Getting Patents & Economic Data to Speak to Each Other:

An 'Algorithmic Links with Probabilities' Approach for Joint Analyses of Patenting & Economic Activity

Travis J. Lybbert, University of California, Davis

Nikolas J. Zolas, U.S. Census Bureau

WIPO

Max Planck Institute

11 March 2013

NSF

'Patent Heterogeneity' Matters

- Patenting intensity, patent strategies, and patent values vary dramatically across industries
  - Distinct differences across sectors: e.g., manufacturing, pharmaceuticals, electronics, chemicals, agriculture, etc.
  - Important differences *within* many sectors
- This patent heterogeneity has important economic implications for research and policy
  - 'One size fits all' patent policy debate
  - Determinants of patenting decisions
  - Economic effects of patents and patent strategies

- **Understanding these evolving, policy-relevant relationships requires joint analysis of patent and economic data at 'relevant resolution'**

Linking Patents to Economic Activity

- We devise an algorithm-based *platform* for empirical patent analysis:

A high resolution linkage between patents and economic data

- This platform
  - Can support a host of research questions addressed from a variety of methodological angles
  - Enables joint analyses at different levels of resolution
  - Is easily updated and applicable to a range of classifications

- The full paper and the concordance files we construct are available on WIPO's *Economics Publications* site

  Google "WIPO Lybbert"

Levels of Linkages

1. **Macro-level linkage**: Aggregate patent data linked to aggregate economic data by country, country pair and/or year
   - Quick and coarse…and common
   - Misses substantial differences across industries and firms

2. **Meso, industry-level linkage:** Patents linked to disaggregated economic data based on industrial relevance
   - Leverages existing high resolution patent and economic data
   - Broader scope, greater empirical flexibility complements firm-level analyses

3. **Micro, firm-level linkage**: Firm-level patent filings linked to firm-level economic activity (e.g., FDI, market share, exports, etc.)
   - Enables rigorous analysis of firm-level strategies
   - Encouraging progress on this front, but such data remain relatively sparse
   - Even fully linked firm data must be analyzed in broader industry context

Building a Meso, Industry-Level Linkage

Patents are classified by Intl. Patenting Classification (IPC)

- Facilitate prior art searches
- Classifies technical features
- Hierarchical and extremely high resolution

Economic data are classified by trade and industry classifications

- Standard International Trade Classification (SITC)
- International Standard Industrial Classification (ISIC)
- Hierarchical and high resolution

Section B: Performing Operations; Transporting

B 64: Aircraft; Aviation; Cosmonautics

B 64 D: Equipment for fitting in or to aircraft; Flying suits; Parachutes; …

SITC Section 7: Machinery and Transport Equipment

79: Other transport equipment

792: Aircraft and associated equipment; spacecraft (including satellites);…

Building a Meso, Industry-Level Linkage

Patents are classified by Intl. Patenting Classification (IPC)

- Facilitate prior art searches
- Classifies technical features
- Hierarchical and extremely high resolution

Economic data are classified by trade and industry classifications

- Standard International Trade Classification (SITC)
- International Standard Industrial Classification (ISIC)
- Hierarchical and high resolution

- Several other datasets borrow SITC and ISIC structure (R&D, capital, etc.)
- Insightful analyses possible with robust IPC-SITC and IPC-ISIC linkages
- Linking at coarse level can be done manually but is often too coarse
  MERIT (1994) and DG (2003) concordances
- Linking at higher resolution is more useful but complex
  - Canadian patents (1972-95) cross-classified IPC and cSIC → implicit links
  - Yale Technology Concordance (YTC) and OECD Concordance

Example: PCT Patent US1999016386

- Title: “Spray on bandage and drug delivery system”
- Applicant: 3M
- IPC:
  - A61K (OST “Medical Engineering”) → 8 SITC classes (CR4=65%)
  - A61L (OST “Pharmaceuticals/cosmetics”) → 10 SITC classes (CR4=52%)
- PCT Designations: China, New Zealand, Israel, Slovakia, Romania, Lithuania, Turkey, Kenya (ARIPO), Spain (EPO), Sweden (EPO),…
- NPE: China, New Zealand
- **With probabilistic links from these two IPC subclasses to SITC classes one can compute patent-specific weighted trade flow measures**
  - **US-destination patent-specific economic ties (bilateral trade)**
  - **Internal patent-specific market demand (total imports to destination)**
  - **Internal patent-specific production capacity (total exports from destination)**
  - **Competition (changes in trade and patenting, patenting intensity)**
- **Other useful patent-specific measures emerge with IPC-ISIC linkages**

