.3. Exclusion of Pharmaceutical Industry

In this section, we report Tobin's Q regression function estimation results by dropping the firms in pharmaceutical industry of our sample. The rationale behind these robustness checks is described in section A.1. Following section B.4, two different measures of software intensity are used. Table C.3.1 reports the regression results using the share of software patents as software intensity. The regression results based on the other software intensity variable (share of citations directed to prior software patents) are found in Table C.3.2. The regression results in both tables are qualitatively similar to the results in Table III in the main text.

Table C.3.1. Exclusion of Pharmaceutical Industry (Software intensity: Share of software patents)

lnQ	**Full Sample** NLS	**1981-1988** NLS	**1989-1996** NLS	**1997-2005** NLS
RD/Assets	**-0.026** (0.075)	**0.043** (0.213)	**-0.061** (0.114)	**0.232**** (0.105)
RD/Assets * Software Intensity	**0.360**** (0.153)	**-0.184** (0.332)	**0.120** (0.209)	**1.095***** (0.316)
Software Intensity	**-0.242***** (0.046)	**-0.514***** (0.092)	**-0.215***** (0.066)	**-0.247***** (0.065)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	1566	270	499	797
Adj R-Squared	0.236	0.425	0.157	0.351

The software intensity is based on the *share of software patents*. * p<0.10, ** p<0.05, *** p<0.01

Table C.3.2. Exclusion of Pharmaceutical Industry (Software intensity: Share of citations directed to prior software patents)

lnQ	**Full Sample** NLS	**1981-1988** NLS	**1989-1996** NLS	**1997-2005** NLS
RD/Assets	**-0.072** (0.066)	**-0.055** (0.195)	**-0.086** (0.106)	**0.254**** (0.107)
RD/Assets * Software Intensity	**0.596***** (0.182)	**-0.106** (0.323)	**0.147** (0.210)	**1.198***** (0.337)
Software Intensity	**-0.280***** (0.050)	**-0.512***** (0.096)	**-0.215***** (0.068)	**-0.286***** (0.069)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	1566	270	499	797
Adj R-Squared	0.236	0.401	0.154	0.354

The software intensity is based on the *share of citations directed to prior software patents*. * p<0.10, ** p<0.05, *** p<0.01

C.4. Separate regressions: U.S. and Non-U.S.

In Section C.4, based on the same rationale from Section B.5, we conducted separate Tobin's Q regressions using two subsamples; U.S. firms and non-U.S. firms. Table C.4.1 and Table C.4.2 report Tobin's Q regression results using U.S. firms only and non-U.S. firms only, respectively. As reported in both tables, for both U.S. firms and non-U.S. firms, the firms with higher software intensity have become increasingly rewarded over time by stock market investors with higher market valuations. Again, analogous to Section B.5, this suggests that our regression results are unlikely driven by U.S. firms only that are relatively more software intensive than non-U.S. firms.

Table C.4.1. U.S. firms only

lnQ	**Full Sample** NLS	**1981-1988** NLS	**1989-1996** NLS	**1997-2005** NLS
RD/Assets	**-0.127** (0.080)	**0.113** (0.258)	**-0.295***** (0.067)	**0.492***** (0.139)
RD/Assets * Software Intensity	**0.272**** (0.127)	**-0.100** (0.339)	**0.105** (0.122)	**1.711***** (0.525)
Software Intensity	**-0.246***** (0.044)	**-0.642***** (0.076)	**-0.234***** (0.059)	**-0.237***** (0.084)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	1207	261	362	584
Adj R-Squared	0.553	0.648	0.505	0.683

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

Table C.4.2. Non-U.S. firms only

lnQ	**Full Sample** NLS	**1981-1988** NLS	**1989-1996** NLS	**1997-2005** NLS
RD/Assets	**-0.078** (0.074)	**0.164** (0.368)	**0.096** (0.134)	**-0.267***** (0.093)
RD/Assets * Software Intensity	**1.708***** (0.439)	**-0.172** (0.658)	**1.758**** (0.760)	**2.079***** (0.452)
Software Intensity	**-0.399***** (0.106)	**0.234** (0.229)	**-0.514***** (0.162)	**-0.524***** (0.113)
Industry Dummies	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes
Number of Obs	1081	104	357	620
Adj R-Squared	0.405	0.487	0.468	0.372

