Can Patent Family Size and Composition Signal Patent Value?

Francois P. Kabore, Georgetown U. and CERAP

Walter G. Park, American U.

October 2014

Motivation

- New Method of Valuing Patents
- Why Patent Valuation Important?
  - Firm Decision-Making (Licensing, Litigation)
  - Policy-Making and Regulation
  - Applications: creating indicators of innovation activities; conducting studies on productivity, R&D, technology transfer, etc.
  - Problem: no institutional markets (as for stocks and bonds)

Revealed Value

- Patent Renewal
  Pakes ‘86, Schankerman & Pakes ‘86, Lanjouw et al. ‘98
- Patent Citations
  Trajtenberg ’90, Jaffe & Trajtenberg ‘02, Hall et al. ‘05
- Opposition
  Harhoff & Reitzig ‘04
- Patent Family Size
  Putnam ‘96, Lanjouw & Schankerrman ‘04, Harhoff et al. ‘04

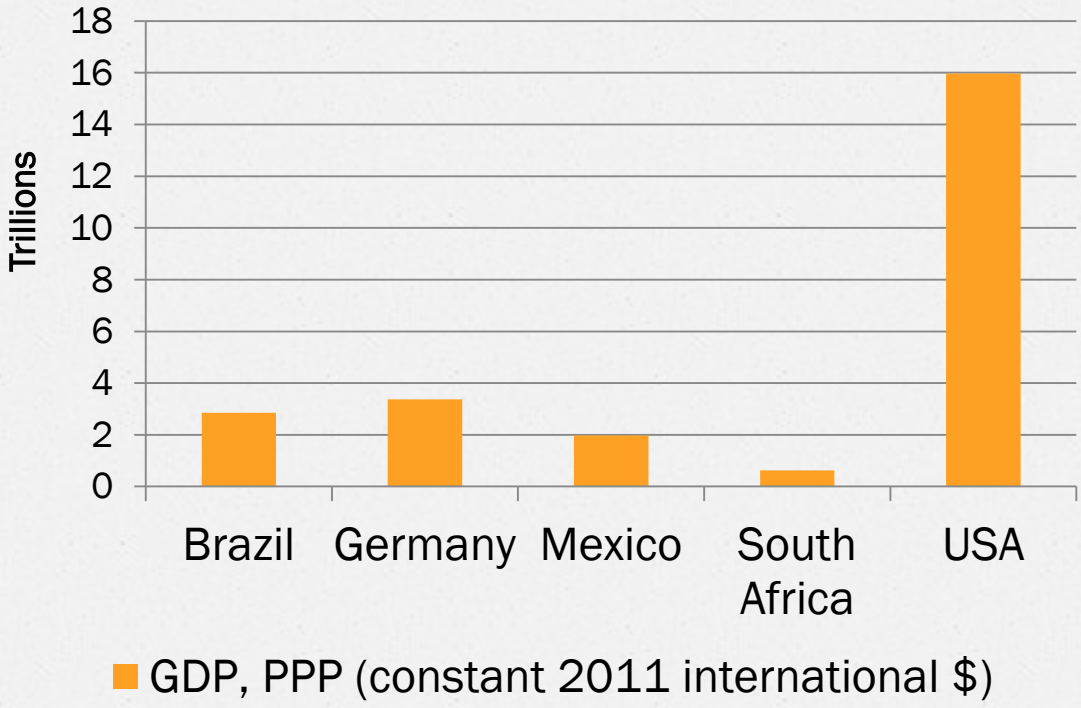
“Ex Post” Measures

Criticisms

- Citations
  - mixed empirical support (Bessen ‘08)
  - degree of importance not accounted for
  - other motivations to cite
- Renewals
  - influenced by cost
- Family Size
  - markets vary in attractiveness