The Canadian Bottleneck Problem: Composite Concordances Rapidly Atrophy

Conventional Composite Concordance Approach

"Algorithmic Links with Probabilities" (ALP) Guiding Principles

1. Use descriptive content of patents as the basis for the concordance
2. Eliminate need for concordance layering by constructing direct links
3. Automate the construction process as much as possible
   - Minimize manual work and subjective judgments
   - Optimize use of search and matching algorithms
   - Facilitate updates to respond to evolving technologies and industries

- **ALP-Data Mining**
- **ALP-Probability Matching**

ALP-Data Mining Approach

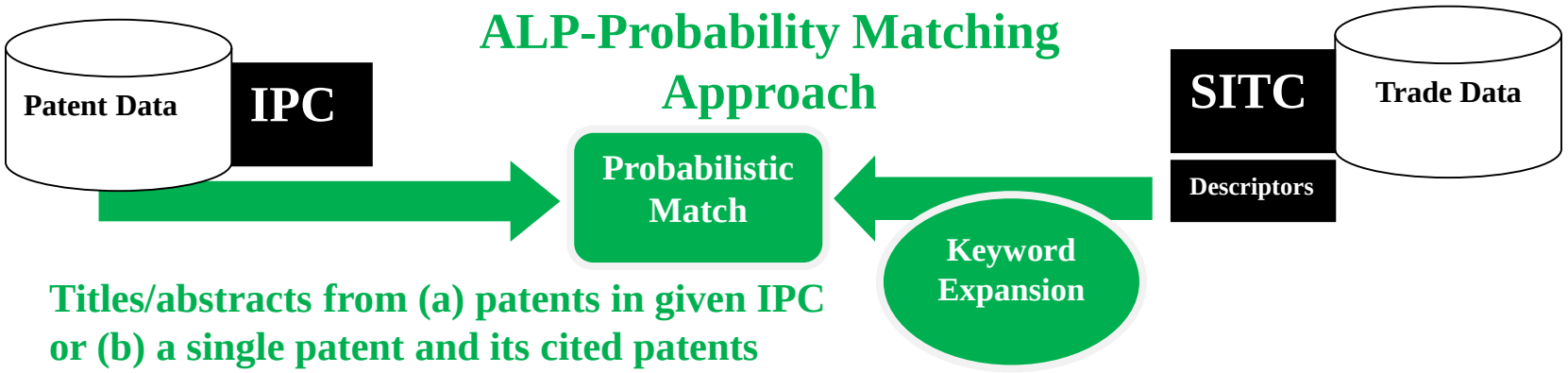
- Derive keywords from SITC descriptors and augment with PATENTSCOPE's 'cross-lingual' synonym expander
- Search PATSTAT for these keywords
- Extract IPCs from patents retrieved by search
- Compile frequency distribution over IPCs
- Filter and weight results to construct concordance

ALP-Data Mining Approach

IPC Frequency for SITC 8484, "Headgear and Fitting Thereof"

