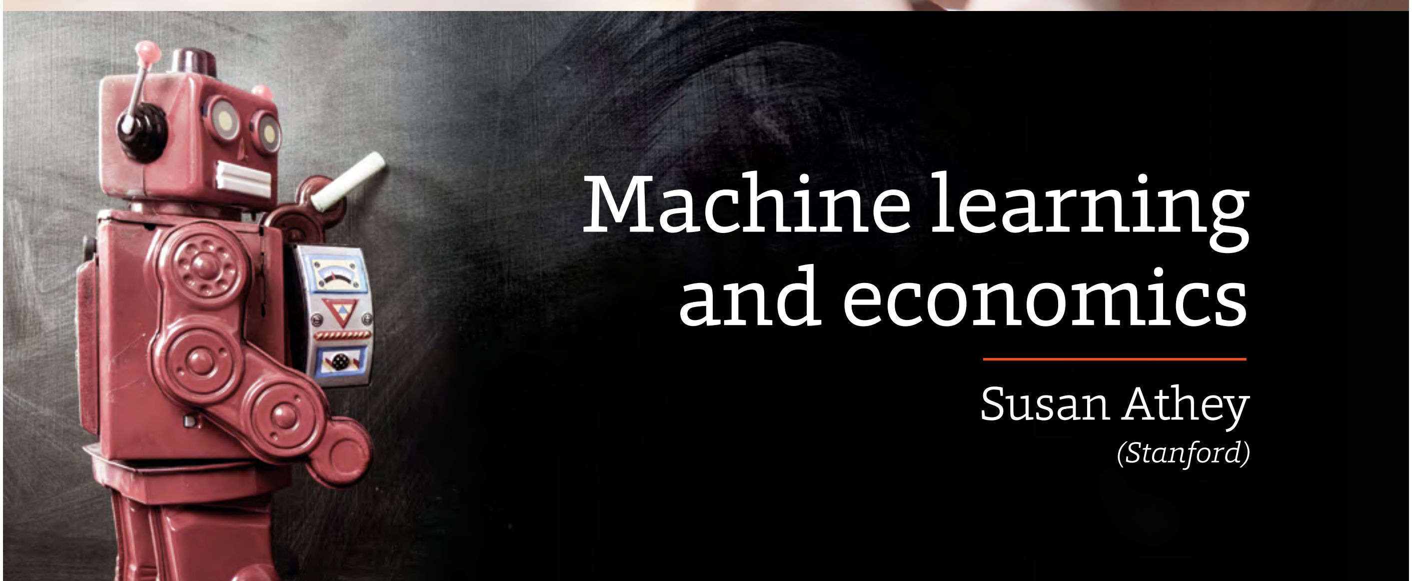
TNIT News

TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

How do search costs shape digital markets?

Daniel Ershov
(TSE)



Machine learning and economics

Susan Athey
(Stanford)

➔ Issue 19 ➔ February 2019

Toulouse School of Economics | Digital Center

Microsoft

A brave new world?

Jacques Crémer

Forward-thinking research is essential to help businesses and regulators keep up with the speed of the digital revolution. In this issue of TNIT News, we feature the work of two TNIT researchers whose work seeks to capture the fast-changing dynamics of the new economic landscape.

TSE's **Daniel Ershov** looks at the effects of consumer search costs on entry, product design, and quality investment in online markets using unique data from the Google Play mobile app store. His results have important implications for anti-trust cases in digital markets.

Meanwhile, Stanford's **Susan Athey** discusses the extraordinary power of machine-learning and AI techniques, allied with economists' know-how, to answer real-world business and policy problems. With a host of new policy areas to study and an exciting new toolkit, social-science research is on the cusp of a golden age. Economics, in particular, will never be the same again.

BYTES

DARON ACEMOGLU WINS MAJOR TSE-IAST AWARD

At Toulouse city hall in October, TNIT member Daron Acemoglu was presented with the 2018 Jean-Jacques Laffont prize.

