PATENT STATISTICS FOR INTERNATIONAL COMPARISONS AND ANALYSIS OF NARROW TECHNOLOGICAL FIELDS

Ivan Haščič, Jérôme Silva and Nick Johnstone

OECD Environment Directorate

Conference on Patent Statistics for Decision Makers
12-13 November 2013, Rio de Janeiro, Brazil

OECD
BETTER POLICIES FOR BETTER LIVES

Motivation

- Technological scope – narrow fields (e.g. many 'environmental' techs)
- Geographic scope – smaller innovators, countries with lower patenting activity (e.g. emerging/transition economies)

➢ How to use patent data 'reliably' in such contexts?

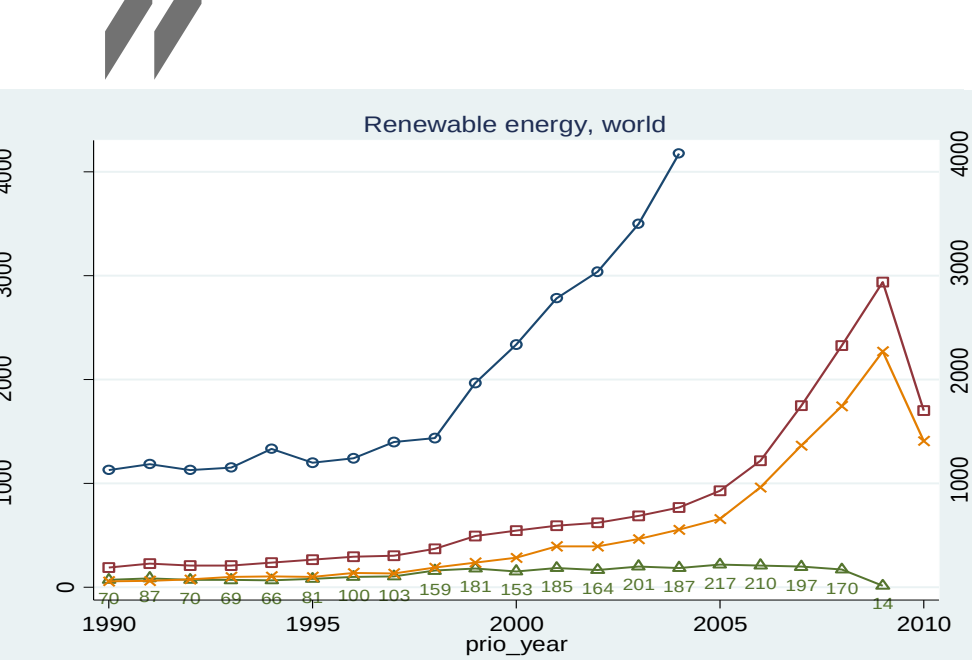
Outline

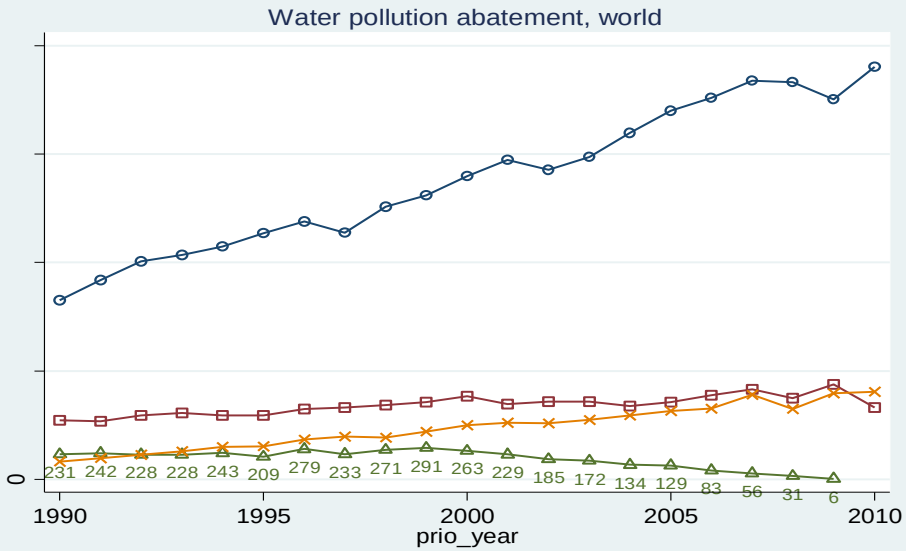
1) Adapting patent indicators to different contexts