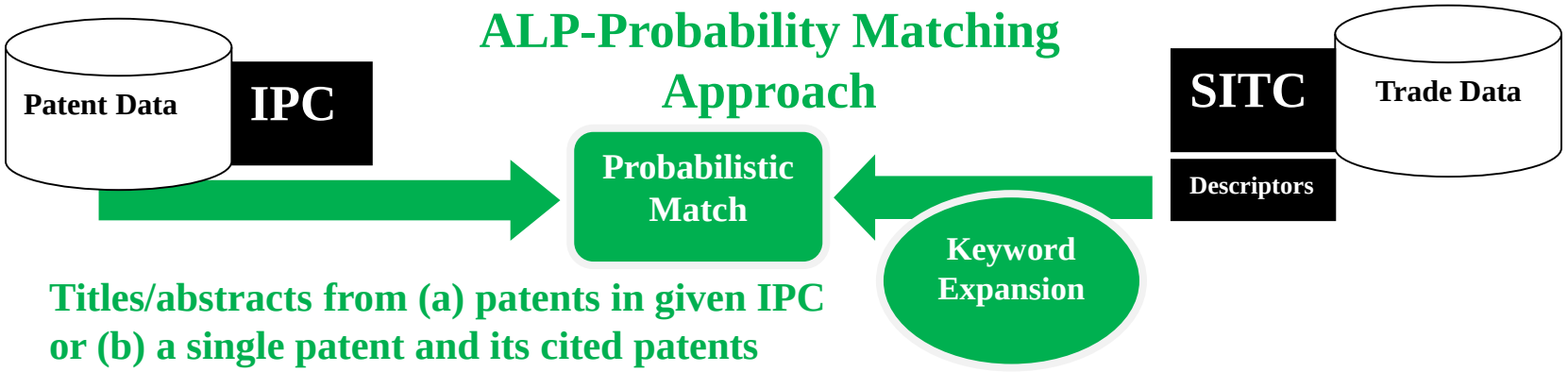
IPC	Raw	Specificity weights	Hybrid weights	IPC Description
A42B	43.1%	53.5%	97.8%	Hats; Head Coverings
A42C	1.5%	23.4%	-	Manufacturing or Trimming Hats
A62B	5.2%	9.9%	2.2%	Devices For Life-Saving
B68B	0.1%	5.1%	-	Harness; Whips Or The Like
F41H	1.7%	5.1%	-	Armour; Camouflage
B63C	1.6%	3.0%	-	Life-Saving In Water;

ALP-Probability Matching Approach



- Keywords derived from SITC descriptors and expanded
- (a) keywords from the titles/abstracts of patents from a given IPC
- (b) keywords from the title/abstracts of a single patent and all patents it cites
  - This could ultimately lead to patent-specific matching
- Probabilistic links based on weights from ranking keyword matches

ALP-Probability Matching Approach



IPC Number	A42B	
IPC Description	Headwear – Hats; Head Coverings	20,988 analyzed

SITC Description	Raw	Specificity	Hybrid
8484 - Headgear and fitting thereof, nes	65.0%	13.8%	72.3%
6576 - Hat shapes, hat-forms, hat-bodies and hoods	19.0%	7.4%	11.4%
6571 - Articles of felt, nes	8.1%	-	-
8421 - Overcoats	7.9%	-	-
6579 - Special products of textile material	-	20.6%	-
6517 - Yarn of regenerated fibres	-	20.1%	6.7%
6577 - Wadding, wicks	-	14.5%	4.6%
6543 - Woven fabric of wool or fine hair, nes	-	12.8%	5.0%
6581 - Textile material used for packing of goods	-	7.9%	-
6121 - Articles of leather used in mechanical appliances	-	3.0%	-

