

# What goes on in court? Identifying contract-related topics decided by United Kingdom courts from 1709 to 2021 using machine learning

Ahmed Izzidien<sup>1</sup>, Holli Sargeant<sup>2</sup>, Felix Steffek<sup>3</sup> (University of Cambridge)

**Abstract** Despite the relevance of contractual conflict in legal practice, there is yet to be a dataset which captures the type of issues and clauses that result in cases being brought before the courts. Such a dataset would be invaluable to a machine learning algorithm that seeks to predict whether new clauses are likely to cause conflict. In this study, we analyse a dataset based on half a million United Kingdom court decisions decided between 1709 and 2021, from which we extract 60,379 cases dealing with contracts. We characterise the language of this dataset using Latent Dirichlet Allocation to approximate legal topic modelling. We augment the data by plotting it with the court names and dates for each case, which allows for a racing bar chart visualisation. This is the first study of its kind to provide easy access to legal researchers on cases dealing with contracts in the United Kingdom.

## Introduction

Legal conflict costs UK organisations £28.5 billion annually and trillions globally (Acas 2021). For many citizens and small or medium-sized enterprises, litigation costs are often out of their reach. Current methods to ameliorate this include alternative dispute resolution. Yet, all such methods attempt to solve the problem after it has occurred instead of preventing it (Zeleznikow 2021).

Artificial intelligence, specifically machine learning, is increasingly deployed in law, yet none of these implementations has considered the problem of predicting conflict based on the wording of contract clauses. As a first step, we build a new corpus of cases specifically concerned with contract clauses. We use machine learning methods to categorise court cases according to their topics and extract their dates and court names. The topic allocation process is unsupervised in that it requires no human annotation or tagging of documents by hand. The algorithm discovers the semantic structure of the cases by examining statistical co-occurrence patterns using a statistical technique known as the Dirichlet mode. In doing so, it identifies hidden (latent) patterns which allow it to assign topics to each case (Blei, Ng, and Jordan 2003). We plot the prevalence of each type of case according to both the date and court. This plot can be accessed as a racing bar chart by scanning the QR code below. We present our methodology, findings, a discussion and limitations.

## Methodology

We begin with a corpus (the Cambridge Law Corpus) which contains around half a million decisions of United Kingdom courts between 1709 and 2021. We proceed as follows: Step 1: We extract cases in which any of the key words: ‘contract’, ‘agreement’ or ‘clause’ occur. Step 2: We perform a topic modelling of these cases using Gensim’s Latent Dirichlet Allocation (LDA) (Rehurek and Sojka 2011). The process considers all cases and attempts to derive fifty topics that best represent them. Each topic found by the LDA is represented as a bag-of-words, for example, one topic found was represented by: (‘say’, ‘make’, ‘notice’, ‘give’, ‘tenancy’, ‘rent’, ‘tenant’, ‘property’, ‘premise’, ‘document’, ‘date’, ‘require’, ‘take’, ‘provide’, ‘landlord’, ‘lease’, ‘pay’, ‘include’, ‘refer’, ‘contain’, ‘account’, ‘regulation’, ‘follow’, ‘partly’, ‘charge’). Step 3: Two legal experts consider each bag-of-words and allocate a functional topic to them. In the previous example, the topic allocated was property. Step 4: Using these functional topics, each case is auto labelled by its most dominant topic. Step 5: An analysis of the data is given in the form of (i) a tabulation of the names of the courts and number of cases (Table 1), (ii) the number of cases per topic (Table 2), (iii) a cumulative line plot (Figures 1 & 2) and (iv) a racing bar chart (please scan the QR code) of the cumulative number of cases since 1709 per functional topic.

## Findings

Using the key words, 60,942 cases were found in the following courts.