The software intensity is based on the *share of software patents*. Regression results using alternative software intensity (share of citations directed to prior software patents) are qualitatively the same. Those regression results are available from the authors upon request. * p<0.10, ** p<0.05, *** p<0.01

D. Share of Software Intensive Firms: U.S. firms vs. Non-U.S. firms

This section supports human resource constraints hypothesis discussed in the main paper (V. Discussion). The purpose of this section is to show that firms headquarted in the United States are disproportionally more software intensive than non-U.S. firms. Figure D.1 demonstrates what percentages of firms from the United States and other countries are categorized as software intensive firms when we use the median software intensity as a threshold: (1) the above-median software intensity firms and (2) the below-median software intensity firms. Figure D.2 and D.3 use 75th percentile and 90th percentile as thresholds, respectively. Figure D.1 illustrates that 56% of U.S. firms and 44% of Non-U.S. firms belong to the group of the above-median software intensity firms.[2] The difference expands in Figure D.2 and Figure D.3. According to Figure D.2, 30% of U.S. firms are above 75th percentile. However, only 19% of Non-U.S. firms are above 75th percentile. Figure D.3 also shows similar (and larger) difference between U.S. firms and Non-U.S. firms – 15% of U.S. firms are above 90th percentile and only 4% of Non-U.S. firms are above the threshold.

The software intensity based on the share of software patents is used to draw the figures. The figures using another software intensity based on the share of citations directed to prior software patents are very similar to the figures in this section. Those figures are available from the authors upon request.

[2] If U.S. firms are not disproportionally more software intensive than Non-U.S. firms, then half of U.S. firms should be included in the group of the above-median software intensity firms.

Figure D.1. Above-median

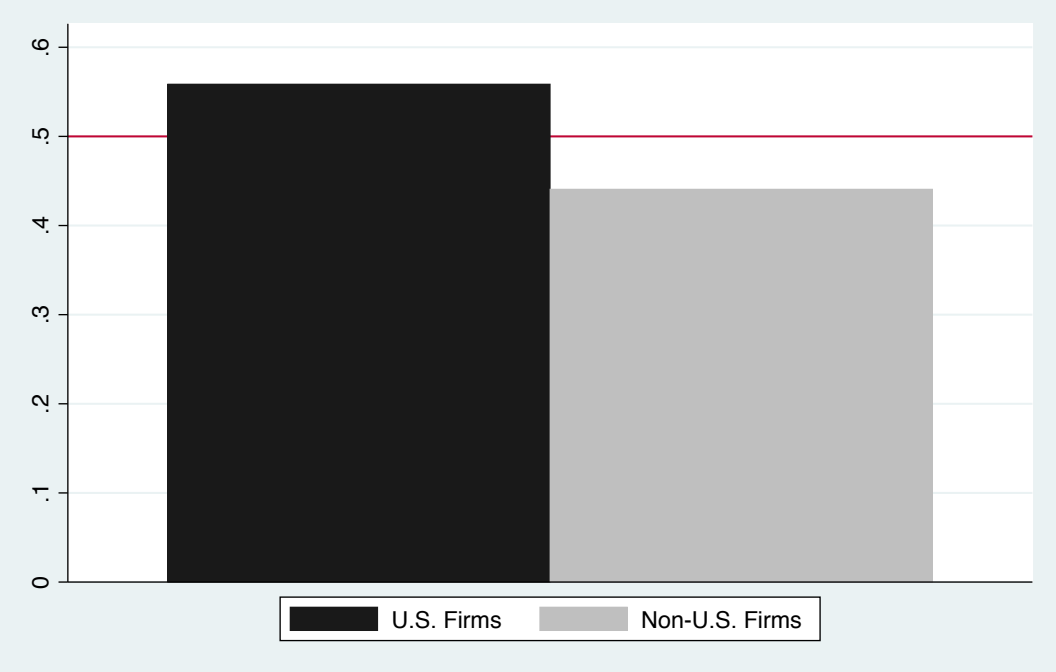
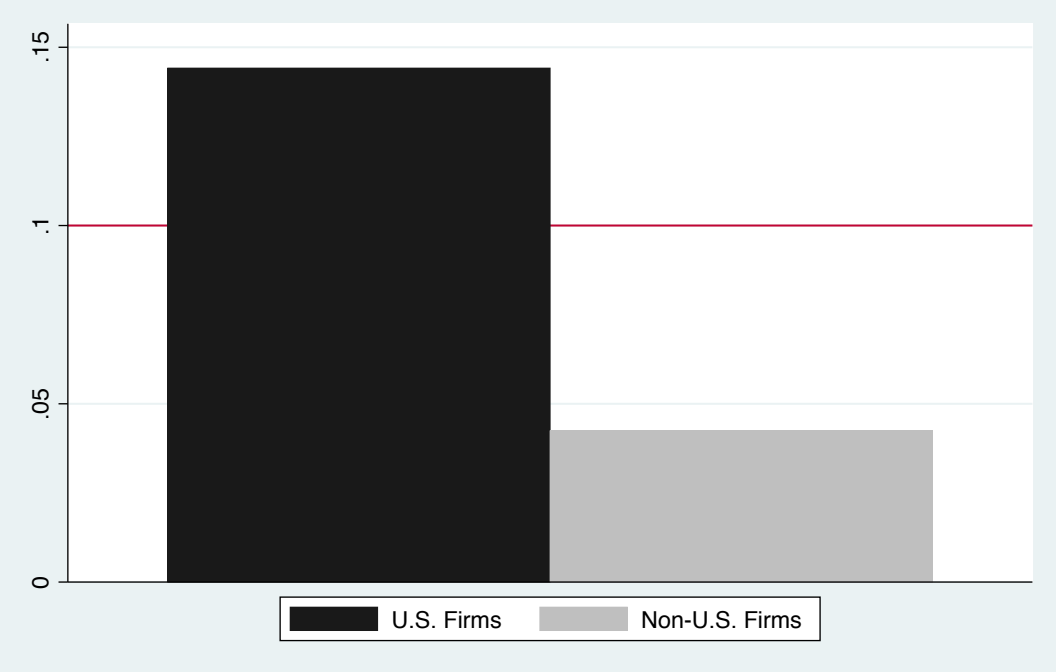