TSE president Jean Tirole described the MIT professor as one of the most influential and outstanding economists of his generation: *"Daron is also an extraordinary public intellectual but the motivation for the prize is his tremendous contribution in changing the course of several fields in economics.*

He is an outstanding mathematical theorist, an outstanding applied economist, an outstanding empiricist. He's a kind of cyborg matched with an extraordinary, wonderful human being."

Read about Daron's paper on AI in TNIT newsletter Special Issue, December 2018

HEIDI WILLIAMS AMONG BEST ECONOMISTS OF THE DECADE

The Economist has picked TNIT member Heidi Williams as one of its eight best economists of the decade.

These young scholars, it says, are the future of the discipline: *"They display an impressive combination of clever empiricism and serious-minded wonkery. They represent much of what's right with economics as well as the acumen of top American universities in scooping up talent."*

The magazine's December 2018 edition praises Heidi and her cohort for displaying the empirical virtuosity of predecessors such as TSE Nobel laureate Jean Tirole, and for doggedness and rigor in tackling subjects of grave public importance.

Read about Heidi's research on the impact of patents on innovation in TNIT newsletter Issue 16

How do search costs shape digital markets?

by Daniel Ershov
(TSE)

Daniel Ershov is an assistant professor at TSE, with a PhD in economics from the University of Toronto. In this issue of TNIT News, he discusses his recent research, which looks at the effects of consumer search costs on entry, product design, and quality investment in online markets using unique data from the Google Play mobile app store.

Contrary to contemporary expectations in the early 1990s, recent studies show that the internet did not eliminate consumer search costs (e.g., Ellison and Ellison 2009). With the proliferation of product varieties, "discoverability" online is a major concern for consumers, firms and regulators. Consumers complain about not being able to find products in virtually every online market. Firms are concerned about investing in high-quality products and entering into markets where they cannot be discovered by consumers. Regulators are concerned that platforms, by changing search algorithms and consumer search costs, can influence firm entry, investment incentives and the degree of online competition. These considerations played a major role in the European Commission's €2.4 billion fine for Google.

While there is a large existing literature on consumer search costs, it focuses on the effects of search costs on prices and largely ignores non-price effects such as entry or product quality. Non-price effects are important, particularly in the numerous online markets where prices are uniform (e.g., iTunes) or are zero (e.g., SoundCloud). In a working paper titled "The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market", I examine how consumer search costs in online markets affect market structure, product variety, quality, and consumer welfare. I empirically study these effects using new data from the Google Play mobile app store, a large online market where most products are free to download. App stores have a large number of products: thousands of new apps appear every week, and it is costly for consumers to search for new products.

A NATURAL EXPERIMENT

App stores are broadly separated into "game" (e.g., Angry Birds) and "non-game" (e.g., Facebook). Surveys suggest that consumers primarily search for apps by browsing through categories in the app store (e.g., "Travel Apps"). I take advantage of a natural experiment: a re-categorization of part of the Google Play store. In March 2014, Google increased the number of game categories from 6 to 18. Industry observers and anecdotal user responses suggest that this reduced consumer search costs. Before the change, consumers browsing through the categories would see different app types together (e.g., Family Games and Action Games). Consequently, consumers looking for an app type would not necessarily find it easily.

The re-categorization of the store was a surprise to game developers. However, it did not affect the "non-game" area of the store and non-game developers. With non-game apps as a "control" group, I use difference-in-differences to capture several key effects. First, 33% more games enter relative to non-games after the re-categorization. Second, most entry effects are driven by "niche" app types that were more difficult to find before the re-categorization. Third, the quality of the new games - as measured by consumer ratings and other quality proxies - fell after re-categorization relative to new non-games. Last, I also show clear evidence of the search cost mechanism by examining game and non-game downloads: I find that downloads of games increased relative to non-games, and that older game apps benefit most since they became easier to discover.