Our Approach

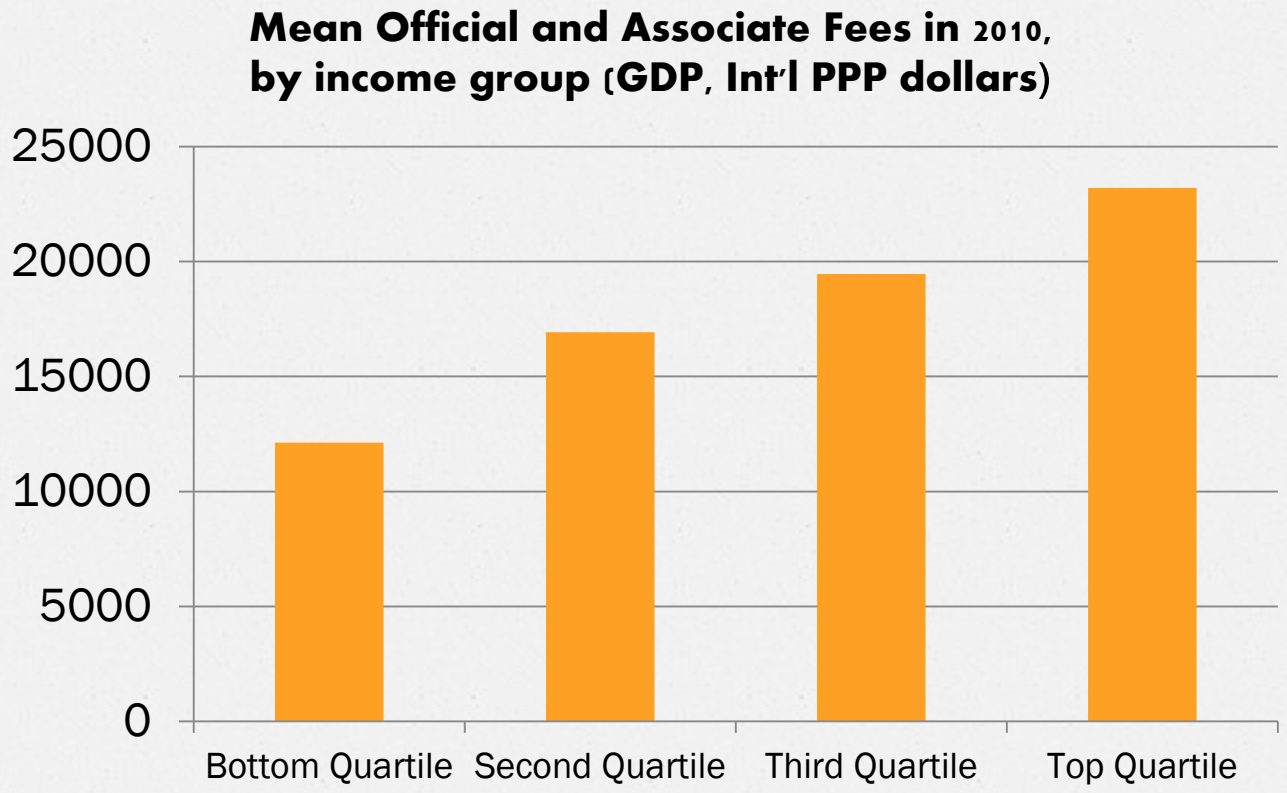
- GDP-weighted Family Size



Why Market Size matters

- Expectation of higher returns
- Imitation Risk
- Competing Innovations
- Cost of procurement & enforcement

Cost of Patenting by Market Size



Indicators

- Family Size

$$V_{jt} = \sum_{n=1}^{N} I_{nt}$$

- GDP-weighted Family Size

$$V_{jt} = \sum_{n=1}^{N} \omega_{nt} I_{nt}$$

where j indexes the priority patent and I_{nt} = 1 if country n is in the patent family at publication time t.

Data

- PATSTAT and Patent Renewal Status (PRS)
  - Matched to obtain ‘duration’
  - Worked backwards to obtain forward citations
- Five source countries: France, Germany, Japan, UK, and USA, and all destinations.
- Sample period: 1980 – 2001.
- 1,208,345 observations. All IPC (A – H) fields.