Figure D.2. Above-75th percentile

Figure D.3. Above-90th percentile

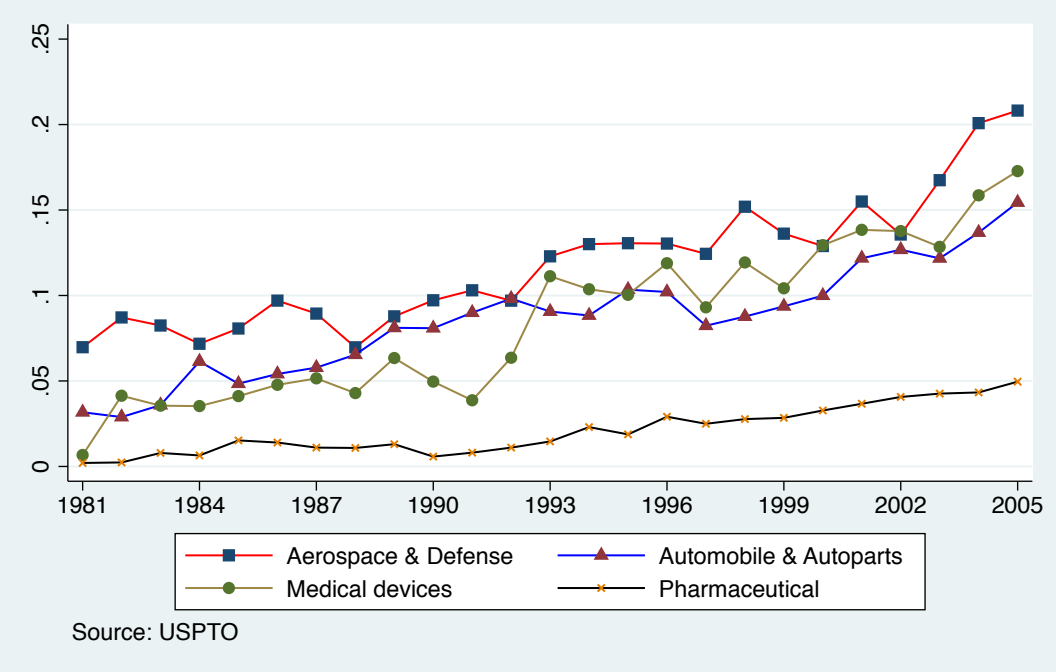


E. Software Intensity of Patent Portfolios of Firms: By Industry

E.1. Share of software patents (Grant years: 1981-2005)

In the main paper, Figure E.1 shows that firms in our sample have been increasingly to produce software patents over time. As expected, this software intensity metric has increased considerably in all four industries. The aerospace and defense industry seems to be the most software-intensive industry followed by medical devices, automobiles and auto parts, and pharmaceuticals.