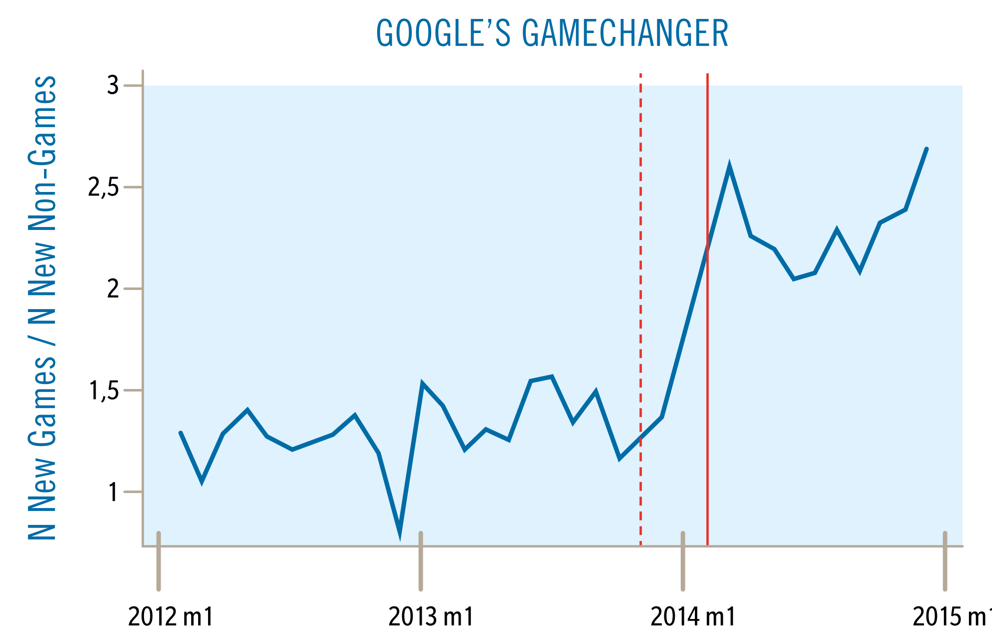
TNITNews
TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

> **"Regulators are concerned that platforms, by changing search algorithms and consumer search costs, can influence firm entry, investment incentives and the degree of online competition. These considerations played a major role in the European Commission's €2.4 billion fine for Google**

MARKET ENTRY

These results confirm existing theoretical predictions but are new to the empirical literature. The entry results are particularly intriguing since they shed some light on previous findings on entry in online markets. For example, previous research such as Brynjolfsson, Hu and Smith (2003) shows that online retail massively increased product variety over brick-and-mortar retail. My findings suggest that search costs alone could be responsible for much of that change, since even a change in search costs on an online platform (rather than the creation of a platform) generates large entry effects.

The overall impact of the re-categorization on consumer welfare is not easily measured since most apps are free. In addition, the reduction in consumer search costs and increase in product assortment are happening simultaneously so it is not obvious what is driving changes in welfare. The different effects of re-categorization can also point in opposite directions. Reduction in marginal search costs for consumers should enhance welfare. Furthermore, if consumers like variety, the additional entry should benefit them. On the other hand, each additional entrant also generates a negative externality due to search costs - a congestion effect that can increase total consumer search costs. Consumers would then not fully benefit from additional variety in the market. Moreover, consumers should also like quality. Conditional on the number of products, a greater share of low-quality products would reduce consumer welfare. In the presence of search costs, a larger share of low-quality products could make it harder to find high-quality products. This would also offset welfare gains.



The graph shows the ratio of the number of game apps on Google Play relative to the number of non-game apps. Despite the large absolute changes in game and non-game entry over time, relative game and non-game entry was nearly constant until Google's announcement in December 2013 that it would increase the number of game categories. This re-categorization took place in March 2014, producing a clear shock in the game/non-game ratio. The time series removes monthly (i.e. December) fixed effects.