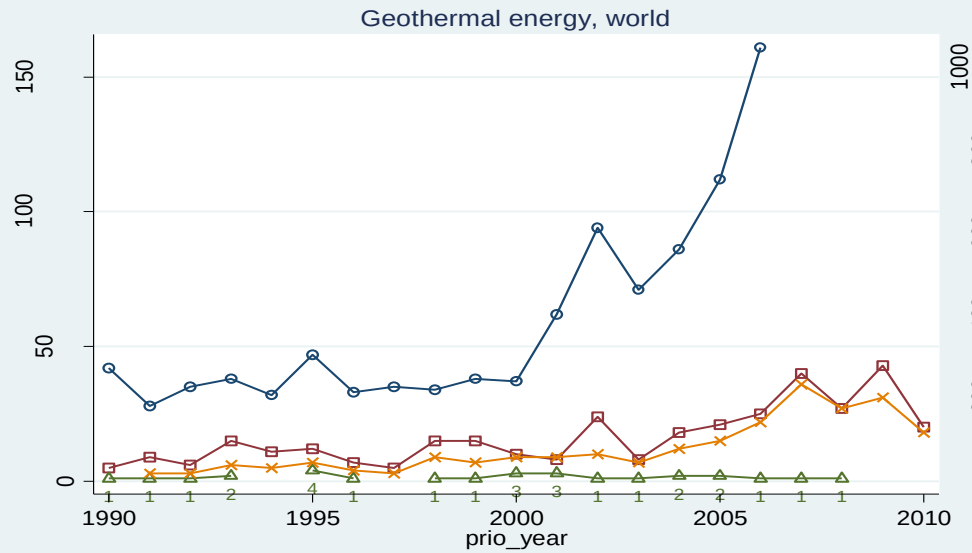
2) Idiosyncratic issues in construction and analysis of patent statistics
   a) Coverage
   b) Designation
   c) Missing info
   d) Identification