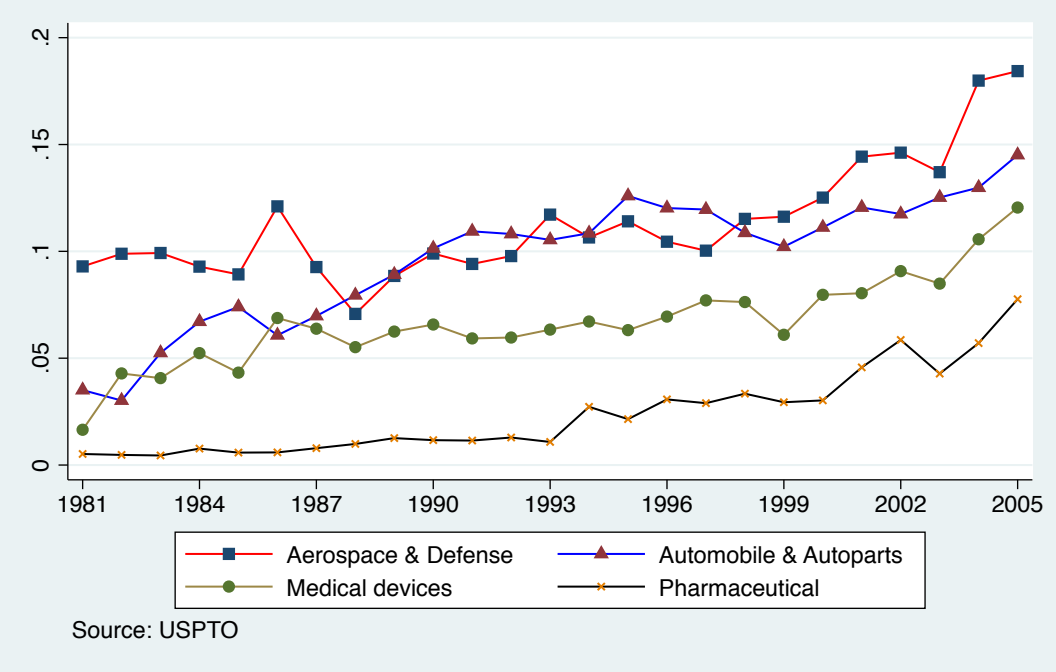
Figure E.1. Share of software patents (Grant years: 1981-2005)



E.2. Share of citations directed at software patents

Figure E.1 shows that patents generated by the firms in our sample have been increasingly to depend on software technology over time. There is substantial variation across industries. Specifically, we find aerospace and defense industry to be the most software-intensive industry, followed by automobiles and auto parts, medical devices, and pharmaceuticals. The differences across industries in absolute terms have not decreased over time.[3] However, these differences across industries in relative terms have decreased as innovation in all of our sample industries has become increasingly reliant on software as an input into the production of new-patented inventions.[4]

Figure E.2. Share of citations directed at software patents (Grant years of citing patents: 1981-2005)



[3] For example, the difference between the aerospace and defense and pharmaceutical industries in 1990 was 9% (10% - 1%). The deference became 10% (18% - 8%) in 2005.

[4] Let's compare the differences between the aerospace and defense and pharmaceutical industries in 1990 and 2005. The share increased about twofold in the aerospace and defense from 1990 to 2005 (10% to 18%). However, the share increased eightfold in the pharmaceutical during the period (1% to 8%).

F. ICT Education and Immigration Flows

F.1. Description of Data

The data which describe available flows and stocks of software-skilled human capital comprise of three components: (a) annual graduates with software-related degrees, (b) annual immigration flows of immigrants in software-related occupations, and (c) annual assignment of software-related offshore talent from India. The availability and degree of official disaggregation of these data vary significantly among countries and time periods. It was therefore necessary for us to make several assumptions about the structure and evolution of various data series in order to be able to construct longitudinal time series for all of the countries included in our analysis (i.e. USA, UK, Germany, France, and Japan). This Appendix describes the data construction process for each country in some detail.