Table 1. Sample Statistics of Patent Value Indicators by Source Country

Source Country	Statistic	Duration	Family Size (Count)	Family Size weighted by GDP	Family Size weighted by Private GDP	Citations received within first 5 years	Citations received within first 8 years
All 5	Mean	8.79	1.27	5.9E+12	4.9E+12	0.80	1.16
	Min	0	1	6.0E+10	4.4E+10	0	0
	Max	20	29	2.4E+13	2.1E+13	66	106
France	Mean	8.85	1.42	4.5E+12	3.8E+12	0.72	1.05
	Min	1	1	1.1E+11	9.9E+10	0	0
	Max	20	6	2.4E+13	2.1E+13	50	60
Germany	Mean	8.78	1.28	3.9E+12	3.2E+12	0.54	0.79
	Min	0	1	6.0E+10	4.4E+10	0	0
	Max	20	8	2.4E+13	2.1E+13	42	75
Japan	Mean	9.97	1.30	7.4E+12	6.2E+12	1.24	1.71
	Min	0	1	1.4E+11	1.2E+11	0	0
	Max	20	29	2.4E+13	2.1E+13	51	75
U.K.	Mean	8.85	1.48	5.8E+12	4.8E+12	0.90	1.35
	Min	1	1	1.4E+11	1.1E+11	0	0
	Max	20	11	2.4E+13	2.1E+13	66	106
U.S.A.	Mean	8.29	1.21	6.6E+12	5.6E+12	0.76	1.13
	Min	0	1	6.0E+10	4.4E+10	0	0
	Max	20	16	2.4E+13	2.1E+13	62	103

Figure 2. Percentage Distribution of Patents by Duration (years)

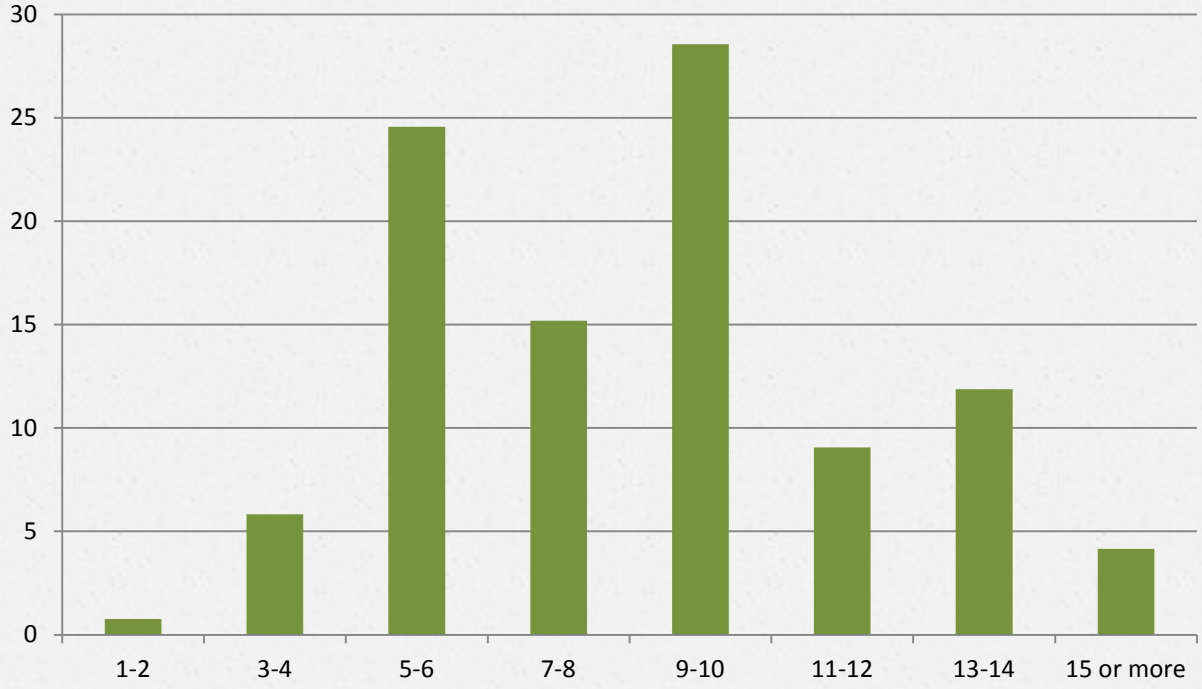


Table 3. Patent Valuation Indicators Grouped by Duration Rank and Family Value Rank