3) Implications for analysis

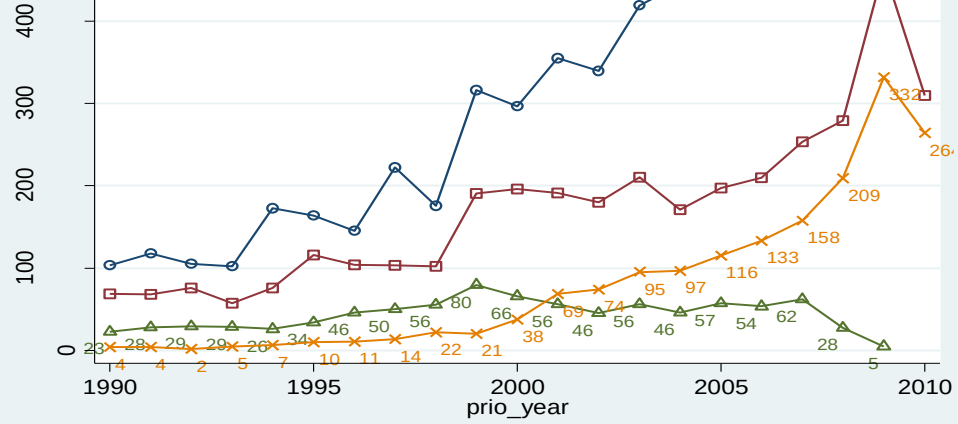
1) Indicator: technological context

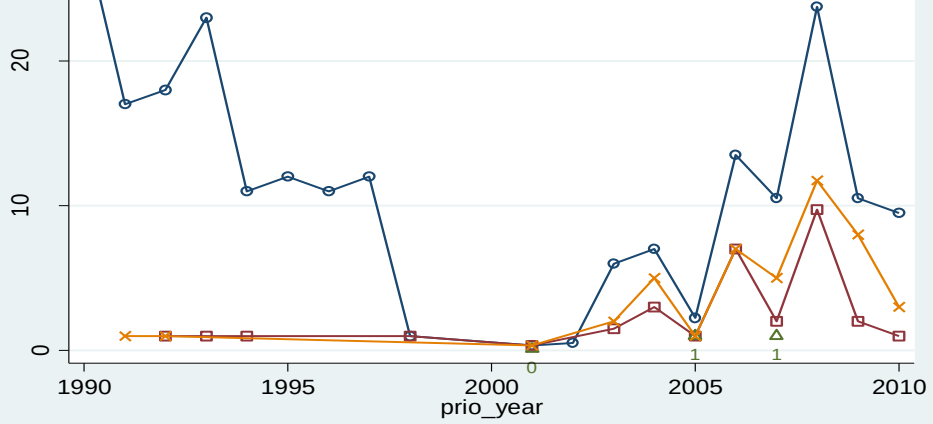


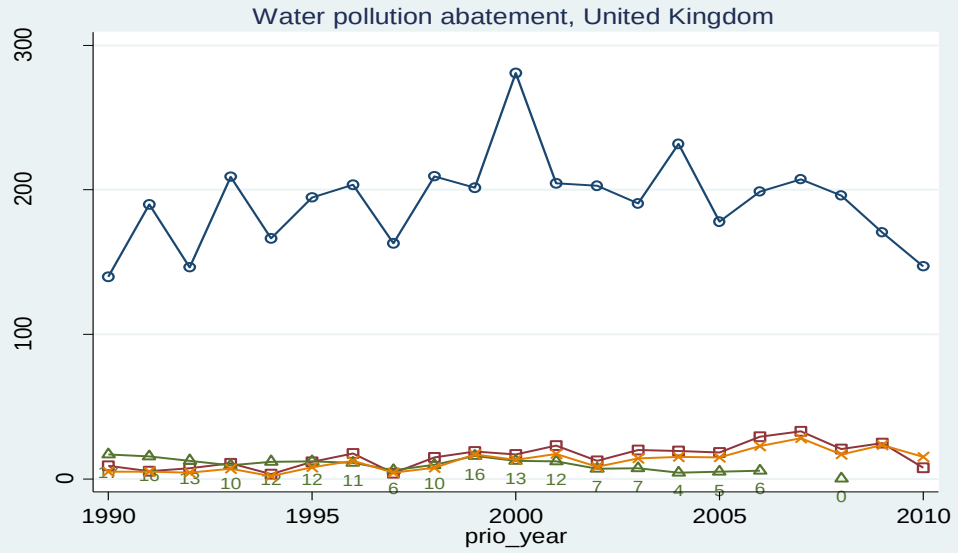


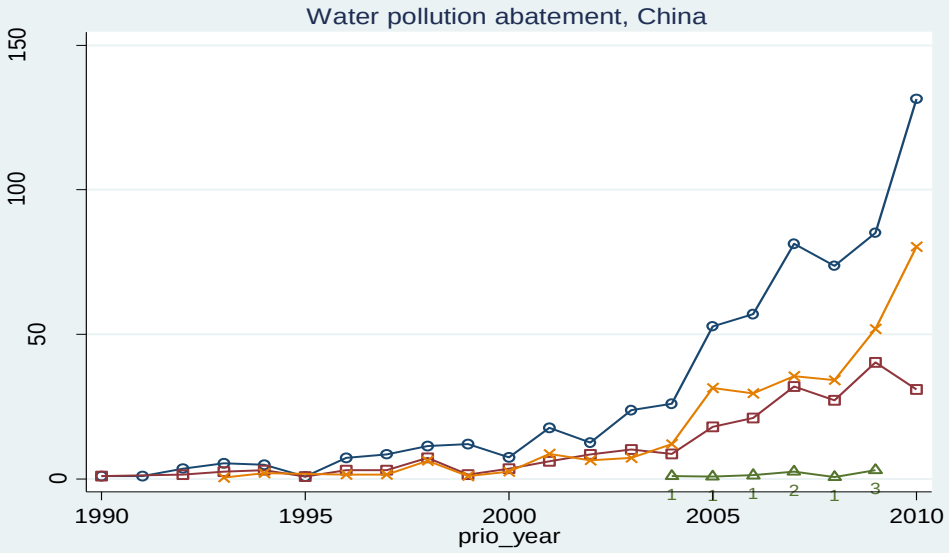


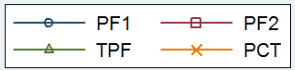
1) Indicator: country context



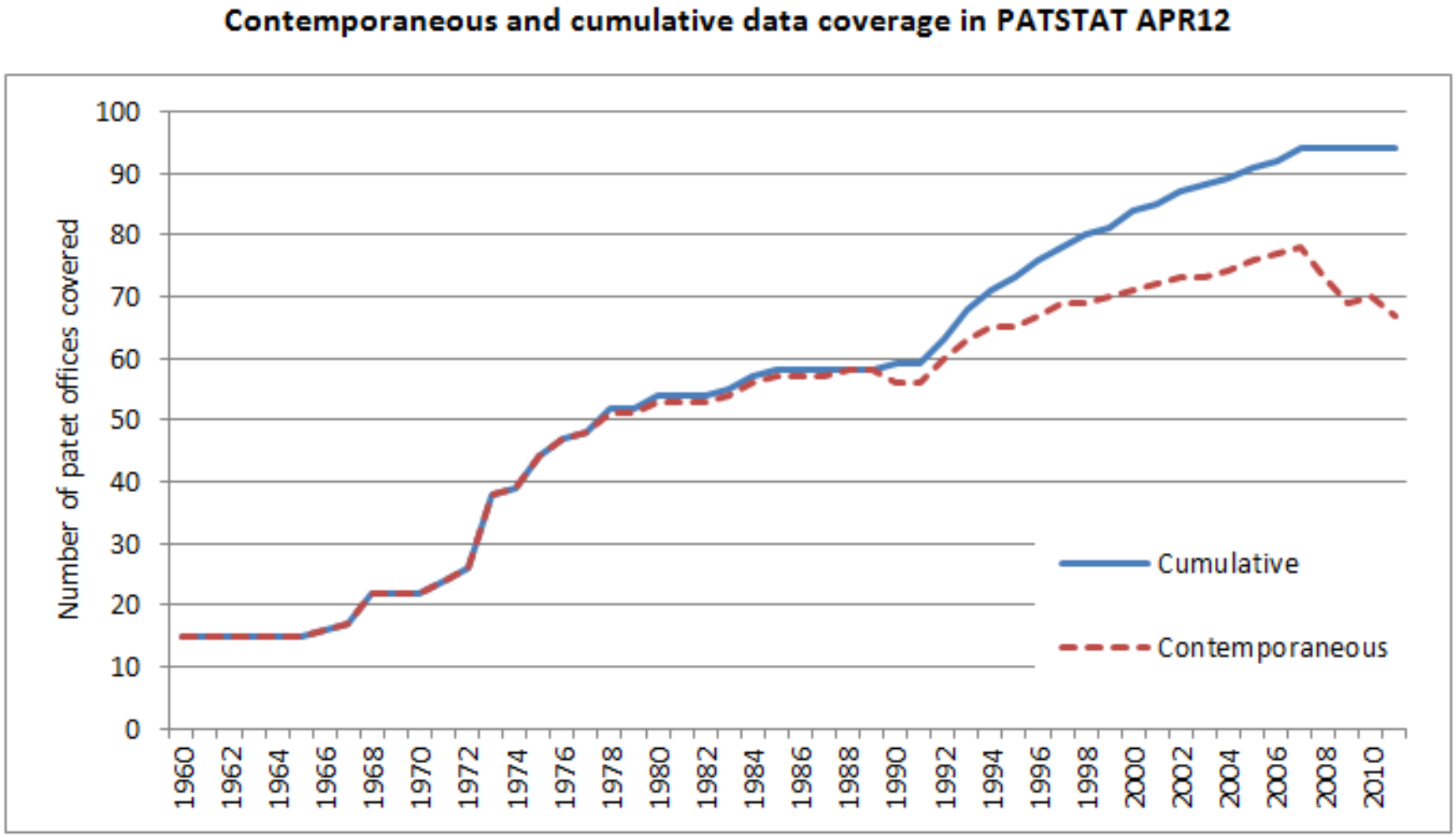








2a) Spatial and temporal coverage



2a) Spatial and temporal coverage

APR12 PATSTAT, 1980-2009:

- Complete coverage (=30): 39 offices (e.g. EP, IB, JP, US)
- Complete coverage (<30): 16 offices (e.g. DD, CS, YU)
- Partial coverage (<30): 40 offices (e.g. AR, IN, MA)
- No coverage with data on 19 application offices (e.g. TN, TH, AZ)
- No coverage with data on 129 inventor countries (e.g. VE, SA, IR)