The United States: Data on cohorts of software-related graduates come from NSF's annual Survey of Earned Doctorates and the biannual National Survey of Recent College Graduates. Due to the biannual nature of the latter survey, we used linear interpolation in order to impute numbers for missing years. The annual data on flows of immigrants in software-related occupation come from USCIS annual reports on "Characteristics of Specialty Occupation Workers (H-1B)". Data for years prior to 1999 come from a previous paper written by the authors. Data on available offshore software-skilled labor in India come from NASSCOM's annual reports, which contain a breakdown of Indian BPO industry's exports by destination.

The United Kingdom: Data on annual cohorts of software-related graduates come from statistical tables published by UK's Higher Education Statistics Agency. Data on annual flows of migrants in software-related occupations come from Prof. John Salt's periodic reports on the state of immigration in the United Kingdom. Linear interpolation was used to impute immigration flows data for years for which official statistics were not available. Data on available offshore software-skilled labor in India come from NASSCOM's annual reports, which contain a breakdown of Indian BPO industry's exports by destination.

Germany: Data on annual cohorts of software-related graduates come from the Federal Statistical Office of Germany. Immigration data come from annual reports published by Germany's Federal Office for Migration and Refugees. Data on available offshore software-skilled labor in India come from NASSCOM's annual reports, which contain a breakdown of Indian BPO industry's exports by destination.

France: Data on annual cohorts of software-related graduates come from OECD's statistical tables on "tertiary-type A" education. Due to the unavailability of data for years before 1999 and after 2010, linear extrapolation was used to populate the missing data fields. Data on annual flows of immigrants come from various reports published by the French Ministry of the Interior. Due to sparse information on immigration for some years in the sample, linear extrapolation and interpolation were used to impute immigration flows data for years for which official statistics were not available. Data on available offshore software-skilled labor in India come from NASSCOM's annual reports, which contain a breakdown of Indian BPO industry's exports by destination.

Japan: Data on graduates are taken from the Japanese Ministry of Education, Sports, and Welfare's Basic School Survey. Data on newly registered foreign workers come from the Annual Report of Statistics on Legal Migrants published by the Japanese Ministry of Justice. Data on available offshore software-skilled labor in India come from NASSCOM's annual reports, which contain a breakdown of Indian BPO industry's exports by destination.

From Flows to Stocks: In order to calculate stocks of available software-related labor for the above countries at various points in time, we employed a perpetual inventory method. We made the following assumptions in order to be able to convert flow measures into stocks: (1) flow of new graduates in the 10 years prior to 1995 equaled that in the year 1995; (2) 80% of new graduates go into software-related employment every year; (3) every year 10% of existing software-related employees move out of software-related employment for good; (4) 10% of existing software-related immigrants leave the country and/or software-related employment each year; (5) 40% of total H1-B issued in a year are new petitions for initial employment; (6) flow of new ICT immigrants in the 10 years prior to 1995 equaled that in the year 1995.

F.2. Supplementary Figures: Flow

We present Figure 5 in the main text. The figure clearly show that the United States has significantly larger "flow" of ICT labor force than the other four major countries. The "flow" of Figure 5 is the sum of ICT graduates, ICT immigrants, and offshoring to India. It is worth showing figures with each component of "total flow".