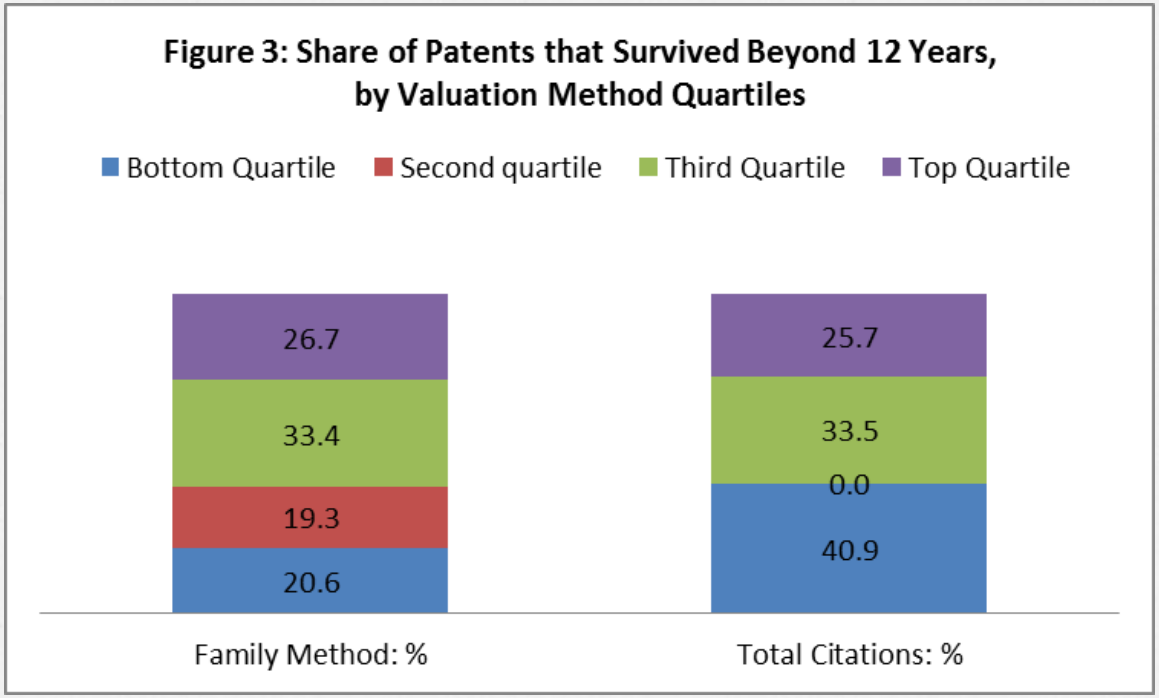
A. Grouped by Duration Quartiles

Percentile of Duration	Percent of Observations	statistic	Duration	Family Size (Count)	Family Size weighted by GDP	Family Size weighted by Private GDP	Citations received within 5 years	Citations received within 8 years
≤ 25%	31.1	Mean	5.16	1.22	5.03E+12	4.20E+12	0.51	0.77
25%< & ≤ 50%	26.0	Mean	8.05	1.24	5.50E+12	4.62E+12	0.62	0.90
50%< & ≤ 75%	24.0	Mean	10.26	1.22	6.54E+12	5.51E+12	0.90	1.29
> 75%	18.9	Mean	13.92	1.45	7.05E+12	5.91E+12	1.39	2.00
Total	100.0	Mean	8.79	1.27	5.89E+12	4.94E+12	0.80	1.16

B. Grouped by GDP-weighted Family Size Quartiles

Percentile of Family Size, GDP-weighted	Percent of Observations	statistic	Duration	Family Size (Count)	Family Size weighted by GDP	Family Size weighted by Private GDP	Citations received within 5 years	Citations received within 8 years
≤ 25%	25.2	Mean	7.70	1.09	1.46E+12	1.18E+12	0.28	0.41
25%< & ≤ 50%	25.3	Mean	8.15	1.37	4.84E+12	4.02E+12	0.87	1.37
50%< & ≤ 75%	27.1	Mean	9.68	1.20	7.09E+12	5.96E+12	0.96	1.37
> 75%	22.4	Mean	9.64	1.43	1.06E+13	9.00E+12	1.12	1.52
Total	100.0	Mean	8.79	1.27	5.89E+12	4.94E+12	0.80	1.16

Figure 3: Share of Patents that Survived Beyond 12 Years, by Valuation Method Quartiles



Methodology

"Horse Race" Test:

- Split Sample in Half (using stratified sampling)
  A. Call them z = 0 sample and z = 1 sample
  B. Estimate Model of Duration using z = 0 sample, using various valuation indicators
  C. Forecast Duration with the z = 1 sample, using various valuation indicators. *Out-of-sample*.
  D. Switch: Estimate with z=1 and forecast with z=0.
  E. Determine which indicator best predicts duration.
  F. Repeat where dependent variable is "citations", and test which indicator best explains **citations**.