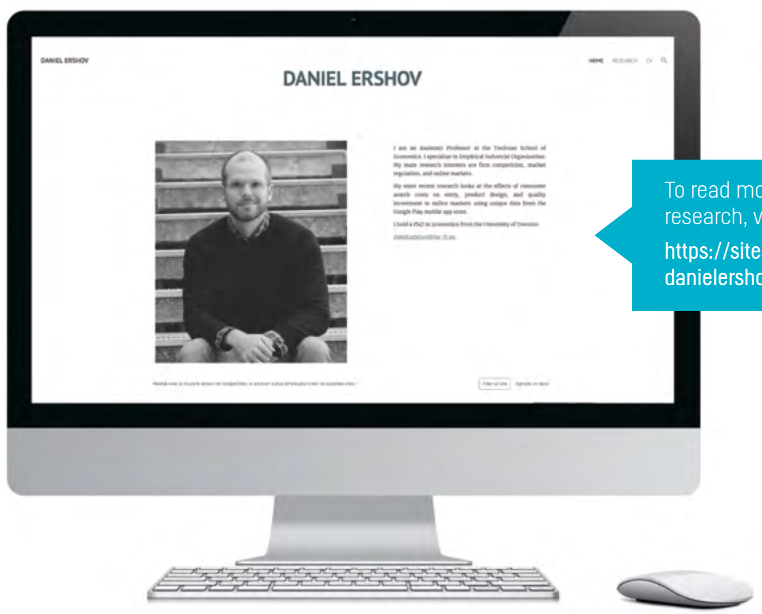
> **These results are the first evidence suggesting that when online consumer search costs increase, consumer welfare can decrease via two channels: a direct decrease due to higher search costs, and an indirect decrease due to a foreclosure effect that reduces product variety**

WELFARE GAINS

To measure and decompose the welfare implications of the re-categorization, I set up a structural model of consumer search and demand, and firm entry. I find that the welfare of each US Android consumer increases by $1-1.4 per month following the re-categorization. This is approximately equal to the price of the median paid app. Since there are over 100 million Android consumers in the US, aggregate welfare gains are over $1 billion per year. I show that over 75% of the welfare gains come from reduced marginal search costs. Consumers also experience large gains from increased product variety, but these are mostly eaten away by the congestion externality. Nonetheless, increased product variety contributes about 25% of the welfare gains, overwhelming the small negative effect of the change in quality.

ANTI-TRUST CASES

These results have important implications for anti-trust cases in online markets. They are the first evidence suggesting that when consumer search costs increase (the inverse of the decrease I observe), consumer welfare can decrease via two channels: a direct decrease due to higher search costs, and an indirect decrease due to a foreclosure effect that reduces product variety.



To read more about Daniel's research, visit **https://sites.google.com/view/danielershov/home**

Machine learning and economics

by Susan Athey
(Stanford)

Susan Athey is the Economics of Technology Professor at Stanford and has been conducting research with Microsoft for many years. She is the first female winner of the John Bates Clark Medal, one of the most prestigious awards in economics. Drawing on her recent working paper titled 'The Impact of Machine Learning on Economics', she talks to TNIT News about the technological revolution that is transforming economics and society.

You have done a lot of work as a consultant to Microsoft. How has this experience affected your interest in machine learning (ML)?

In 2007 I was hired by Microsoft's then-CEO, Steve Ballmer, to help Microsoft beat Google. I landed in the search advertising platform with little prior experience or exposure to a business like this; indeed, there really weren't other businesses remotely like internet search. I collaborated with a group of engineers who had done a lot of work to build the search advertising platform, but mostly didn't have much background in economics, marketplaces, or advertising. But they did know a lot about ML, and they worked in decentralized teams improving specialized ML algorithms. Potential changes were tested using randomized controlled experiments, with many experiments running simultaneously, and many product changes would be approved in a single meeting after reviewing the results of the experiments. My time in the search engine was a life-altering experience for me. I saw the technology, the algorithms, and the experimentation platform. I thought, "Wow, this is going to change everything."

When I went back to my economics colleagues, their reactions ranged from, "Why did you sell out to do consulting?"; to a pat on the head, "You've lost it"; to politely listening. When I specifically described ML, I encountered a lot of negativity, such as "Oh, that's just prediction." I tried to convince econometricians to work in this area. But even my husband Guido Imbens (who is an econometrician) was not interested at that time. Fast forward a few years and now he's working part-time for Amazon, seeing first-hand the power of the new methods and processes, and we're collaborating on a variety of projects bringing in ML techniques.