Figure F.2.1. ICT Graduates

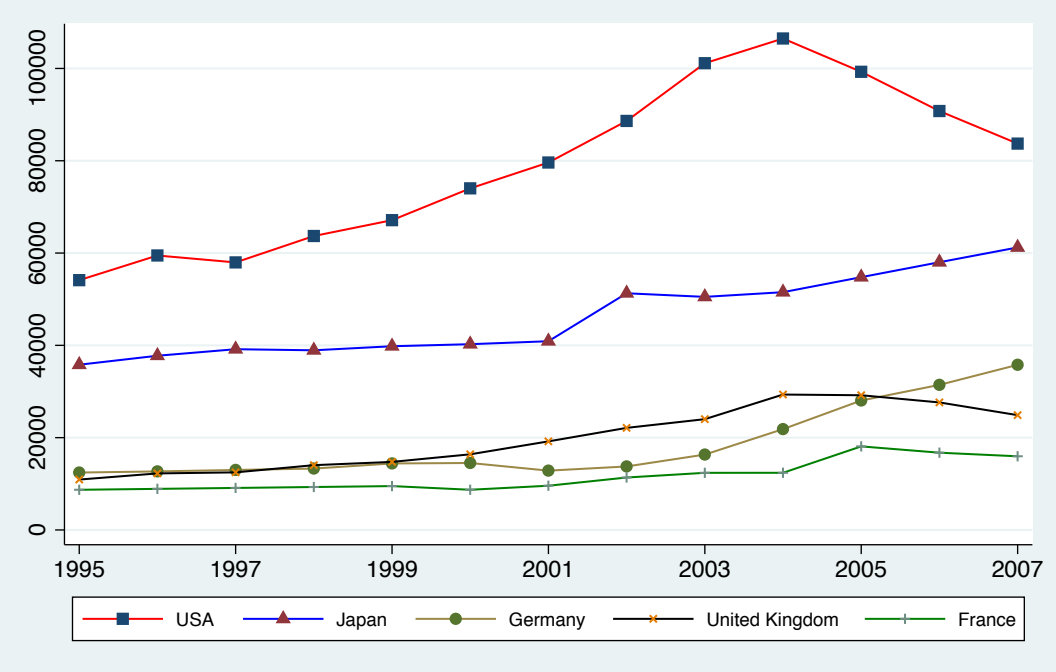


Figure F.2.2. ICT Graduates and ICT Immigrants

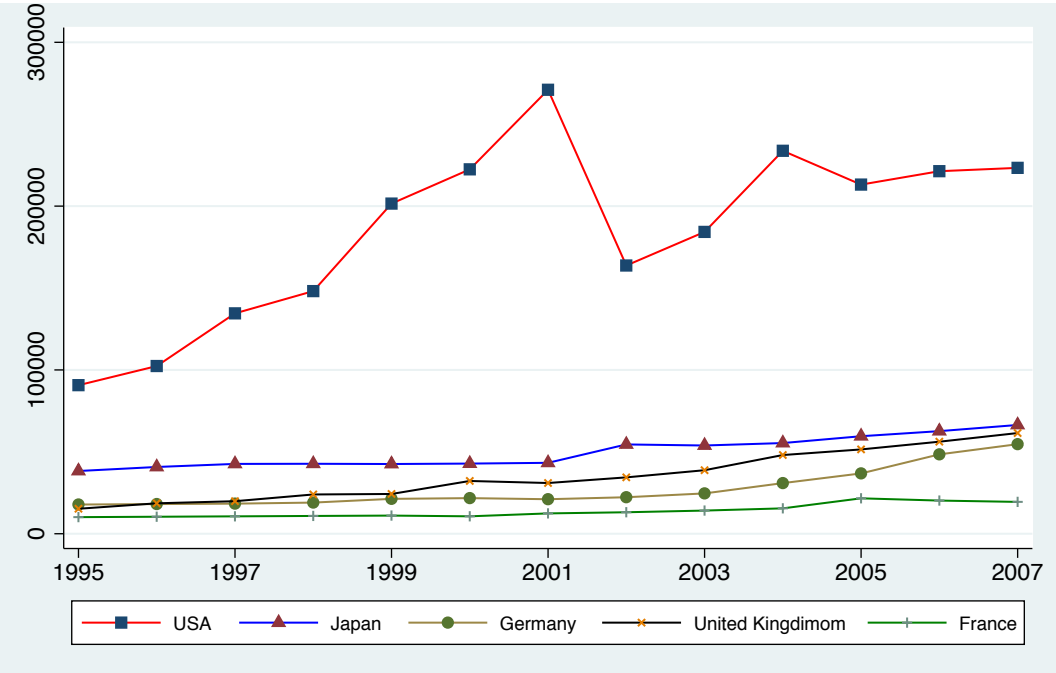
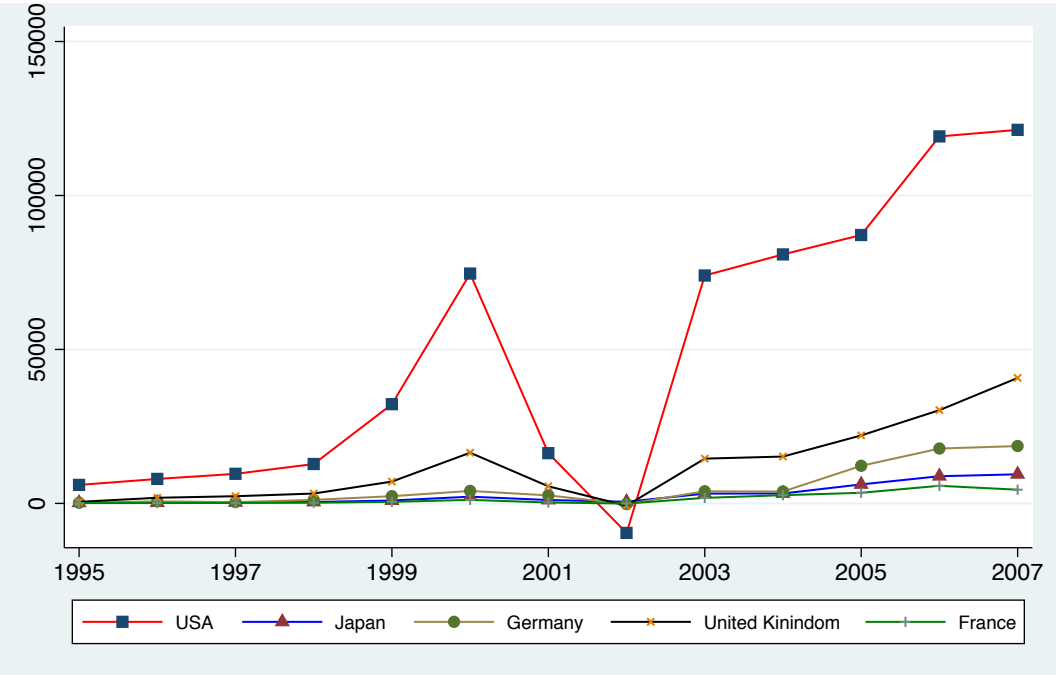