Forecast Accuracy

- Predicted Equation: $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$
- RMSPE = $\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y}_i - Y_i}{Y_i}\right)^2}$
- Theil's U = $\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i\right)^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i\right)^2}}$

Results

Table 4. Joint Effects of Patent Value Indicators on Duration of Patent Life

Dependent Variable: ln (Duration)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample:	Full Sample	Full Sample	Full Sample	Full Sample	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z =1	Split Sample z =1	Split Sample z =1	Split Sample z =1
ln (family size)	0.025***	0.027***	0.026***	0.028***	0.025***	0.028***	0.026***	0.028***	0.025***	0.027***	0.025***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln (GDP-weight family size)	0.100***	0.099***			0.108***	0.106***			0.090***	0.089***		
	(0.001)	(0.001)			(0.001)	(0.001)			(0.001)	(0.001)		
ln (priv GDP-wgt family size)			0.097***	0.096***			0.105***	0.103***			0.087***	0.086***
			(0.000)	(0.000)			(0.001)	(0.001)			(0.001)	(0.001)
ln(citation counts 5 years)	0.052***		0.052***		0.056***		0.056***		0.047***		0.047***	
	(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
ln(citation counts 8 years)		0.048***		0.049***		0.052***		0.053***		0.043***		0.044***
		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)
Observations	1,208,345	1,208,345	1,208,345	1,208,345	654,043	654,043	654,043	654,043	554,302	554,302	554,302	554,302

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Coefficients of all the fixed effects (country, year, technology) and interaction terms (tech x year, tech x country, country x year) are omitted to conserve space

Results

Table 5. Regression Model for Predicting Duration, using the z = 0 Split Sample

Dependent Variable: ln (Duration)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample used to create forecasting equation:	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z =1	Split Sample z =1	Split Sample z =1	Split Sample z =1	Split Sample z =1	Split Sample z =1
ln (family size)		0.131***						0.104***				
		(0.002)						(0.002)				
ln (GDP-weight family size)			0.121***						0.101***			
			(0.001)						(0.001)			
ln (priv GDP-wgt family size)				0.118***						0.098***		
				(0.001)						(0.001)		
ln(citation counts 5 years)					0.096***						0.077***	
					(0.001)						(0.001)	
ln(citation counts 8 years)						0.091***						0.072***
						(0.001)						(0.001)
Observations	654,043	654,043	654,043	654,043	654,043	654,043	554,302	554,302	554,302	554,302	554,302	554,302

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Coefficients of all the fixed effects (country, year, technology) and interaction terms (tech x year, tech x country, country x year) are omitted to conserve space

Table 6. Measures of Forecast Accuracy

i. Using z = 0 sample to predict Duration in the z = 1 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Citations, first 5 Years	(6) Citations, first 8 years
RMSPE	0.234	0.231	0.213	0.214	0.227	0.226
-- Ratio to Col. (3)	1.099	1.085	1.000	1.005	1.066	1.061
Theil U	0.081	0.080	0.076	0.076	0.080	0.080
-- Ratio to Col. (3)	1.061	1.048	1.000	1.000	1.048	1.048

ii. Using z = 1 sample to predict Duration in the z = 0 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Citations, first 5 Years	(6) Citations, first 8 years
RMSPE	0.267	0.264	0.244	0.245	0.260	0.259
-- Ratio to Col. (3)	1.094	1.082	1.000	1.004	1.066	1.061
Theil U	0.088	0.087	0.081	0.081	0.086	0.085
-- Ratio to Col. (3)	1.085	1.073	1.000	1.000	1.061	1.048

Results

Table 7. Regression Model for Predicting Citations using Selected Patent Value Indicators

Dependent Variable:	Citations 5 Years	Citations 5 Years	Citations 5 Years	Citations 5 Years	Citations 5 Years	Citations 5 Years	Citations 8 Years	Citations 8 Years	Citations 8 Years	Citations 8 Years	Citations 8 Years	Citations 8 Years
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample Used:	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 1	Split Sample z = 1	Split Sample z = 1	Split Sample z = 0	Split Sample z = 0	Split Sample z = 0	Split Sample z = 1	Split Sample z = 1	Split Sample z = 1
ln (family size)	0.174***			0.135***			0.169***			0.124***		
	(0.003)			(0.003)			(0.003)			(0.003)		
ln (GDP-weight family size)		0.169***			0.160***			0.207***			0.198***	
		(0.001)			(0.001)			(0.001)			(0.001)	
ln (patent duration)			0.204***			0.182***			0.244***			0.216***
			(0.002)			(0.002)			(0.002)			(0.002)
Observations	654,043	654,043	654,043	554,302	554,302	554,302	654,043	654,043	654,043	554,302	554,302	554,302