As digitization moves through the economy, there are so many interesting problems that many different skill sets and approaches will be required. It's very exciting that many more economists are now having their own experiences with tech companies, ranging from search engines, to Uber and Lyft, to marketplaces like AirBnb and Rover. Now, collectively, economists are getting technology and digitization very fast, and we are changing the tech economy while also being profoundly influenced in both questions and methods by the experiences. I recently wrote a paper for the *Journal of Economic Perspectives* called "Economists (and Economics) in Tech Companies" that surveys these recent trends. In October 2018, the National Association of Business Economists hosted a conference and job fair for tech companies to recruit economists. More than 20 tech employers participated, including Amazon and Microsoft, and more than 200 PhD economists submitted résumés.

What impact will ML have on research and policy?

My first prediction is that adoption of off-the-shelf ML methods, for intended tasks such as prediction, classification and clustering, will become pervasive. That's already happening. It's completely standard to use ML for textual analysis in political science and computational social science.

There have already been a number of successful policy applications of ML prediction methodology. Examples by Harvard economist Sendhil Mullainathan with a variety of coauthors include predicting whether an elderly patient will die within a year to determine whether to do a hip replacement operation. Similarly, if you can predict who will show up for court for their trial, you can let more defendants out on bail. ML research by Harvard economists Edward Glaeser, Andrew Hillis, Scott Kominers, and Michael Luca has helped cities to predict health-code violations in restaurants, in order to better allocate inspector resources. There is a rapidly growing literature using ML together with satellite imagery and street maps to predict economic quantities such as poverty, safety, and home values. This methodology can be used to compare outcomes over time at a very granular level. Large-scale imagery and sensors may lead to new types of analyses of productivity and well-being.

Why is ML useful for these types of applications?

Traditional econometric models have been specified "by hand"; in practice, researchers would try dozens of alternatives and select a few to display. This has the advantage of incorporating domain knowledge about the setting; but it has a number of disadvantages. One is that the researcher may not find the model that best fits the data.

My own research focuses on using ML methods to attack problems of causal inference. Applications include estimating how treatment effects or parameter estimates vary with exogenous covariates; estimating optimal targeted treatment assignment policies; and efficiently controlling for confounders when estimating average treatment effects. Numerous job market papers in marketing in 2018 used ML tools, and several economics papers have applied methods I've developed for estimating heterogeneous treatment effects. For example, Günter Hitsch and Sanjog Misra analyze targeting using ML methods.

Can ML improve scientific credibility?

ML improves credibility for several reasons. To start, let's assume the researcher has decomposed their problem such that it is appropriate to use a predictive ML method to solve part of the problem. Then, ML provides a rigorous and systematic approach to find the best model to maximize goodness of fit in a held-out test data set. When the researcher chooses an algorithm, the algorithm in turn chooses the best specification of the model to fit the model (or in a causal model, to best optimize another objective). This eliminates a lot of problems that can arise when a researcher evaluates a lot of models then cherry-picks the ones that give the most appealing results. The researcher can increase transparency, since the researcher can fully describe the algorithm and the model selection can be replicated, while the process a researcher uses to experiment by hand with different models is difficult to document. The model selected by the ML algorithm may be substantially more complex than a hand-selected model, and it may find interaction effects that would be difficult to hypothesize about in advance. Having a model optimized to the data reduces the incentive of researchers to try many alternative specifications without adjusting confidence intervals. When properly applied (for example, using techniques that have been proposed in recent statistics and econometrics papers such as sample splitting or cross-fitting), the ML methods both provide better fit, and also contribute to more reliable confidence intervals (since traditional methods are often prone to human specification searching without correction).

Another area where ML can have a positive impact is that it enables an increased emphasis on stability and robustness to assess the credibility of studies. We can use ML to test a lot of different models; this enables assessing robustness of key parameter estimates across different models. As a result of digitization, we're often operating in settings where there are lots of exogenous changes. Large tech firms release new algorithms every week, and conduct thousands of experiments per year. E-commerce firms and even physical stores change prices regularly, and observational data from scanners and transaction logs can provide data that contains many price changes; this is something I've exploited in a series of research papers combining ML and structural estimation methods. Thus, as digitization increases, we have lots of ways to test the credibility of our models, including credibility of counterfactual predictions.