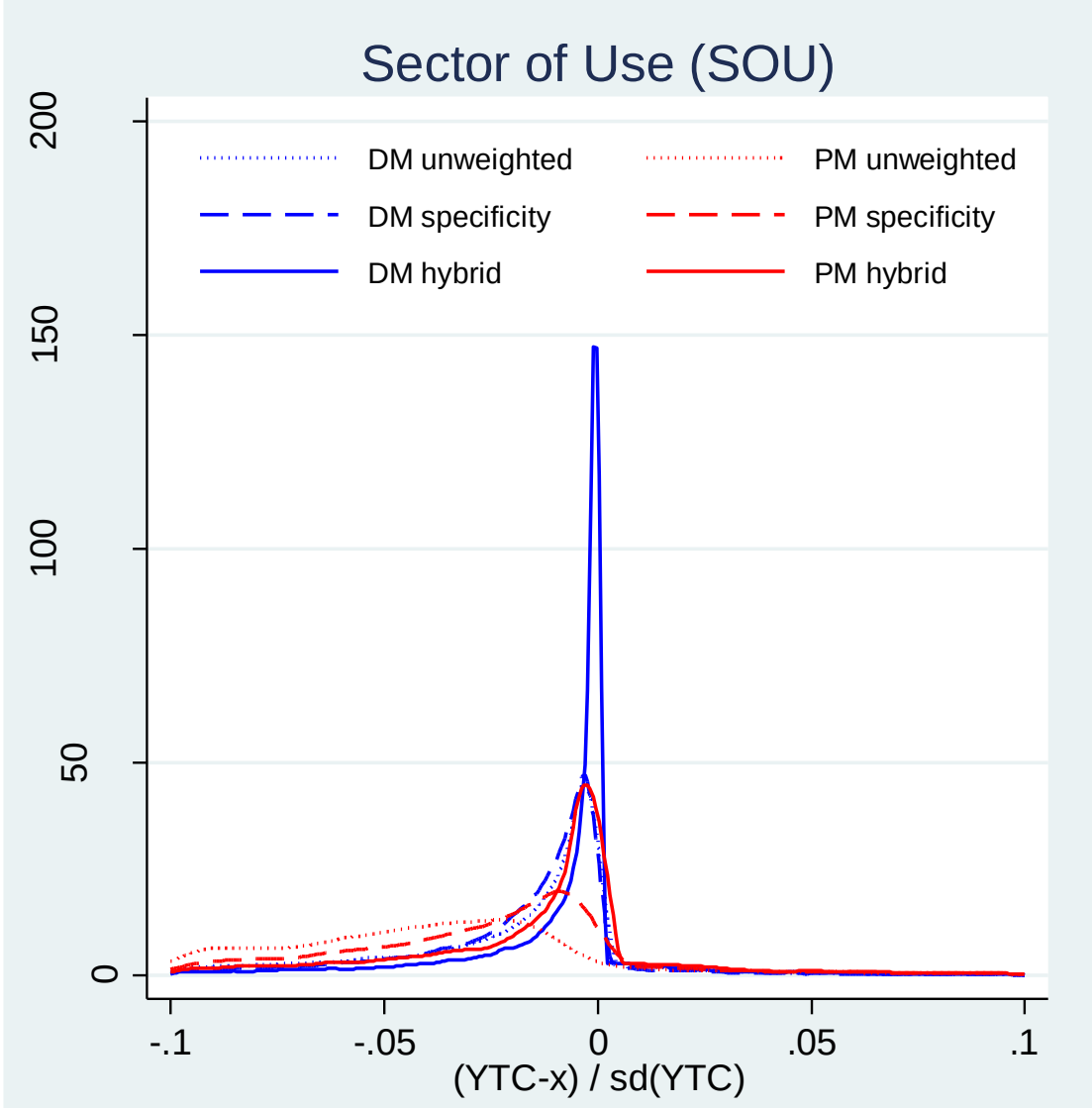
‘Ground Truthing’ ALP concordances

- Manual inspection with summary statistics
- Direct comparison with existing concordances
  - DG (2003) concordance: IPC-ISIC
  - YTC concordance: IPC-cSIC

