TAX4026
Computational Science of Taxation

Course Book 2019/2020

Introduction to the course

Course aim

The aim of Computational Science of Taxation is that students learn to think interdisciplinary between tax and technology. Computational science is a multidisciplinary field and it aims to understand complex systems by developing computational models and simulations. Computational taxation is to tax law what bioinformatics is to medicine and econometrics and business analytics are to economics. The focus question of this course is how computational models and methods may help to understand and improve the tax domain and complexity in taxation?

Students who successfully complete this course will be able to build bridges between the tax domain and the technology domain. They will have the conceptual knowledge and personal competences to be able to co-create innovative computational tax solutions and work in multi-disciplinary teams of tax lawyers, business & public policy advisors, and computer and data scientists.

This course is part of the national Tax and Technology series of courses organized by Maastricht University, Vrije Universiteit Amsterdam and Tilburg University. For more information, visit taxandtechnology.com.

Intended learning outcomes (ILOs)

Upon successful completion of this course, a student is able to:

1. Describe and explain major historical, current and future developments in computer science and their impact in the past or potential to reinvent the tax domain of the future.
2. Explain taxation as a computational model;
    a. Translate taxation problems as computational models;
    b. Describe, explain and apply the data science process;
    c. Describe and explain how computational models can assist in reasoning about taxation problems.
3. Explain how computational tax models can be valorised in practice and/or have social impact.
4. Create and explain (written and oral) a proof of concept of a computational tax solution to peers and experts.
5. Work in a team on a computational tax project.

Teaching and learning activities (TLAs)

Classes

Classes take place on Tuesdays at the Faculty of Law in rooms B0.115 and B0.118. We start at 11:00 sharp, work until 12:30 and have a lunch break until 13:30, after which we resume work until 15:30. There is no class on 25 February 2020 due to Carnival (23-25 February 2020). The class on Tuesday 17 March 2020 is a hackathon.

Classes comprise mainly presentations, class discussions, simulations and workshops. In general, the time between 11:00 and 12:30 will include more presentations and interactive discussions and the time between 13:30 and 15:30 will predominantly have a workshop format with hands-on activities.

Project-based learning

In addition to your work for the classes you will work on a team project. This project requires you to present a tax problem, think about computational methods to model the problem and explain how your ideas can be valorised in practice or can have social impact. You will be assessed on the team paper that presents your project and an individual video pitch that sells your project.

Project teams

Project teams will be made during the first class by the teaching team. No discussion possible: make it work.

Required reading

The required reading is indicated for every week. Read it before you come to class. As you will notice, there is only required reading for weeks 1 to 5. The required reading for weeks 6 and 7 will be announced, if any, during the course. Almost all sources can be accessed online, most are available through the UM e-library or Google Scholar. If you encounter broken links in this course book, then follow a simple two-step procedure. First, search the source yourself. Second, notify the teaching team and, if available, send along the new link.

Teaching team

Dr Marcel Schaper (*course coordinator*)
Associate Professor of International Taxation and Tax Technology
mgh.schaper@maastrichtuniversity.nl
Office A2.012-B, Faculty of Law

Mr Joey van de Pasch, LL.M
Lecturer
jjghm.vandepasch@maastrichtuniversity.nl

Mr Maniek Santokhi, M.Sc
Lecturer
m.santokhi@maastrichtuniversity.nl

Assessment

Assessment comprises three elements (weight in final course grade):

1. An individual digital exam to test your theoretical knowledge (50%) on Thursday 2 April 2020 at 13:00-16:00 in the MECC (Westhal). For more information about digital exams, visit the dedicated website.
2. A written team paper reporting your team project (30%); and
3. An individual oral pitch (on video) of the group project (20%).

These three elements are graded with an integer on a 1-10 scale. The course grade is determined by weighing the grades and rounding this weighted grade to the nearest integer. A passing grade is a course grade of 6 or higher. Scores for the three elements remain valid in the academic year 2019/2020.

The deadlines are:

- Draft team paper to be submitted on Tuesday 17 March 2020 at 15:30 at the latest.
- Final team paper to be submitted on Friday 3 April 2020 at 23:59 at the latest.
- Individual video pitch to be submitted on Friday 3 April 2020 at 23:59 at the latest.

The technical details of submission will be communicated during the course.

What if I fail the course?