2a) Spatial and temporal coverage

Country/office	Theoretical coverage 1980-2009		Empirical coverage in PATSTAT APR12 as Appln Authority	as Inventor Country
Korea	30	complete	1921667	1633271
Brazil	30	complete	408144	74490
Mexico	30	complete	176070	17390
Israel	30	complete	132758	127959
South Africa	30	complete	132370	23184
Hong Kong, China	30	complete	69446	23408
Turkey	30	complete	43447	27571
Egypt	30	complete	8474	2285
ARIPO	25.2	complete	5524	-
Russia	16.9	complete	479655	243024
Argentina	28.7	partial	59471	8662
Guatemala	27.6	partial	1010	212
India	27.4	partial	41207	73617
Singapore	26.6	partial	54160	30356
China	24.3	partial	3587728	2141368
Ecuador	20.0	partial	7098	678
Philippines	19.2	partial	14701	3050
Peru	17.3	partial	10752	1116
Morocco	16.2	partial	11621	1728
Colombia	14.9	partial	13458	3040

2a) Spatial and temporal coverage

Country/office	Theoretical coverage 1980-2009		Empirical coverage in PATSTAT APR12 as Appln Authority	as Inventor Country
Malawi	14.8	partial	428	18
OAPI	14.7	partial	6819	-
Zimbabwe	14.4	partial	2094	268
Panama	13.6	partial	2411	1002
Chinese Taipei	12.2	partial	583999	566918
El Salvador	9.8	partial	1329	320
Malaysia	9.2	partial	6321	9516
Ukraine	8.6	partial	48114	54103
Indonesia	5.0	partial	14326	2146
Chile	3.8	partial	3445	2739
Armenia			55	548
Azerbaijan			51	1252
Sudan			31	162
Tunisia			22	1128
Venezuela				2682
Saudi Arabia				2285
Iran				2255
United Arab Emirates				1021
Bolivia				220

2a) Spatial and temporal coverage

How to distinguish 'missing' observations from zeros?