ML is good at predictions. Why do we need anything else?

Most real business problems or policy problems are not straight prediction problems. Off-the-shelf ML methods predict which customers will click on an advertisement, which consumers are likely to quit, or which restaurants will fail a health inspection. That's not the same as knowing how to allocate resources. The people who click are not necessarily the ones for whom an advertisement is most effective; an advertisement for a travel website might simply remind a consumer to check the status of their flight rather than induce them to make a new purchase. The people who quit your platform are not necessarily the ones to try to keep; some will quit anyway, for example because they are moving out of the service area of the firm or no longer need the service. Some restaurants may have an easier time than others improving their health rating.

The main idea is that, although prediction is often a large part of a resource allocation problem, there is an important gap between units that are at risk and those for whom intervention is most beneficial. Determining which units should receive a treatment is a causal inference question, and requires different types of data. Either randomized experiments or natural experiments may be needed to estimate heterogeneous treatment effects and optimal assignment policies.

The ML community has developed this incredibly effective hammer, designed for prediction and classification problems, and they have been very focused on hitting every prediction nail out there.

TNIT News
TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

Prediction might get you about 80% or 90% there for some problems; for others, where correlation and causality are confused, it could lead you in the wrong direction. For example, if drinking red wine is predictive of longevity, a predictive model might lead people to conclude they should drink more red wine, when in fact it is possible that red wine is harmful to health, but that people who drink red wine are different from non-drinkers in ways that are hard to control for. I wrote an article in *Science* in 2017 calling for more caution in using predictive models. More broadly, interest is growing, both in computer science and the social sciences, in taking that next step to make data-driven resource allocation most effective. And the intersection of ML and causal inference is a very exciting and rapidly growing area of econometrics and statistics.

Has data become more important than theory?

I've been writing a lot of papers recently on the development of new econometric methods based on ML, designed to solve traditional social-science estimation tasks. ML has this feature that because it can be evaluated on a test set very easily, you don't need statistical theory. If all I'm trying to do is use X's to predict Y, then if I have a held-out test set (that is, a subset of data not used to train the model), I directly measure how well that model does at predicting. This is one reason this area has advanced so quickly. There's one agreed-upon metric, which is goodness of fit in a held-out test set; then, the algorithm can be a black box, and you just race to succeed at making better and better black boxes.

That's very different from causal inference, where if the assumptions required to identify a causal effect are not satisfied in your data (for example, there are unobserved confounders that affect both treatments and outcomes), you will get the same incorrect answer in a training set and a test set. Red wine might predict long life in your training dataset, and also in your test dataset, but still red wine might be harmful to health. The methodology of evaluating models on test sets is useful for evaluating performance where the main problem is that models "overfit" to training datasets (which easily happens in the rich world of ML models), but the approach of using test sets to evaluate performance needs to be modified to be useful in evaluating the performance of a causal model. Economists are used to throwing away most of the predictive power of a model to get unbiased estimates of causal effects. It's a very different way of thinking about things. A few research teams, including myself and coauthors, are exploiting this difference to do interesting, new, statistical science, because there was a big hole left to fill. We are finding that insights from statistics about semi-parametric efficiency (estimation when functional forms are not known) can be combined with ML methods to derive algorithms for estimating parameters of interest - things like treatment effects, optimal treatment assignment policies, or demand elasticities - efficiently. ML methods can provide a dramatic improvement in real-world performance of estimation methods, but statistical theory still guides how they should be used.

More broadly, for more complex counterfactual scenarios, like auction designs that have not yet been tried, it is very hard to make significant progress without bringing in theoretical and behavioral models. With a behavioral model combined with estimates of preference parameters, it is possible to make predictions about scenarios that have not been observed in the data. Artificial intelligence (AI) algorithms also potentially need to reason about scenarios that they have not encountered before; theoretical models can help ensure that counterfactual predictions are reasonable.