If you fail the course, you can only retry elements for which you scored insufficient (5 or lower). If you pass the course, you cannot retry any element, even if you have an insufficient score for one or more elements.
If you failed the course and scored a 5 or lower for the exam, you may resit the exam (written or oral at the discretion of the teaching team) and if you obtain a higher score, this higher score replaces your prior score.
Since the paper and pitch are a team effort, the resit opportunity allows you to make individual improvements if you failed the course and failed any of these two elements. However, because of the prior team effort, a resit score of 6 is the highest score you may be given in the resit. If you failed the course and scored a 5 or lower for the group paper and/or pitch, you may make individual improvements to the paper or make a new video. If these improvements are such that the quality of the paper and/or pitch becomes sufficient, you obtain a resit score of 6 for the respective retried element. No higher scores will be given.

Don’t be afraid of mathematics

You will obviously be confronted with some mathematics: the course is called ‘Computational Science of Taxation’. We expect you to understand and learn on a conceptual level. Mathematics can help you in understanding concepts, but if you do not understand the mathematics, this does not limit your learning experience in this course.

BYOD

Bring your own computer device to class. You really do not need to have the most recent MacBook Pro; any recent notebook suffices. Make sure that it is fully charged before you come to class. There are much fewer electricity sockets available in the tutorial rooms than there are students in the class (the faculty is a monumental building which makes it beautiful but also logistically difficult).

Code of Conduct

Maastricht University, Faculty of Law

The Faculty of Law provides top quality challenging and rewarding education. It seeks to do so in an exceptionally inspiring environment for both students and staff. In order to create such a safe study and working environment the following code of conduct is established that applies to all students enrolled at the Faculty of law regarding activities relating to their course of study:

1. Respect

The Faculty seeks to provide an inclusive, diverse non-hierarchical learning community that fosters both individual and collective learning and growth. This community is based on mutual respect. Students shall respect ethnic, sexual, religious and cultural diversity in the academic community. Students expressly engage in a collaborative and respectful attitude towards fellow students, staff members and all other professionals encountered during the course of the study.
Students shall at all times refrain from the use of physical, verbal, psychological violence or violence of any other kind in their relationships with fellow students and members of staff.

2. Responsibility

The student is an active member of the student community of Maastricht University. This expressly includes being an active participant in education that is in line with the UM teaching philosophy and refraining from activities that undermine that philosophy.
The student approaches the educational experience with an open mind and broad interest and takes in learning opportunities that present themselves in an active and constructive manner.
The student strives for improvement, is able to incorporate criticism and is able to provide and receive feedback. Students are willing and able to reflect critically on their own thoughts and behaviour.

3. Academic and scientific integrity

The student refrains from any form of fraud in all study related activities. Students only connect their name to work that they have effectively contributed to. The student recognises the contribution of other students in group work.
Students represent themselves as students in external contacts and will not accept responsibilities that supersede their developing knowledge and competences.
This article applies without prejudice to the Code of Conduct for Scientific Integrity as adopted by the European Science Foundation and the Nederlandse Gedragscode Wetenschapsbeoefening as adopted by the VSNU.

4. Confidentiality

Students expressly accept the obligation to treat any file, case or other documentation or information that they are confronted with in the course of their study as confidential. The student respects, at any time, the privacy of individuals concerned, including fellow students and staff members.

5. Use of ICT facilities and social media

Students conform to the UM rules on acceptable use of ICT facilities. Students expressly conform to a responsible use of social media, including but not limited to matters relating to points 1 and 4 of this code of conduct. Students respect the intellectual property of the UM. It is expressly prohibited for students to place images or other personal information of fellow students and staff members on social media or to publish such information in any other form.

6. Legal community

Students understand and accept the fact that they are part of the legal professional community and therefore conform to the ethical, legal, procedural and other standards that are commonly shared in the legal professional world. Students conform to these standards in their presence, behaviour and use of language.

7. Enforcement.

The student will abide by the norms that follow from this code as described above and will fully accept the consequences of not doing so. Those consequences may include, but are not limited to, disciplinary proceedings that may result in measures described in articles 7.42a and 7.57h WHW and the UM Enrolment Provision, namely denying access to buildings, premises and/or facilities of the UM or ending the enrolment of the student concerned. Enforcement of article 7.57h WHW will take place in accordance with the rules laid down in the Algemene huisregels en ordemaatregelen van de Universiteit Maastricht.

8. Adoption and Entry into force.

This code of conduct has been adopted in accordance with article 2.1.3 sub h of the UM Mandatenregeling by the Dean on behalf of the Executive Board on 25-06-2019 [RU19.00251]. This code of conduct enters into force on 25-06-2019. This code of conduct is drafted in both Dutch and English. The Dutch version is deemed original.