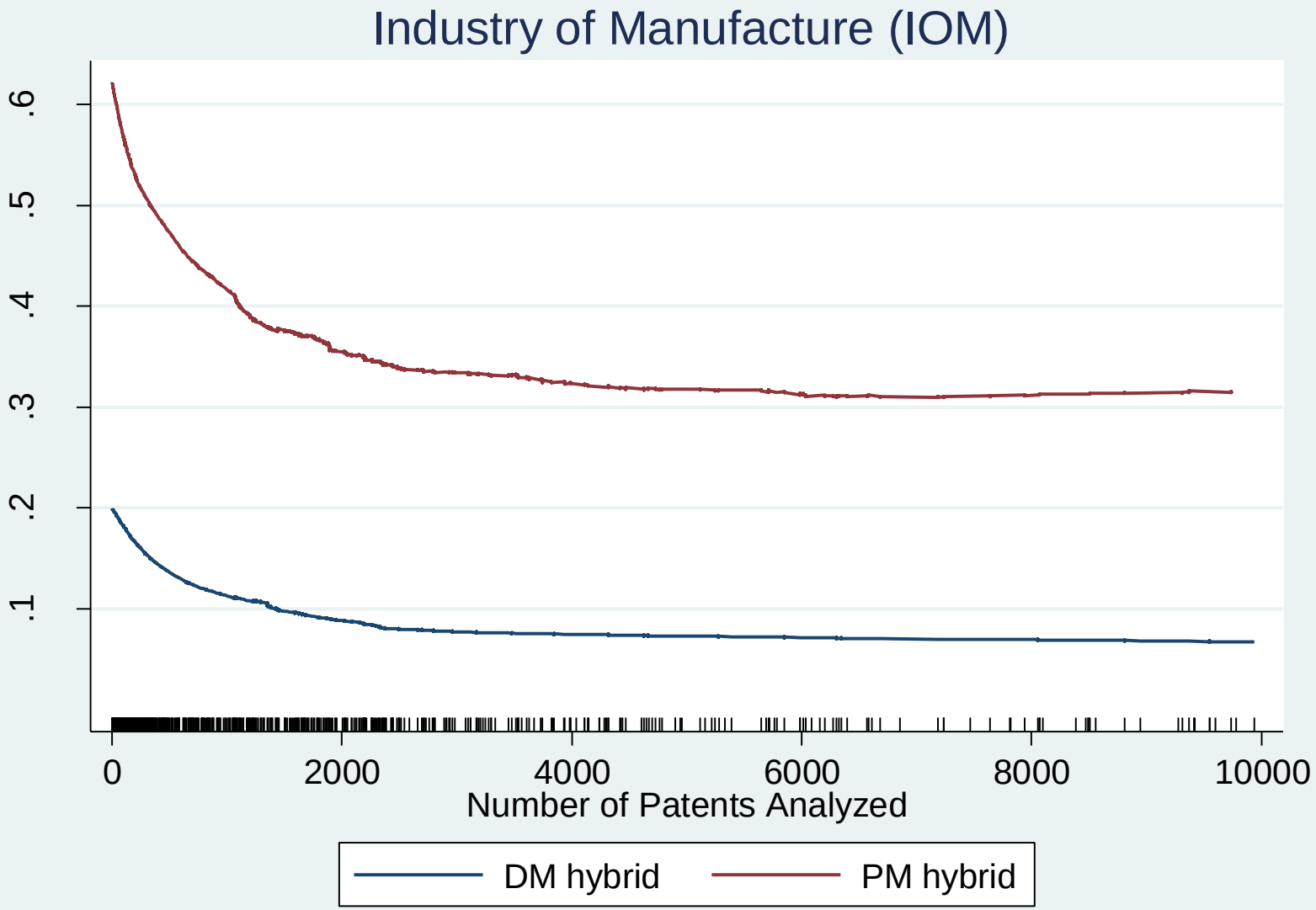
How well can a ALP concordance replicate patent examiners’ cross-classification of patents?

How do ALP concordances compare to the YTC?

- We construct cSIC ALP concordances based on Canadian patents that were cross-classified (1975-1995)
- When YTC puts zero weight on a given IPC-cSIC link, 80% of ALP weights are also zero
- Excluding these matched zeros, most differences are less than 5% of the std dev of the YTC



Do differences fade with 'large numbers'?



