Using Text-as-data Analysis to Assess the Design and Impact of Trade Agreements

Wolfgang Alschner[a], Julia Seiermann[b] & Dmitriy Skougarevskiy[c]

22 February 2017

PRELIMINARY WORK - PLEASE DON'T QUOTE WITHOUT PERMISSION OF THE AUTHORS

[a] University of Ottawa, wolfgang.alschner@uOttawa.ca

[b] The Graduate Institute and UNCTAD, julia.seiermann@unctad.org

[C] The Graduate Institute and European University at St. Petersburg, dmitriy.skugarevskiy@graduateinstitute.ch

Motivation

- Regional trade agreements (RTAs) are increasingly common and increasingly "deep"
- Current analysis of regional trade agreements: legal (detailed) versus economic (coarse)
- Empirical research focuses on few variables (constrained by necessity of manual coding)
- With increasing scope and depth of RTAs, existing empirical methods may not be fine-grained enough to capture meaningful differences
- Use text-as-data methods to
  - Reduce the trade-off between detail and scope
  - Increase precision of estimates of impact of RTAs on trade flows, welfare…

Literature: Capturing Trade Agreement Design

Indicator	Details		Reference
Type	FTA, customs union, common market, economic union		Baier/Bergstrand/Clance (2015), Baier/Bergstrand/Feng (2015)
Length	No. of provisions		Egger/Nihai (2015)
Depth	Binary indicator: "Deep" versus "shallow"		Aichele/Felbermayr/Heiland (2015)
	WTO-equivalent, WTO-plus, WTO-minus, WTO-extra		Marceau (2009), Ruta (2016)
	Additive index of "deep" features:		Dür/Baccini/Elsig (2014), Design of Trade Agreements (DESTA) database
	- Reduce all tariffs to zero - Services trade - Investments - Standards	- Public procurement - Competition - Intellectual property rights	

→These methods rely on resource-intensive manual coding of agreements

→ Our approach relies on computational text analysis ("text-as-data")

Textual database of RTAs

- Corpus of RTA full texts
- Basis: WTO RTA database
  - 476 treaties notified to the WTO, signed by 202 countries, 1948-2015
  - FTAs, customs unions and regional economic integration agreements
  - 60% currently in force
- Plus: TPP agreement (not in force yet)
- Transformed all texts (excluding annexes and schedules) into a unified marked-up text format (XML)