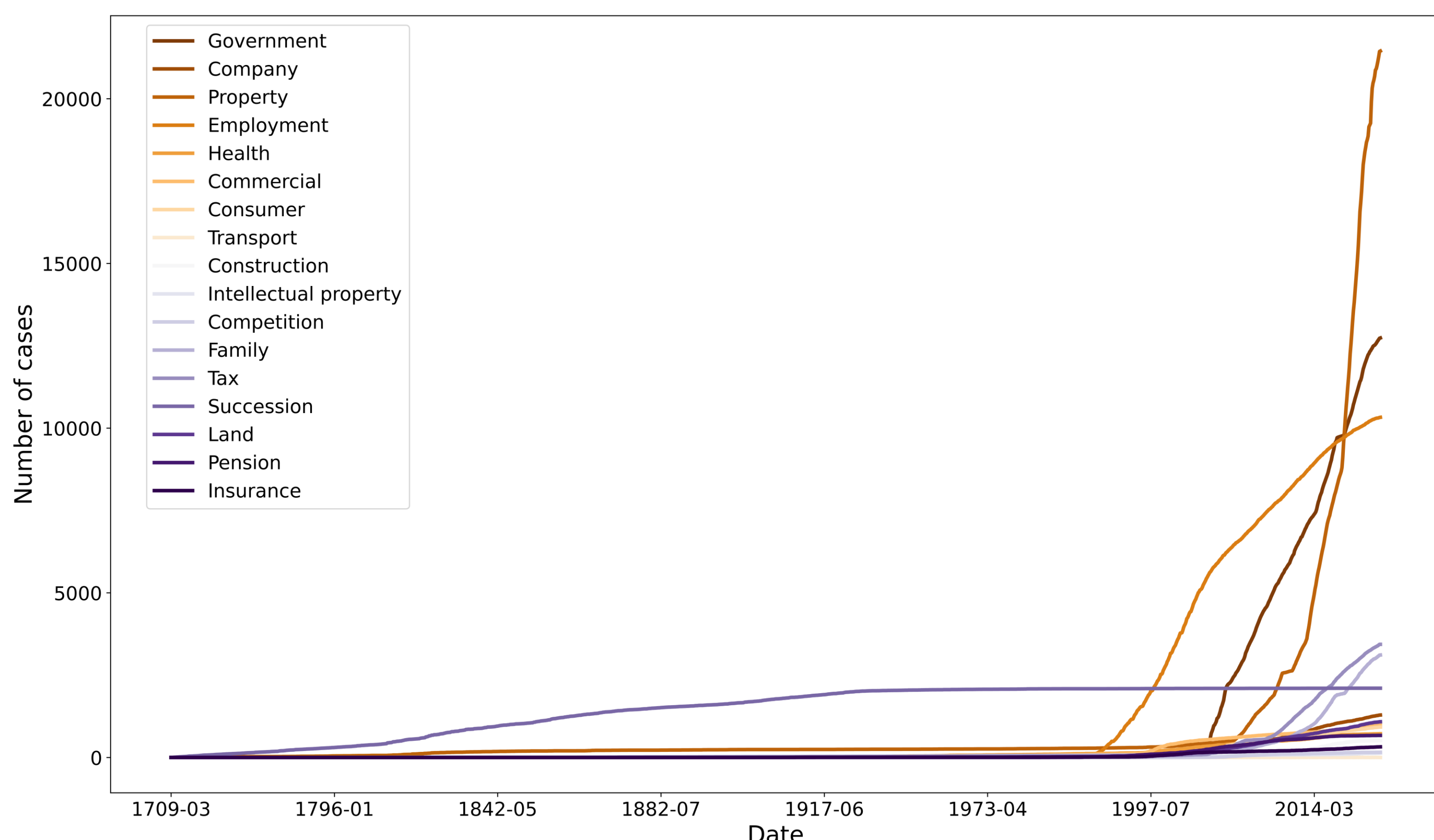
**Table 1:** Cases containing ‘contract’, ‘agreement’ or ‘clause’ by court name