Week 1: The History of Computational Tax

Within economics it is accepted that mathematical knowledge has a complementary effect on providing answers to fundamental economic issues (e.g. the balance of supply/demand and optimisation for markets). Eventually, this insight has developed into a field that we now call econometrics. More or less the same applies to the field of biology, where bioinformatics is the result of incorporating mathematics and computational science to aid in answering the fundamental research questions.

The impact that technology can have on tax has been known for many years. Yet a formal enrichment of the discipline by mathematics and computational science is a recent development spurred by the current technological climate of cheap and readily available data and computing power. Computational science is a multidisciplinary field that aims to understand complex systems by developing computational models to analyse or predict its behaviour. During this course the fundamental tax problems that potentially can be solved by using technology will play a key role and are of great importance: we work to uncover a new field of computational science of taxation.

During the first week, the aim is to get an overview of the technological developments of the past 70 years. From the early attempts of computing, the introduction of the integrated circuit in the 1950s, the first Personal Computer, the explosion of data, all the way up to today, when computational science, especially Artificial Intelligence, seems to be revolutionizing almost everything around us. Key to this chronological introduction to computing is the inevitability of why history unfolded itself the way it did.

This unfolding is characterized by the growth of computer power, as dictated by Moore's law, and the growth of data, as dictated by our creation and consumption of it. The discrepancy in growth, where we are not able to process all data, forces us to be smart about the way we cope with this. In comes computational science, a discipline that provides the tools and techniques to deal with this discrepancy. Tax too is a complex system not immune to the data growth and hence benefits from incorporating these smart approaches in a fundamental way.

Required reading (read in this order):

- C. Tolga, *Introduction to bioinformatics*, miRNomics: MicroRNA Biology Computational Analysis, Humana Press, Totowa, NJ, 2014, p. 51-52. (available through Google Scholar)
- P. Ceruzzi (2003). *A history of modern computing* (2nd ed., History of computing). London, Eng.: MIT Press. Read: 1-12 (available at Inner City Library: SB QA 69.03)
- R. Schaller, *Moore's law: past, present and future*, IEEE spectrum 1997, p. 52-59. (available through Google Scholar)
- S. Russell, Artificial intelligence: a modern approach, Pearson Education Limited 2016, p. 1-30. (available through Google Scholar or at Inner City Library: SB QA 69K1)

Homework: Your Computational Tax Problem

Think about a problem within tax that is, in your opinion, solvable computationally (i.e. with technology). Write a one page paper about the problem within the field of tax and your proposed solution and take the ethical part in consideration as well. Make sure to hand in your one pager before the first class. The detailed submission requirements will be communicated separately. During class, you will discuss your ideas in your project team. You will select one idea to pitch to the entire class in not more than 5 minutes.

Be creative and think outside the box! We challenge you to come up with a solution that most probably pushes the boundaries of what is possible. During class and the upcoming weeks, we will touch on these solutions again and discuss their feasibility further.

Week 2: Computational Simulation of Taxation

Computational thinking is a problem-solving method to express problems and their solutions in ways that a computer could execute[1]. It involves abstraction, automation and analyses. This week, you will learn how to express taxation problems as computational simulations and study different computational approaches to solve them.

Current simulation models for tax, used by for instance the OECD and its Dutch counterpart CPB, are based on a technique called Microsimulation. These models use individual economic units (e.g. a single person or family) and apply the true tax calculations and consequences. Yet, these group of individual economic units are just a subsample of the total (the entire economy) it tries to simulate. The results are ballooned to true proportions according to the representation of units within the sample. Then a future calculation (say an output for 2030) is computed by assuming how certain properties of the individual units will evolve over time. Popular models for this are TRIM3 and MICSIM.

Microsimulations are a very straightforward approach to simulating (economic) situations. We can further build on these with concepts from computational science: markov chains, reinforcement learning, physics-based approaches, evolutionary/genetic algorithms etc. These agent-based techniques will be applied to tax related propositions.

Tax evasion is such a proposition that will be explored. One of the computational models we will be using comes directly from physics: the *Ising model of ferromagnetism*. Yes, you can model the tax evading behaviour of taxpayers as a magnet. An agent can be an individual or a group and our agents are taxpayers. These agents have, in the simplest model, a binary choice to define their behaviour: comply with tax obligations or evade. The Ising model is very useful in modelling the behaviour of agents with a binary choice as a 'spin' of two states, just like the 'north' and 'south' magnetic dipole moments resulting from atom spins in magnets.