Indicator: Textual distance

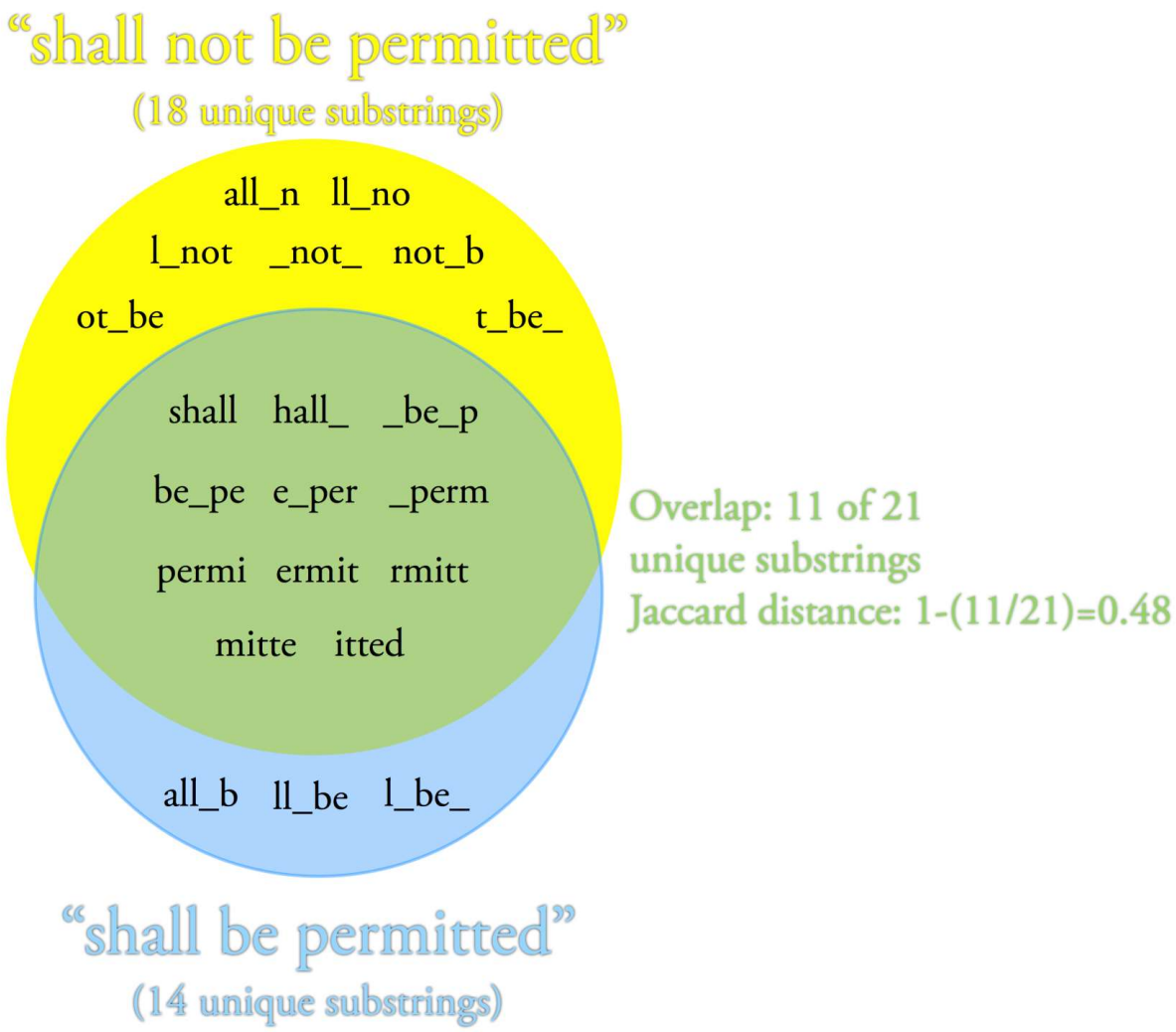
- Split each document into 5-character components
  - E.g. "free trade" → split into 6 components: "free ", "ree t", "ee tr", "e tra", " trad", "trade"
- Compute the Jaccard distance between pairs of treaties
  - Dissimilarity between two sets of 5-character components (A_i and A_j):

$$d_{ij} = \frac{\left|A_i \cup A_j\right| - \left|A_i \cap A_j\right|}{\left|A_i \cup A_j\right|}$$

- Textual similarity (how much the treaties resemble each other)

$$s_{ij} = 1 - d_{ij} = 1 - \frac{\left|A_i \cup A_j\right| - \left|A_i \cap A_j\right|}{\left|A_i \cup A_j\right|}$$

- *Note: Only possible with same language documents (for now: English)*

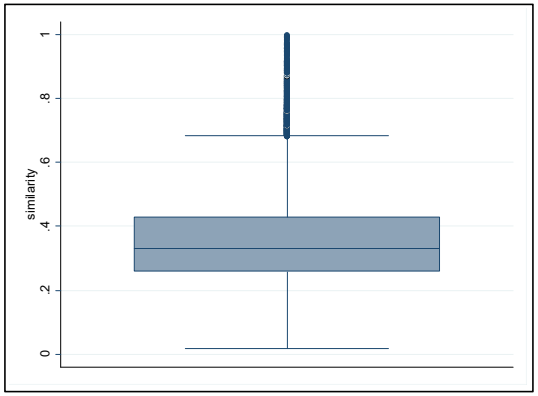

"shall not be permitted"
(18 unique substrings)
all_n ll_no
l_not _not_ not_b
ot_be t_be_
shall hall_ _be_p
be_pe e_per _perm
permi ermit rmitt
mitte itted
all_b ll_be l_be_
"shall be permitted"
(14 unique substrings)
Overlap: 11 of 21
unique substrings
Jaccard distance: 1-(11/21)=0.48