Court	Cases
Court of Common Pleas	4
United Kingdom Investigatory Powers Tribunal	6
United Kingdom VAT & Duties Tribunals (Insurance Premium Tax)	7
United Kingdom VAT & Duties Tribunals (Landfill Tax)	7
England and Wales High Court (Exchequer Court)	7
Intellectual Property Enterprise Court	8
England and Wales High Court (King’s Bench Division)	14
England and Wales High Court (Admiralty Court)	16
England and Wales County Court (Family)	19
Northern Irish Courts	33
England and Wales Family Court (High Court Judges)	36
Special Immigration Appeals Commission	41
England and Wales Court of Protection	43
England and Wales High Court (Family Division)	46
United Kingdom Upper Tribunal (Tax and Chancery Chamber)	47
United Kingdom Financial Services and Markets Tribunals	48
United Kingdom VAT & Duties Tribunals (Customs)	51
England and Wales Court of Appeal (Criminal Division)	59
First-tier Tribunal (Health Education and Social Care Chamber)	62
England and Wales High Court (Patents Court)	68
English and Welsh Courts	96
United Kingdom Information Tribunal including the National Security Appeals Panel	116
England and Wales Care Standards Tribunal	117
England and Wales High Court (Queen’s Bench Division)	122
England and Wales High Court (Administrative Court)	133
First-tier Tribunal (General Regulatory Chamber)	137
England and Wales High Court (Commercial Court)	140
United Kingdom Competition Appeal Tribunal	188
United Kingdom VAT & Duties Tribunals (Excise)	201
England and Wales High Court (Technology and Construction Court)	209
England and Wales Family Court (Other Judges)	237
England and Wales High Court (Chancery Division)	243
England and Wales Land Registry Adjudicator	286
United Kingdom Special Commissioners of Income Tax	318
United Kingdom Upper Tribunal (Immigration and Asylum Chamber)	372
United Kingdom Asylum and Immigration Tribunal	421
England and Wales Lands Tribunal	470
England and Wales Court of Appeal (Civil Division)	582
Information Commissioner’s Office	627
United Kingdom Supreme Court	632
United Kingdom Upper Tribunal (Lands Chamber)	775
United Kingdom Social Security and Child Support Commissioners	821
The Judicial Committee of the Privy Council	849
United Kingdom Upper Tribunal (Administrative Appeals Chamber)	884
United Kingdom VAT & Duties Tribunals	950
United Kingdom House of Lords	3486
First-tier Tribunal (Tax)	4117
First-tier Tribunal (Property Chamber)	4360
England and Wales Leasehold Valuation Tribunal	5001
United Kingdom Employment Appeal Tribunal	9128
United Kingdom Employment Tribunal	11071
United Kingdom Immigration and Asylum (AIT/IAC)	13251