- Construct coverage weights by office/year

  e.g. w=0.164 if batches for 60 days in a given year

- Assignment rules

Document type	Coverage weight
Singleton priority	w^{PRIO}
Claimed priority	max (w^{PRIO}, w^{DUPL})
Duplicate application	w^{DUPL}

- Could be more fine-grained if estimated econometrically
- Benefit:
  - Clearly identify 'true' zeros
  - Identify non-zero counts with 'low' reliability (sample selection based on threshold coverage)

2b) Designation of national jurisdictions in regional patent filings

Alternatives:

- Do nothing (most common)
- Designated countries in PAT_EP
- Publication kind codes in PATSTAT
- PRS Legal Status database for PATSTAT

Candidate statistics and legal status search strategy

Candidate statistic		*Search strategy*
1. Propensity to designate states at application	AK-A	prs_code=AK and publ_kind_code=A%
2. Propensity to designate states at payment of designation fee	AKX-RBV	prs_code=AKX or prs_code=RBV
3. Propensity to designate states at grant	AK-B	prs_code=AK and publ_kind_code=B%
4. Propensity to pay post-grant fees (annual maintenance fees)	PGFP	prs_code=PGFP

2b) Designation of national jurisdictions in regional patent filings

Our approach:

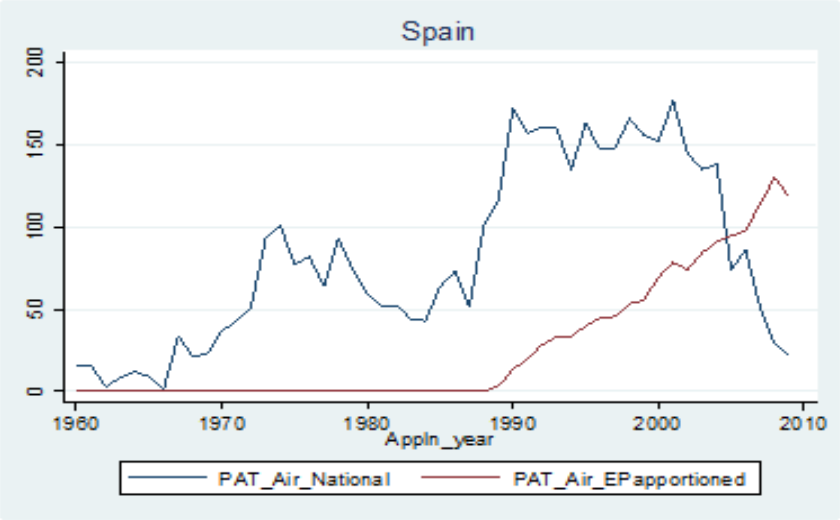
- Data on “payment of designation fees” to construct designation propensities over time for TOTPAT and apply these on EPAT
- Could be estimated econometrically for a more fine-grained attribution
- This is useful for:
  - Estimation of patent family size to construct indicators (dyadic patent family = weighted CP)
  - Apportionment of patenting within the EPO area to construct “patent stocks”

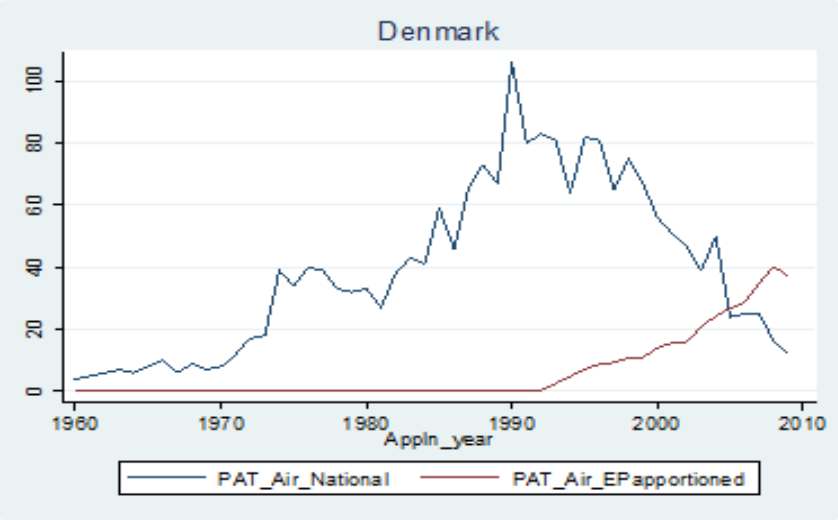
Estimation of patent family size using EP designation propensities