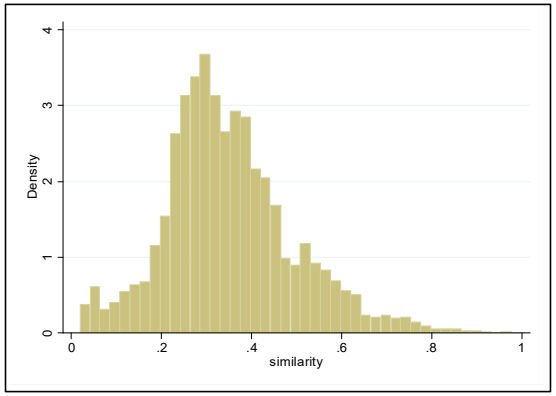
Descriptive statistics: Textual similarity, treaty-level

Range=0.02-0.99

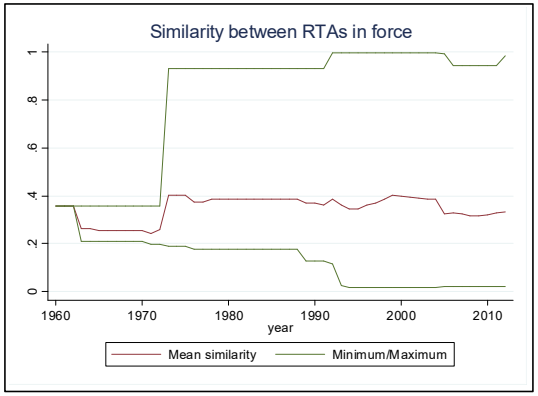
Median=0.33

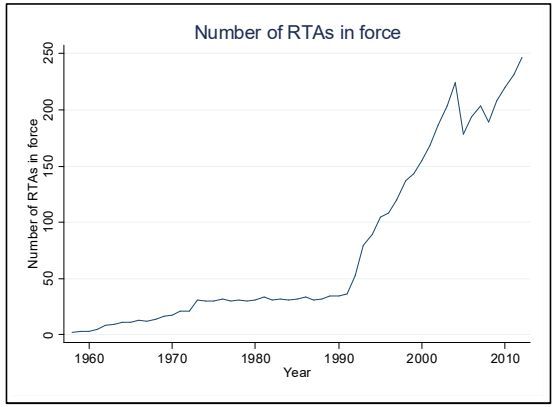
Mean=0.35





Mean similarity relatively stable over time





Most similar pairs of agreements

RTA 1	In force	RTA 2	In force	Similarity
EFTA - Czech Republic Agreement	1992	EFTA - Slovak Republic Agreement	1992	.9992
Slovak Republic - Estonia	1998	Czech Republic - Estonia	1998	.9947
EC - Algeria	1976	EU - Algeria	2005	.9933
Slovak Republic - Israel	1997	Czech Republic - Israel	1997	.9929
US - Colombia	2012	US - Peru	2009	.9858
EC - Czech Republic Europe Agreement	1995	EC - Slovak Republic Europe Agreement	1995	.9834

Disaggregation: Chapters/categories

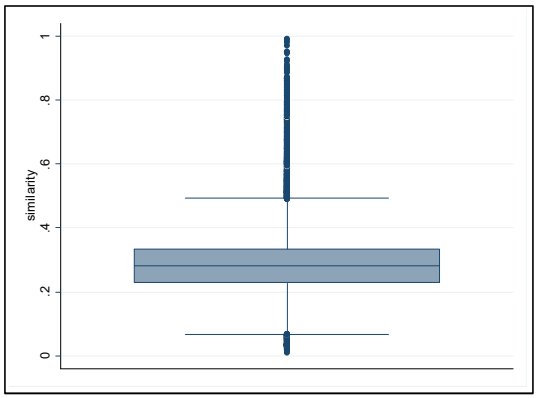
- Chapters with related content may not have similar names; in some agreements, there are no chapter names
- Solution: Identified 56 categories covered in RTAs based on chapter names
- Assigned each chapter in each RTA to 1 category
- Compute textual distance/similarity at the chapter-level