Required reading (read in this order):

- https://boreas.urban.org/T3IntroMicrosimulation.php.
- Computational Modelling: Technological Futures, p. 13-48.
- MICSIM: A behavioural microsimulation model for the analysis of tax-benefit reform in the Netherlands, chapters 1, 2, 3, 6 and 7.
- M. Mitchell (2009). *Complexity: A guided tour*, Oxford England: Oxford University Press. read: chapter 9 and 15. (available through UM e-library).
- Zaklan, G., Westerhoff, F., & Stauffer, D. (2009). Analysing tax evasion dynamics via the ising model. *Journal of Economic Interaction and Coordination*, 4(1), 1-14. doi:10.1007/s11403-008-0043-5. (available through UM e-library)
- M. Pickhardt and A. Prinz (2014). Behavioral dynamics of tax evasion - a survey. *Journal of Economic Psychology,* 40, 1-19. doi:10.1016/j.joep.2013.08.006. read: 1-5. (available through UM e-library)

[1] Wing, Jeannette, *Computational Thinking Benefits Society*, *40th Anniversary Blog of Social Issues in Computing, 2014.*

- J. Ruhl, D. Katz, M. Bommarito, *Harnessing Legal Complexity*, Science, 31 Mar 2017: Vol. 355, Issue 6332, p. 1377-1378, DOI: 10.1126/science.aag3013. (available through UM e-library)

Workshop: working with simulation models

Students will be exposed to the inner workings of several simulation models implemented in Python. In a controlled environment the student is expected to gauge the parameters and assess its response. No particular math or coding ability is required.

Week 3: Taxation as Data Science

In computational science, we try to understand the dynamics of a complex system by simulating its behaviour, and from that extrapolate to make predictions. Yet, with data science it is all about extracting patterns from (big) data originating from a complex system in order to predict future behaviour. This shift in the resulting computational models and methods being used is what will be discussed this week.

Core to this week is the data science pipeline. This pipeline describes all necessary steps needed in order to turn data into insight. First data is collected from all relevant sources. After its storage the data is treated in order to apply a particular model over it. Lastly, the gathered insights from the model ought to be reported. We will be exploring the practical implementation of such a pipeline as well as review the inner workings of several computational models most often used within the tax data science literature.

Finding the optimal tax revenue is one of the propositions we will be using in a data science context. This problem will be treated more from an econometric standpoint and thus concepts such as a utility function, regression and heteroscedasticity will be discussed. We will contrast this approach to that of building a machine learning model to quantify this 'tax gap' via actual tax audits in order to find the optimal tax revenue.

Part of the non-optimality is also due to tax fraud. Similar tax declarations should yield a similar tax base. By clustering over features extracted from these tax declarations we can quantify this similarity and hence find fraud. These powerful clustering techniques will further be explored in the context of finding similarity in tax burdens across European countries.

Besides tax fraud the effects of double tax treaties on foreign direct investments in the context of 'treaty shopping' will also be touched. If the international tax system is treated as a network, where its nodes are either a parent, host or intermediary countries each having their respective relevant withholding tax rate and corporate income tax, we can find the shortest path via the edges between the nodes. This shortest path would then indicate the minimum tax costs of repatriation of dividends over the network of countries and give insight into where to 'shop'.

Required reading (read in this order):

- J. Kelleher and B. Tierney (2018). *Data science*, MIT Press, Chapters 1-3. (available through UM e-library)
- S. Skiena (2017). *The Data Science Design Manual*, Springer, Chapter 1. (available online)
- Introduction to Regression
- D. de Roux et al, *Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach,* Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2018, skip paras. 3.1 and 3.2. (available online)

- M. Van't Riet, A. Lejour, *Ranking the stars: Network analysis of bilateral tax treaties*, No. 290, CPB Netherlands Bureau for Economic Policy Analysis, 2014, p. 1-33. (available online)

Workshop: Python Notebook

The data science pipeline will be the focus during this workshop. Students will be challenged to apply their knowledge with respect to this week's theory in practice. Concretely this means that in a controlled Python environment the students are going to be exposed to a data science pipeline within a tax context. No particular math or coding ability is required.

Week 4: Computational Reasoning

In the previous two weeks we have explored simulation models and data-driven models. Both types of models try to find patterns or predict behaviour mostly through statistical means. Yet, what if we build models purely based on causal reasoning on a knowledge base instead of correlation? This week we will explore these causal systems and their applicability on tax and text-based analytics.

Reasoning starts with basic propositional logic. It is a language where we combine terms through expressions such as 'not', 'and', 'or', 'if, then'. Through this combination we can reason about whole systems of truths. We can extend this language to first-order logic which allows for even more complex truth modeling. We will also discuss fuzzy-logic, used for capturing imprecise information, and introduce ontologies for knowledge representation. These languages are the foundation of so called Expert Systems. These systems (models) encode human knowledge in order to be applied to practical problems.