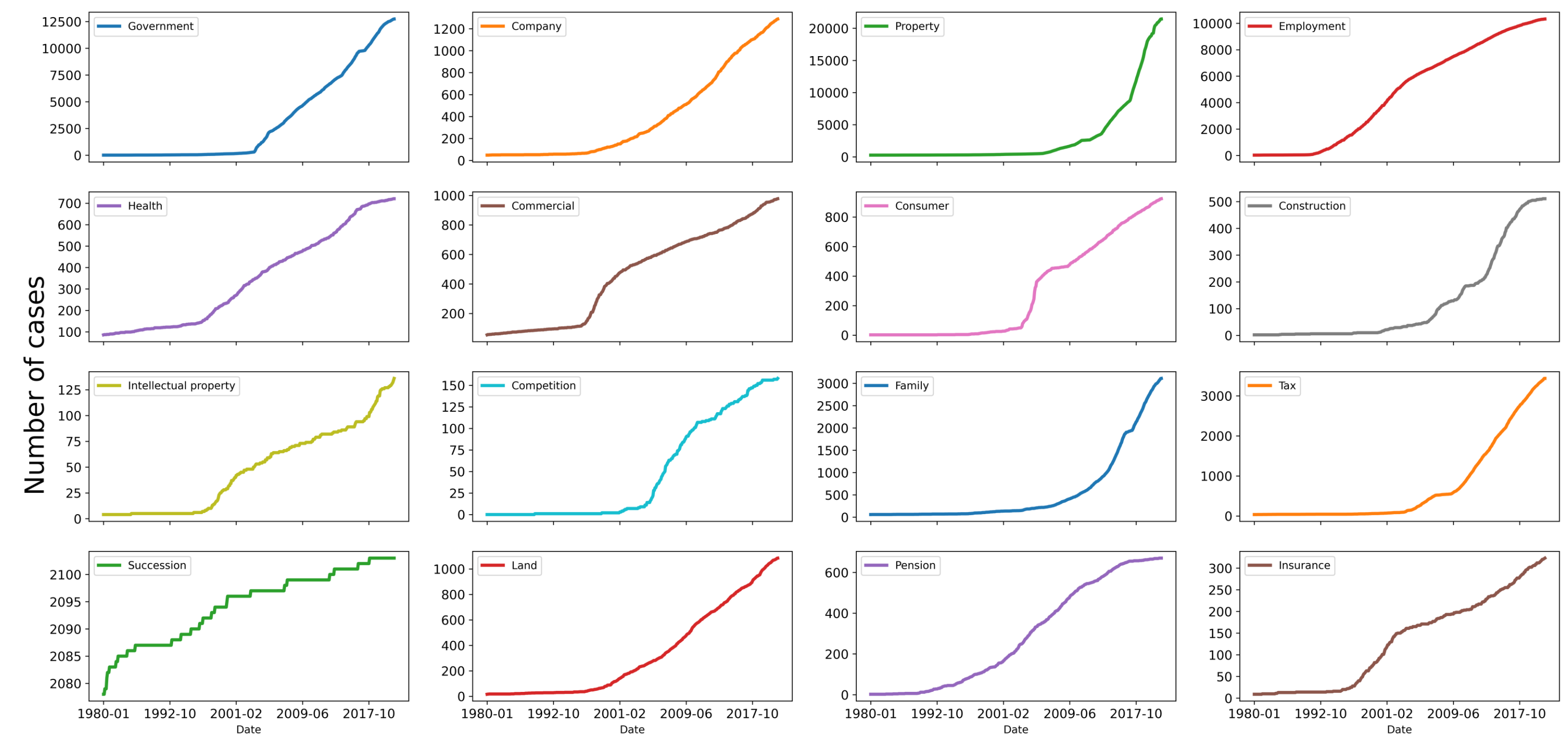
**Table 2:** Number of cases per topic

Topic	Number of cases
Intellectual property	136
Competition	158
Insurance	323
Transport	416
Construction	511
Pension	670
Health	721
Consumer	924
Commercial	978
Land	1085
Company	1289
Succession	2108
Family	3111
Tax	3436
Employment	10327
Government	12738
Property	21448

To visualise the racing bar chart use your smartphone camera to scan this QR code



**Figure 1:** Cumulative number of cases per topic.



**Figure 2:** Cumulative number of cases by topic, plot individually.

## Discussion

Our findings reveal that different functional subject matters have appeared before courts in the United Kingdom at different rates across different historical periods. There may be various reasons why these rates increase in specific years. The representation of court decisions in the Cambridge Law Corpus as well as the definition of functional topics may play an important role. In addition to increased reporting of cases, the introduction of new laws and changes in procedural law (which may allow more people to bring cases to court) are likely contributors. For example, succession law was one of the earliest rises in topics. Succession law has been a well-established area of law since before Roman times (Sloan 2020). Perhaps early cases tested and recorded the established law before it levelled out over time as fewer people challenged succession law, and there are no immediately apparent structural reasons to encourage more litigation.

We have identified a high number of cases, in which contracts play a role, that concern property, employment and government. Figure 2 demonstrates that the considerable rise in property cases began in the late 1990s and has significantly increased since. Several pieces of legislation were introduced since the 1980s that developed property law in the United Kingdom (Landlord and Tenant Act 1985, Law of Property (Miscellaneous Provisions) Act 1989, Leasehold Reform, Housing and Urban Development Act 1993, among others). At the same time, the courts continued to develop the law of trusts and personal property.

Similarly, government cases rose notably in the early 2000s. It is more difficult to speculate on this change. Potentially, this is due to cases concerning public matters mentioning a contract without these being cases dealing with contractual conflicts; or it could reflect an increase in government action taking the form of a contract. For instance, the Public Contracts Regulations 2006 introduced new rules for public and private procurements, further addressed in the Public Contracts Regulations in 2015.