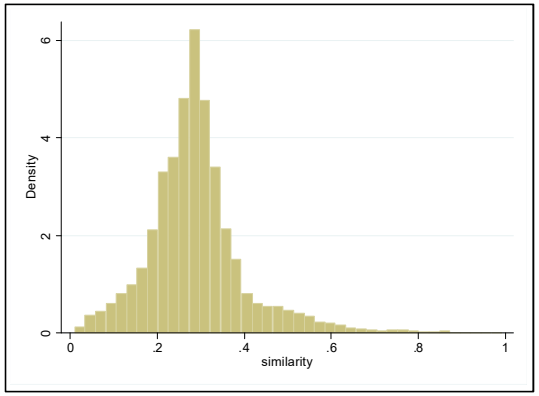
Descriptive statistics: Textual similarity, chapter-level (Trade in goods chapters)

Range=.01-0.99

Median=0.28

Mean=0.29

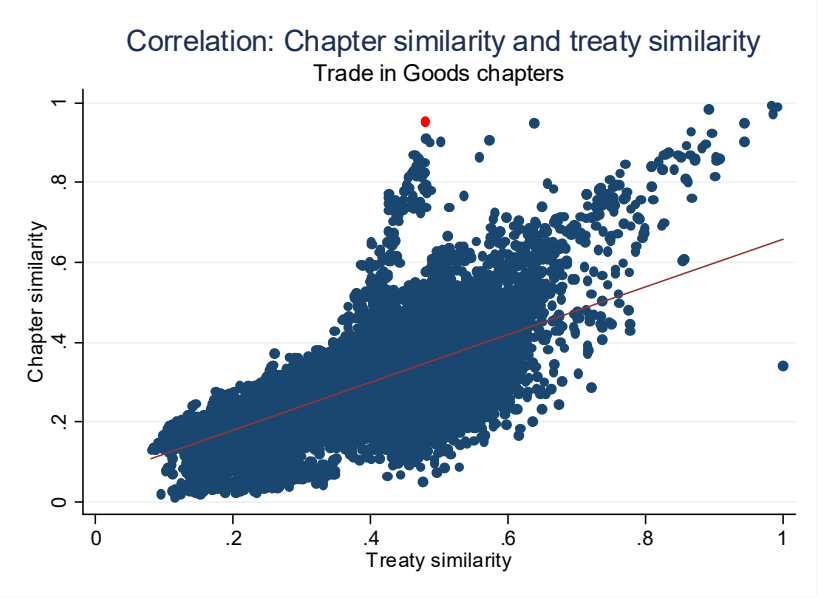




Most similar pairs of chapters

RTA 1	In force	RTA 2	In force	Chapter similarity	Treaty similarity
EC - Czech Republic Europe Agreement	1995	EC - Slovak Republic Europe Agreement	1995	.9926	.9834
EU - Algeria	2005	EC - Algeria	1976	.9889	.9933
US - Bahrain	2006	US - Oman	2009	.9815	.8928
US - Peru	2009	US - Colombia	2012	.9718	.9858
EC - Slovenia Europe Agreement	1999	EC - Slovenia Interim Agreement	1997	.9519	.4801

Correlation: Chapter and treaty similarity



- High chapter similarity also occurs between "dissimilar" agreements
- Example (red): EC - Slovenia Europe Agreement and EC - Slovenia Interim Agreement
- This highlights the importance of disaggregate analysis

Conclusion

- Text-as-data methods (similarity and, potentially, other indicators) can complement conventional empirical methods by offering a more detailed understanding of trade agreement design and its impact on trade and welfare
- (A lot of) Further research is needed, including
  - Assessment of the relation between textual similarity and impact similarity
  - Chapter-level and article-level analysis
  - Machine-coding and machine-learning to complement similarity measures
- Text-as-data analysis cannot replace in-depth legal analysis, but it can reduce the trade-off between detail and coverage of trade agreements (and other texts to be studied empirically)

References

Aichele, R., G. Felbermayr and I. Heiland (2015). Going Deep: The Trade and Welfare Effects of TTIP. CESifo Working Paper No. 5150.