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Coefficients of all the fixed effects (country, year, technology) and interaction terms (tech x year, tech x country, country x year) are omitted to conserve space

Table 9. Measures of Forecast Accuracy for Predicting Citations using a Truncated Sample

Case I. Predicting Patent Citations in the first 5 years of a patent's life using a sample of patents whose duration exceeded 5 years

i. Using z = 0 sample to predict Citations5 in the z = 1 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Duration
RMSPE	0.503	0.509	0.490	0.490	0.506
-- Ratio to Col. (3)	1.027	1.038	1.000	1.000	1.033
Theil U	0.443	0.440	0.426	0.426	0.440
-- Ratio to Col. (3)	1.041	1.034	1.000	1.001	1.032

ii. Using z = 1 sample to predict Citations5 in the z = 0 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Duration
RMSPE	0.509	0.510	0.493	0.493	0.508
-- Ratio to Col. (3)	1.031	1.034	1.000	1.000	1.029
Theil U	0.448	0.443	0.427	0.428	0.442
-- Ratio to Col. (3)	1.047	1.036	1.000	1.001	1.034

Case II. Predicting Patent Citations in the first 8 years of a patent's life using a sample of patents whose duration exceeded 8 years

i. Using z = 0 sample to predict Citations8 in the z = 1 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Duration
RMSPE	0.445	0.450	0.442	0.442	0.449
-- Ratio to Col. (3)	1.006	1.018	1.000	0.999	1.016
Theil U	0.385	0.380	0.367	0.367	0.382
-- Ratio to Col. (3)	1.049	1.037	1.000	1.001	1.043

ii. Using z = 1 sample to predict Citations8 in the z = 0 Sample

Additional RHS Variable	(1) Base Case (no patent value indicators)	(2) Family Size	(3) GDP-weighted Family Size	(4) Private GDP-weighted Family Size	(5) Duration
RMSPE	0.447	0.450	0.443	0.442	0.449
-- Ratio to Col. (3)	1.011	1.016	1.000	1.000	1.014
Theil U	0.389	0.382	0.367	0.367	0.384
-- Ratio to Col. (3)	1.060	1.041	1.000	1.001	1.046

Discrete-time Survival Analysis

Table 10: Logistic Survival Analysis

Impact of patent value indicators on the probability of a patent lapse

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP-weighted Family Size	-0.0924***					-0.0029***				
	(0.001)					(0.0012)				
Private GDP-weighted Family Size		-0.091***					-0.0029***			
		(0.001)					(0.0012)			
Family Size (Raw count)			-0.261***					0.0518***		
			(0.004)					(0.049)		
Citation Counts (first) 5 years				-0.151***					0.0063***	
				(0.002)					(0.0017)	
Citation Counts (first) 8 years					-0.137***					0.0052***
					(0.001)					(0.0016)
Country Fixed Effects	Not Included	Not Included	Not Included	Not Included	Not Included	Included	Included	Included	Included	Included
Year Fixed Effects	Not Included	Not Included	Not Included	Not Included	Not Included	Included	Included	Included	Included	Included
Technology Fixed Effects	Not Included	Not Included	Not Included	Not Included	Not Included	Included	Included	Included	Included	Included
Observations	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465	10,616,465
LR Chi (2)	4130207.4	4303675.5	4125116.7	4130827.3	4131769.1	4297166.3	4297166.2	4297267.4	4297172.9	4297170.9

Applications (in progress)

- Country-wide measures:

$$V_t^i = \sum_{j=1}^{J(i,t)} V_{jt} = \sum_{j=1}^{J(i,t)} \left(\sum_{n=1}^{N} GDP_{nt} I_{nt}^j \right)$$

$$V_t = \sum_{i=1}^{5} V_t^i$$

Relation to R&D

All Five Source countries (US, UK, Germany, Japan, France) pooled