One of the practical problems to be explored is TAXMAN, an early attempt of creating a tax expert system. This system is capable of determining whether a particular corporate transaction is of a certain type as defined by the US Revenue Code Section. We will review its inner workings as well as discuss its biggest shortcoming: information must be presented to the system in a very strict format.

This strict format does not fit reality. Tax law is encoded in textual form instead of strict propositions and interpreting that text is the base of the fiscal profession. With the help of ontologies we will look at how to interpret text in a less rigid fashion. We will contrast this type of text analytics with the machine learning based 'Natural Language Processing' and review several studies that make use of this.

Lastly we will discuss the marriage between reasoning and machine learning. This relatively new (in vogue) field called Neural-Symbolic Computing aims at integrating two most fundamental abilities: the ability to learn from the environment (machine learning), and the ability to reason from what has been learned (logic).

Required reading (read in this order):

- L. McCarty (1977). Reflections on "taxman": An experiment in artificial intelligence and legal reasoning. *Harvard Law Review*, 90(5), 837-893. (read: 837-850). (available through UM e-library)
- B. Alarie, A. Niblett and A. Yoon, Using Machine Learning to Predict Outcomes in Tax Law, December 15, 2017. (Available at SSRN: https://ssrn.com/abstract=2855977or http://dx.doi.org/10.2139/ssrn.2855977)
- E. Ash and O.Y. Marian, The Making of International Tax Law: Empirical Evidence from Natural Language Processing (January 11, 2019). UC Irvine School of Law Research Paper No. 2019-02. (Available at SSRN: https://ssrn.com/abstract=3314310)

- A. Yoon (2016). The post-modern lawyer: Technology and the democratization of legal representation. *University of Toronto Law Journal*, 66(4), 456-471. doi:10.3138/UTLJ.4007. (available through UM e-library)
- C. Stevens, *The next generation of legal expert systems - new dawn or false dawn*?, In International Conference on Innovative Techniques and Applications of Artificial Intelligence Springer 2010, p. 439-452. (available through Google Scholar)

Workshop: OECD Model Tax Convention on Income and Capital

Tax logic systems form the core of this week's workshop. By using ontologies, data can be assessed, categorized and translated into insights. During this workshop students have to build their own ontology based on two articles of the OECD Model Tax Convention on Income and Capital. Article 1 (persons covered) and 5 (permanent establishment) have to be translated into an ontology. By doing this, the impact that ontologies can have on the (international) tax practice will become clear. No particular math or coding ability is required.

Week 5: Computational Tax in Practice

This week, a guest from industry will come and talk about their experience of applying computational models in practice. During this talk, students will get a taste of what it really means and requires to bring tax technology solutions to business and clients. From the very initial idea all the way up to the delivery of the end product.

To fully understand the impact that tax technology solutions can have for business and clients, the guest will also elaborate on the day to day activities in the tax practice and what it feels like to be part of the change.

During the talk there is enough time for questions and answers. The details about the guest lecturer will be communicated during the course.

Week 6: Hackathon Week

Hackathon is a collaborative sprint: you make a lot of progress in a short amount of time by working together. The live learning activities on Tuesday are dedicated to your project this week. You will meet with your group and work together to deliver a draft paper by the end of this week's classes (i.e. at 15:30). This is also the time to make the first steps in designing the pitch. The teaching team will be available for brainstorm, feedback and questions.

The draft paper should communicate your computational tax problem, the modelling approach and methods, and your ideas for valorisation and/or social impact. You will need to apply all the knowledge that you learned in this course thus far. You will receive feedback from the teaching team on this draft paper.

The draft paper should be submitted before Tuesday 17 March at 15:30. The technical details of submission will be communicated during the course.

Week 7: The Future of Computational Tax

This last week of the course, we will discuss the future of computing, why it is bounded to the P vs NP problem, and think about the implications it has on Computational Science of Taxation. Given that is a new course (and a new field of science) we will adapt this week's content to the experiences in previous weeks and our interactions with you. Further details will therefore be communicated later during the course.

Workshop: Open Lab

The workshop time (13:30-15:30) is an open lab. That means that you can work on whatever you want as long as it is an activity relating to the course (e.g. revision, paper, pitch). The teaching team is available to help you.

Week 8: Exam Week

The digital exam will be on Thursday 2 April 2020 at 13:00-16:00 in the MECC (Westhal). For more information about digital exams, visit the dedicated website.

Be reminded of the deadlines for the project deliverables:

- Final team paper to be submitted on Friday 3 April 2020 at 23:59 at the latest.
- Individual video pitch to be submitted on Friday 3 April 2020 at 23:59 at the latest.