Figure F.2.3. Offshoring to India



F.3. Supplementary Figures: Stock

This section provides "stock" of ICT workers in those five countries.

Figure F.3.1. Total

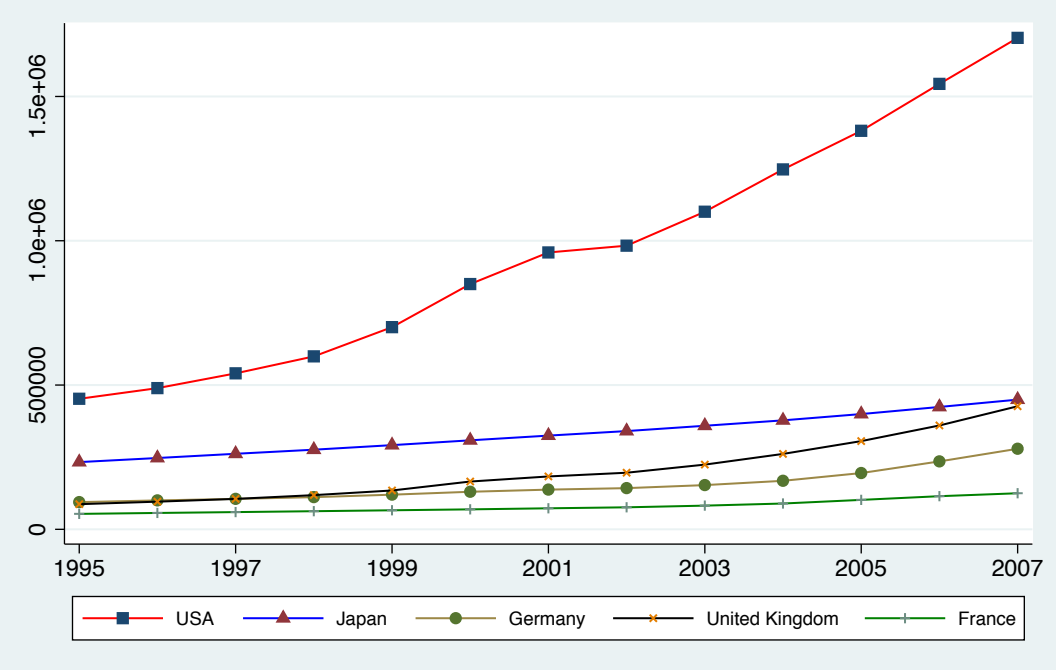