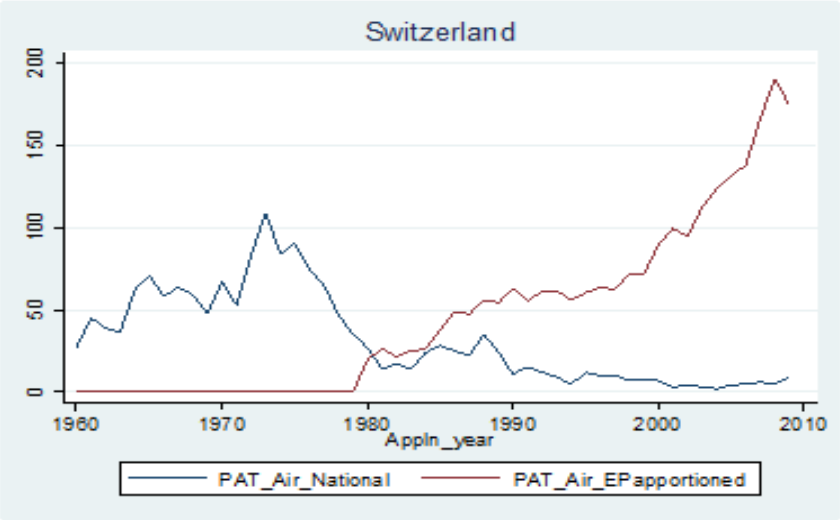
Observed family	Estimated family size in year 2000
EP singleton	2.013
EP + DE	(2.013-0.511) + 1 = 2.502
EP + US	(2.013) + 1 = 3.013

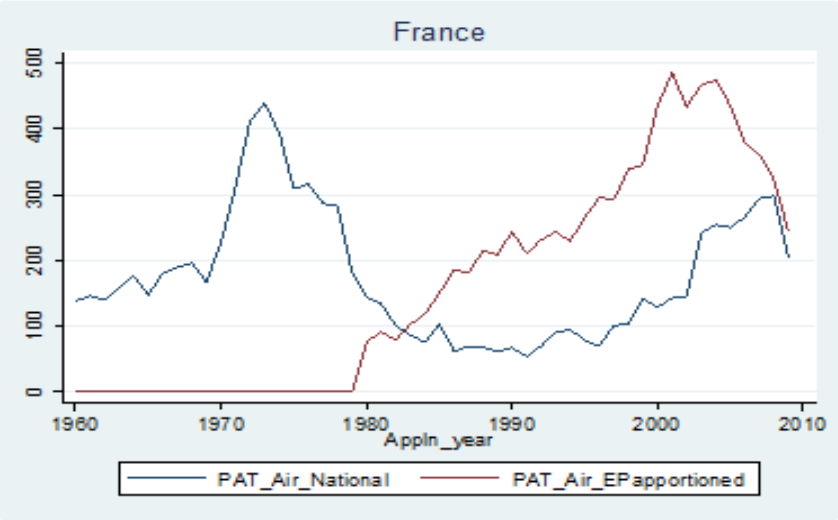
2b) Designation of national jurisdictions in regional patent filings

Patenting activity at the national office versus EP-apportioned filings









2c) Missing information

- Inventors

Benefit of imputing inventor information from duplicate filings

	Priorities with known inventor country	Priorities with known inventor country ***retrieved within PATSTAT***
Renewable energy (Y02E10)	42.3%	46.2% (+3.9)
Geothermal energy (Y02E10:1)	54.0%	58.1% (+4.1)
Wind energy (Y02E10:7)	52.6%	55.7% (+3.1)
Wind motors (F03D)	53.2%	56.3% (+3.1)
(Waste)water treatment (C02F)	35.3%	38.8% (+3.5)