Allee, T. and A. Lugg (2016). Who wrote the rules for the Trans-Pacific Partnership? Research and Politics, July - September 2016, pp. 1–9.

Alschner, W. and D. Skougarevskiy (2016a). The New Gold Standard? Empirically Situating the Trans-Pacific Partnership in the Investment Treaty Universe. Journal of World Investment & Trade. 17 (2016), pp. 330-373.

Alschner, W. and D. Skougarevskiy (2016b). Mapping the Universe of International Investment Agreements. Journal of International Economic Law, 19 (2016), pp. 561-688.

Anderson, J. and E. Van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review, 93(1), pp. 170-192.

Arvis, Jean-François and Ben Shepherd (2013): The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up' problem in gravity models. Applied Economics Letters, 20(6), pp. 515-519.

Baier, S., J. Bergstrand and M. Clance (2015a): Preliminary examination of heterogeneous effects on international trade of economic integration agreements. In: Dür, A. and M. Elsig (2015). Trade Cooperation. The Purpose, Design and Effects of Preferential Trade Agreements, pp. 355-373.

Baier, S., J. Bergstrand and M. Clance (2015b): Heterogeneous Economic Integration Agreement Effects. CESifo Working Paper No. 5488.

Baier, S., J. Bergstrand and M. Feng (2015): Economic integration agreements and the margins of international trade. Journal of International Economics 93(2014), pp. 339-350.

Baier, S, Y. Yotov and T. Zylkin (2016): On the widely differing effects of free trade agreements: Lessons from twenty years of trade integration. LeBow College of Business, Drexel University School of Economics Working Paper Series, WP 2016-15.

Dür, A., L. Baccini, and M. Elsig, "The Design of International Trade Agreements: Introducing a New Database," Review of International Organizations, 2014.

Egger, P. and S. Nihai (2015). Effects of deep versus shallow trade agreements in general equilibrium. In: Dür, A. and M. Elsig (2015). Trade Cooperation. The Purpose, Design and Effects of Preferential Trade Agreements, pp. 374-391.

Fally, T. (2015): Structural Gravity and Fixed Effects. Journal of International Economics, Volume 97, Issue 1, September 2015, pp. 76-85.

Grimmer, J. and B. Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," Political Analysis, 2013.

Larch, M. and Y. Yotov (2016): General Equilibrium Trade Policy Analysis with Structural Gravity. WTO Working Paper ERSD-2016-08.

Lodhi, H., C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins (2002): Text classification using string kernels. Journal of Machine Learning Research 2, pp. 419–444.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Research 27 (2), pp. 209–220.

Marceau, G. (2009): News from Geneva on RTAs and WTO-plus, WTO-more and WTO-minus, in: Proceedings of the 103rd Annual Meeting, vol. 103, pp. 124-128.

Piermartini, R. and Y. Yotov (2016): Estimating trade policy effects with structural gravity. WTO Working Paper ERSD-2016-10.

Ruta, M. (2016): Deep Trade Agreements. Presentation, WTO Public Forum, September 28, 2016.

Santos Silva, J.M.C. and S. Tenreyro (2006): The Log of Gravity. The Review of Economics and Statistics, November 2006, 88(4), pp. 641–658.

Spirling, A. (2012): U.S. Treaty Making with American Indians: Institutional Change and Relative Power, 1784–1911, American Journal of Political Science, January 2012, 56(1)., pp. 84-97.