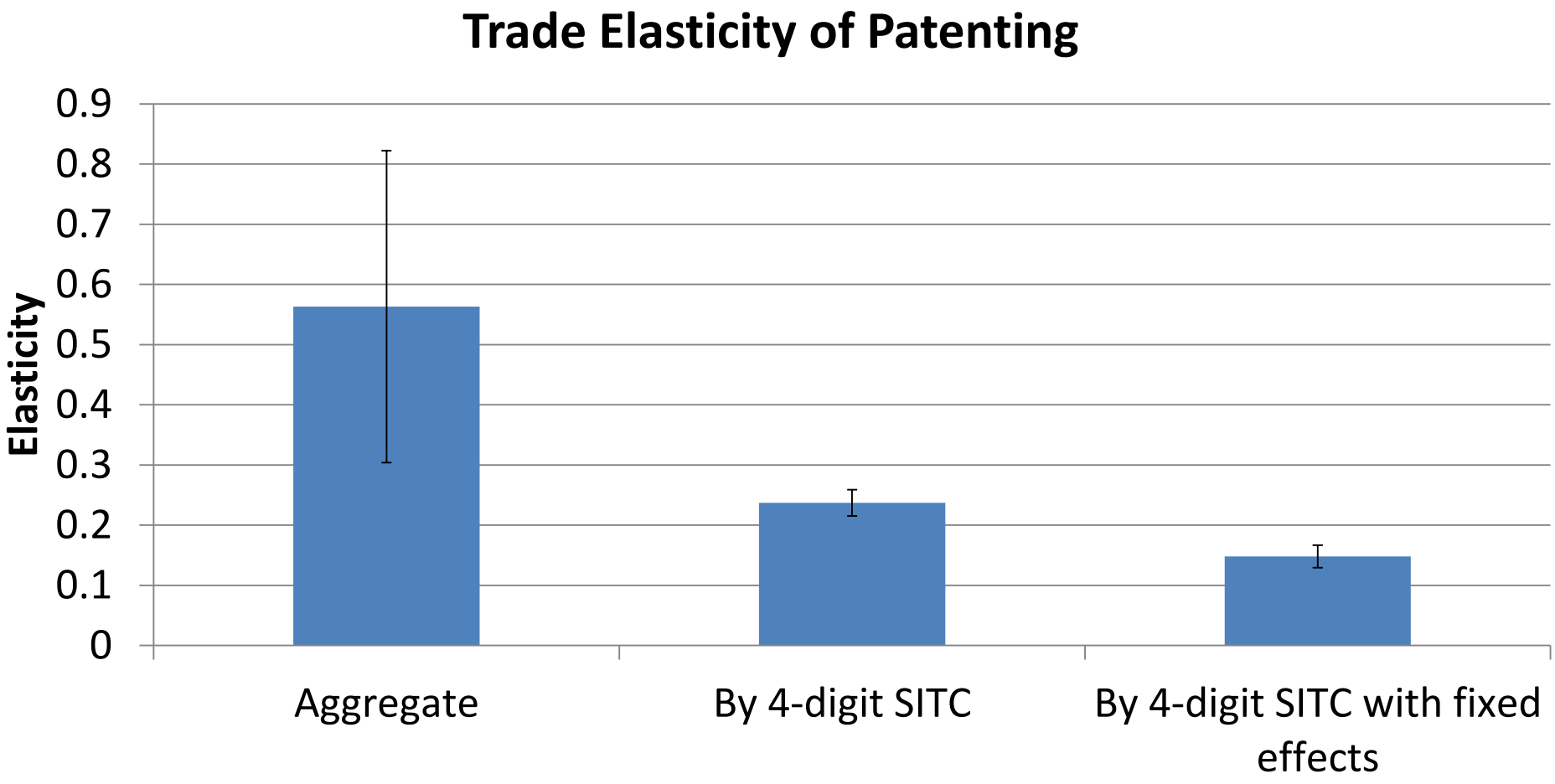
A Disaggregated Model of Patent & Trade Flows

How are bilateral trade and patent flows related?

- Several have modeled this relationship using aggregate bilateral patent and trade flows
- We use the ALP-DM concordance to map annual bilateral patent flows (2001-2005) into 634 different 4-digit SITC classes and estimate an expanded 'gravity' model
  - ALP concordance enables disaggregated, SITC class-level analysis
  - … and allows us to introduce industry-specific variables

How are bilateral trade and patent flows related?