Figure F.3.2. ICT Graduates

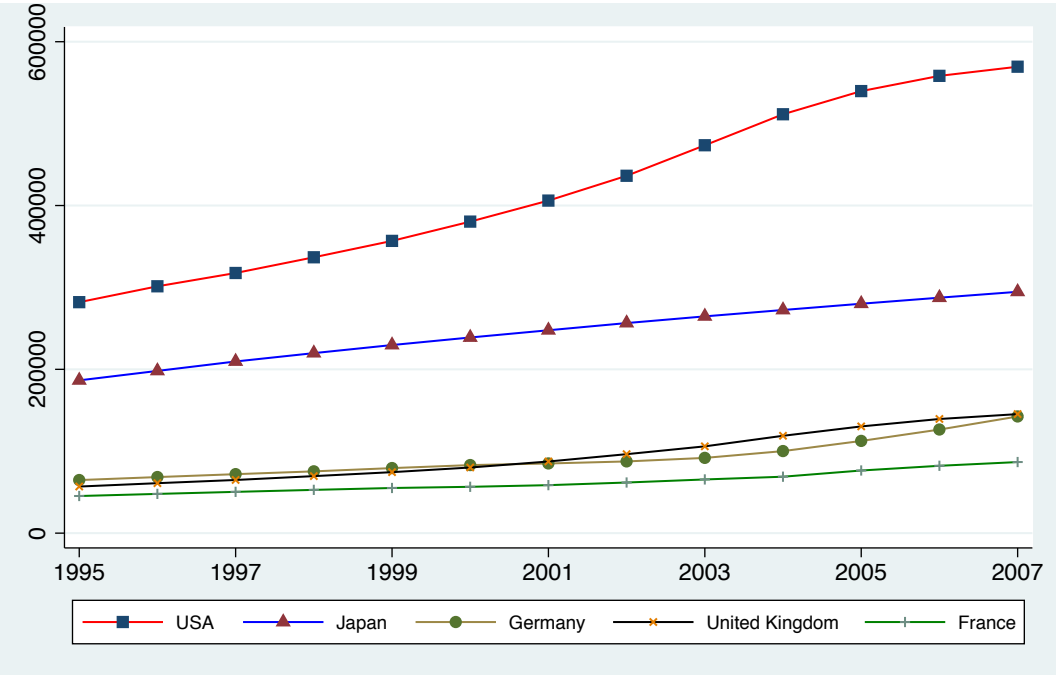


Figure F.3.3. ICT Graduates and ICT Immigrants

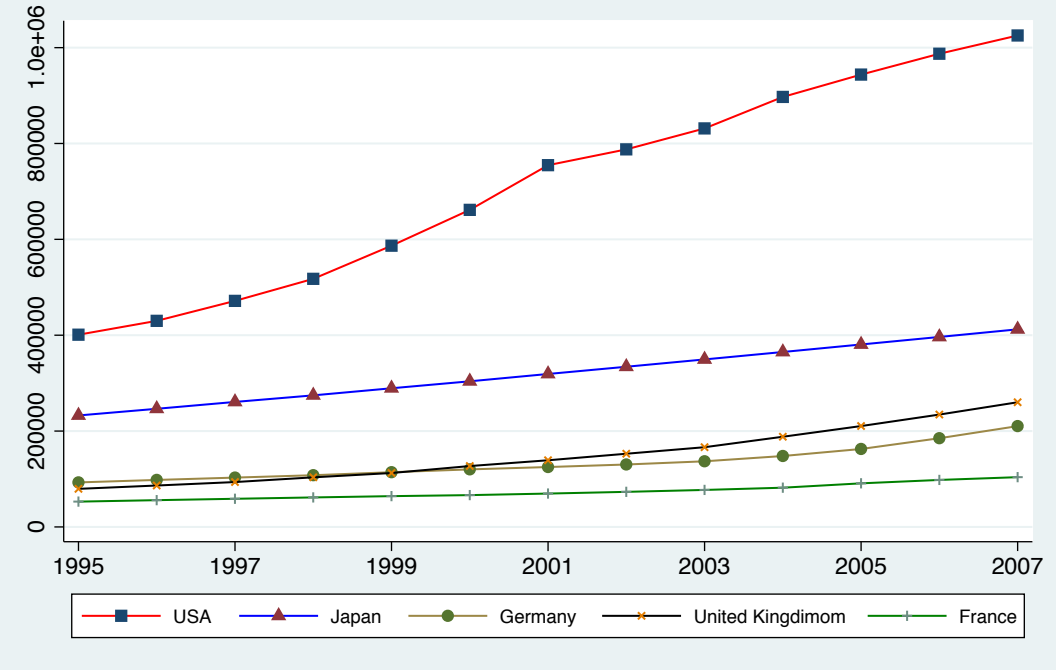


Figure F.3.4. Offshoring to India