Our findings demonstrate the prominence of employment cases over the years, in particular, after the 1990s. The trend may be related to changes in policy and legal practice. Labour laws had existed for centuries, and Industrial Tribunals were established in the 1960s (Deakin et al. 2021). However, under the Employment Rights (Dispute Resolution) Act 1998, their name was changed to Employment Tribunals, and new legislation provided increased jurisdiction (Employment Rights Act 1996, National Minimum Wage Act 1998, Employment Relations Act 1999, Employment Tribunals Rules of Procedure 2004, also coinciding with new anti-discrimination laws in 1995 and the Equality Act 2010).

Commercial and company law curves rise less compared to the topics mentioned above (Figure 2). This may mirror that formal court-based dispute resolution is costly and time-consuming, leading to more out-of-court resolutions in company and commercial matters (Steffek 2013, 38). On the other hand, intellectual property issues are less often addressed as contractual issues before courts in the United Kingdom. Despite a few levelling out of topics, no issue in cases has dropped before the courts. Likely, this can be traced to increased reporting of cases, the litigiousness of society as well as the pervasion of the law and formal dispute resolution processes.

Our findings demonstrate the rates at which cases of different functional subject matters involving contracts have appeared before courts in the United Kingdom. Such observations exemplify how this dataset reveals important empirical information to aid legal research and policy making. The new topic model can further be used by researchers to automatically characterise new cases without the need to re-train the model. The software and racing bar chart, are available on the GitHub link: <https://github.com/AhmedIzzidien/CourtTopics>

## Limitations

At this research stage, we have identified cases that include references to contracts rather than identifying cases addressing contractual conflicts specifically. In future work, we will develop a methodology to identify cases focused on contractual conflicts. The results are influenced by the numbers of cases per court in the Cambridge Law Corpus over time. Therefore, it depends on the reporting of cases and the inclusion in the dataset. For instance, there are considerably more cases heard and reported in recent history compared to earlier periods, say the beginning of the 20th Century.

In Step 2 of the model development, the LDA represented each topic as a bag-of-words. The functional topic is not always clear. This is a limitation of using this method and can be improved with finer model parameter tuning. One challenge identified in Step 3 of our methodology is how to best manually assign the functional topics. We developed and defined the functional subject matters drawing from legal expertise. Some functional areas could have been further developed with subtopics or were distinguished to a more granular degree. For instance, we split some functional areas into different categories than what could have been summarised as a commercial topic. For further analysis, we would like to classify each court case by the doctrinal contract issue, for example. We leave this to further work.

**Acknowledgements:** This research has been funded by the Cambridge Language Sciences Incubator Fund and Cambridge Humanities Research Grants Scheme (Tier II). **Corresponding authors:** Ahmed Izzidien ([ai297@cam.ac.uk](mailto:ai297@cam.ac.uk)) and Felix Steffek ([fs256@cam.ac.uk](mailto:fs256@cam.ac.uk)). <sup>1,2,3</sup> Faculty of Law, University of Cambridge.

## References:

- Acas. 2021. “Estimating the Costs of Workplace Conflict.” <https://www.acas.org.uk/costs-of-conflict>.
- Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3 (Jan): 993–1022.
- Deakin, S. F., Zoe Adams, Catherine Barnard, and Sarah Fraser Butlin. 2021. *Deakin and Morris’ Labour Law*. Seventh edition. Oxford, UK: Hart.
- Rehurek, Radim, and Petr Sojka. 2011. “Gensim–Python Framework for Vector Space Modelling.” *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3 (2): 2.
- Sloan, Brian. 2020. *Borkowski’s Law of Succession*. Fourth edition. Oxford: Oxford University Press.
- Steffek, Felix. 2013. “Principled Regulation of Dispute Resolution: Taxonomy, Policy, Topics.” In *Regulating Dispute Resolution: ADR and Access to Justice at the Crossroads*, edited by Felix Steffek, Hannes Unberath, Hazel Genn, Reinhard Greger, and Carrie Menkel-Meadow, 33–62. Oxford, United Kingdom: Hart Publishing.
- Zeleznikow, John. 2021. “Using Artificial Intelligence to Provide Intelligent Dispute Resolution Support.” *Group Decision and Negotiation* 30 (4): 789–812. <https://doi.org/10.1007/s10726-021-09734-1>.