- Patent classifications

Benefit of imputing IPC symbols using the APPLN_ECLA table

	Nb. of documents identified (appln_id's)		
	search in APPLN_IPC only	search in both APPLN_IPC & APPLN_CPC	search in APPLN_CPC only
(Waste)water treatment (C02F)	433,698	448,427 (+3%)	199,435
Wind energy (Y02E10:7)	-	-	62,702
Wind motors (F03D)	64,339	69,476 (+8%)	43,151
Climate mitigation in transport (Y02T)	-	-	293,670
Electric & hybrid cars (IPC-based)	113,038	128,991 (+14%)	71,357

2d) Identification of relevant documents

Classification of patent documents is not always systematic:

- National systems (ECLA, USPC, FI) – only a subset of docs is classified; harmonization (CPC) helpful!
- IPC – core vs advanced level
- ECLA/CPC – Y02 tags
  - A valuable addition!
  - For historic series based on search algorithms using a variety of attributes in DOCDB (even those not included in PATSTAT)
  - Applications prior to ~2010 (Y02C,Y02E), ~2012 (Y02B,Y02T)
  - Assigned manually thereafter (as any other CPC symbols)

➢ Implications for patent searches

➢ More or less difficult to determine the population from which 'one draws'

2d) Constructing variables to normalize (or control for) non-systematic classification

	If EPAT search strategy is based on:	***…then TOTPAT should be constructed as:***	*Appln_id's*
(1)	IPC symbols	All documents 'identifiable' using IPC	79%
(2)	ECLA symbols	All documents 'identifiable' using EC	49%
(3)	Keyword searches on titles and/or abstracts	All documents (families) with title /abstract in the corresp. language	58% (EN)
(4)	Y02 tags	All documents (families) that could have potentially been 'tagged'	? (this will vary by individual Y-symbol)
(5)	IPC or ECLA symbols	The union of (1) and (2) above.	80%
(6)	IPC, ECLA, ICO, or EN title/abstract	The union of the respective counts	84%
		No restriction	100%

- How to construct a corresponding TOTPAT ?
  - The same indicator (e.g. PF2, TPF, PCT)
  - The same concept (e.g. invention, co-invention, protection, citation)
  - The same type of search strategy (e.g. based on IPC, ECLA, keyword searches, etc – see above)
  - An otherwise identical algorithm as for the EPAT count (i.e. treatment of idiosyncrasies, imputation, other programming details that might affect the final outcome)

3) Implications for analysis

- Descriptive analysis
  - Provide context; normalize
- Econometric analysis
  - Control for idiosyncrasies

3) Implications for analysis

A. Conceptual (economic) reasons:
   - Differences in inventive capacity
   - Differences in propensity to patent
   - Differences in patent breadth and patent ‘quality’
   - Other factors that might affect patenting in general

B. Idiosyncratic (methodological) reasons:
   - Incomplete info due to differences in coverage of patent databases
   - Imperfect info on jurisdictions where patent protection is sought through regional procedures (designation)
   - Extent of missing information on inventors, applicants, patent classifications (incl. after imputation)
   - Differences in ability to identify the relevant documents due to non-systematic assignment of classification symbols

Using TOTPAT deals perfectly with B (there is no other way), imperfectly with A (imperfect as any other proxy).

Conclusion

- Draw attention to issues specific to analysis of narrow tech fields in a cross-country context; esp. emerging/developing economies
- Trade-off between patent quality, data availability and breadth of technological fields
- Need to adapt choice of patent indicators to context; the “optimal” family size for a given application is an empirical question (although PF2 often suitable)
- Need to address idiosyncratic problems in the underlying data
- Do not blindly estimate on the contents of a patent database
- Solution: indicator construction + normalization (control) variables