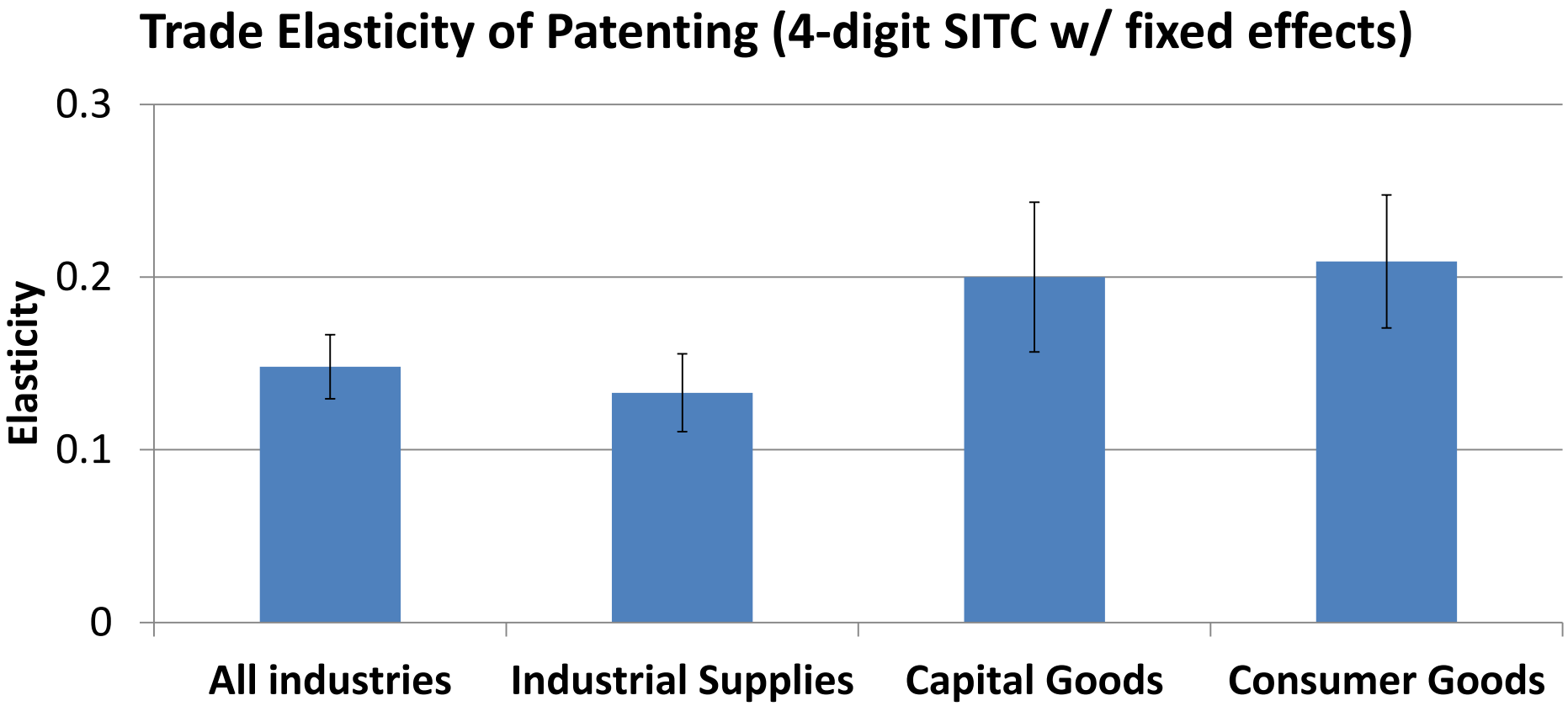
- Trade elasticity of patenting: [% Δ patents / % Δ trade]



Dep Var: Bilateral Patent Flows	Aggregate	Disaggregated by 4-digit SITC
ln Trade		
ln Destination GDP		
ln Origin GDP		
ln Elasticity of Substitution		
ln Distance		
Border Dummy		
Same Language Dummy		
Colonial Dummy		
Year Fixed Effects		
Country Fixed Effects		
Industry Fixed Effects		
Constant		
Observations		
Psuedo R^2		

How are bilateral trade and patent flows related?

- Trade elasticity likely to be heterogeneous across structurally distinct industrial groupings
  - UN constructs Broad Economic Classification (BEC) by SITC



Dep Var: Bilateral Patent Flows	All Industries	Industrial Supplies	Capital Goods	Consumer Goods
ln Trade	0.148***			
	(0.0116)			
ln Destination GDP	1.116***			
	(0.0293)			
ln Origin GDP	0.984***			
	(0.0305)			
ln Elasticity	-0.0236			
of Substitution	(0.0373)			
ln Distance	-0.163***			
	(0.0372)			
Border Dummy	-0.174			
	(0.119)			
Same Language	0.339***			
	(0.0788)			
Colonial Dummy	-0.630***			
	(0.104)			
Year Fixed Effects	Yes			
Country Fixed Effects	No			
Industry Fixed Effects	Yes			
Constant	-57.57***			
	(1.555)			
Observations	2,894,659			
Psuedo R^2	0.582			

Extensions to Sample Analysis

- *What role do patents play in different Global Value Chains (GVCs)?*
  - Lead/lag model
  - Value Added trade data using input-output tables
- *How do trademarks shape trade flows?*
- *How do firms in different GVCs and in different GVC positions exploit trademarks and patents to build and assert comparative advantage?*

In Conclusion

- ALP concordances between patents and economic data
  - Break the Canadian patent 'bottleneck' for high resolution links
  - Perform well compared to existing concordances
  - Can be easily updated as technologies and classifications evolve
  - Enable more rigorous analysis of patent heterogeneity across industries
  - Complement insightful firm-level patent analyses
- ALP concordances should enable richer, policy-relevant research
  - Sample analysis of trade and patent flows suggests some potential empirical insights from the ALP 'platform'
  - Enhanced patent landscapes with 'economic activity layers'
  - Sector-specific analyses of patenting, technology transfer and impacts of policy and institutional changes
  - Dynamic models of the economic impacts of innovation and patent strategies