Will big data change the way we look at the world?

We're going to see a resurgence of the science of productivity and measurement, along with new methods in the design and analysis of large administrative data sets. We will see attempts to bring together disparate sources to provide a more complete view of individuals and firms. Behavior in the financial, physical and digital worlds will be connected, and in some cases ML will be needed to match different identities onto the same individual. We will observe behavior over time, often with high-frequency measurements. For example, children will leave digital footprints throughout their education, interacting with adaptive systems that change the material they receive based on previous engagement and performance.

What effect will ML have on the way economists work?

There will be changes to the organization, funding, and dissemination of economics research. Scholars who do a lot of complex data analysis have already begun to adopt a "lab" model more similar to what is standard today in computer science. A lab might include a post-doctoral fellow, multiple PhD students, pre-doctoral fellows, undergraduates, and full-time staff. Such labs are expensive, so funding models will need to adapt. One concern is inequality of access to resources.

Within a lab, we will see increased adoption of collaboration tools such as those used in software firms; for example, my generalized random forest software is available as an open-source package *(https://github.com/grf-labs/grf)*. Users report issues through GitHub, and can submit requests to pull in proposed changes or additions to the code.

There will be an increased emphasis on documentation and reproducibility, even as some data sources remain proprietary. "Fake" data sets will be created that allow others to run a lab's code and replicate the analysis.

How will economists interact with digital-age policymakers?

We will see changes in how economists engage with government, industry, education, and health. The concept of the "economist as engineer" and even "economist as plumber" will move beyond its traditional home in fields like market design and development. Digitization will bring opportunities for economists to develop, implement and evaluate policies - such as farming advice, online education, health information, government service provision, and personalized resource allocation - that can be delivered digitally. Feedback will come more quickly and there will be more opportunities to gather data, adapt, and adjust.

What are some of the most promising ML innovations?

ML methods lend themselves to incremental improvements, which can be evaluated using randomized controlled trials. Firms like Google and Facebook do thousands of randomized controlled trials of incremental improvements to ML algorithms every year. An emerging trend is to build the experimentation right into the algorithm. Multi-armed bandits balance exploration and learning against exploiting available information about which alternative is best. Bandits can be dramatically faster than randomized controlled experiments because their goal is to find the best alternative, not to accurately estimate the outcome for each alternative.

Balancing exploration and exploitation involves fundamental economic concepts about optimization under limited information and resource constraints. Bandits will help social scientists to optimize interventions much more effectively. Statistical analysis will be commonly placed in a longer-term context where information accumulates over time.

I've been doing research about "contextual bandits", which try to learn the best personalized policies. Contextual bandits involve counterfactual inference and estimation, because bandits create data with non-uniform treatment assignment probabilities, and the goal is to learn the counterfactual mapping between individual characteristics and expected values from alternative treatment assignments. Thus methods from the causal inference literature, like doubly robust estimation, can improve performance.

Contextual bandits have many potential applications in social science. They can be used to prototype field experiments in behavioral economics, as well as to learn personalized treatment assignment policies in applications like health and education.

How should we prepare students for a digital future?

Within 10 years, most students will enter college (or business school) with extensive coding experience. Many will take coding and data analysis in college, and teaching will need to complement this material. In the short run, more students may arrive at econometrics classes having been exposed to ML, and thinking that data analysis is just about prediction or classification problems. They may have a cookbook full of algorithms, but little intuition for how to use data to solve real-world problems. Given the unique advantages economics has at these methods and approaches, many of the new data-science programs are going to realize they will have more marketable and useful students if they bring in economists and other social scientists. At the same time, econometrics will need to focus more on its comparative advantage and on important methods where other fields have pulled ahead.

What can ML experts learn from economists?

At Microsoft Research and in Microsoft's cloud business they're creating a new toolkit for causal inference in ML, that can help businesses make pricing decisions and so on. You can't do that off the shelf with the stuff that you learn in your PhD in ML at Stanford: without the right training and the appropriate types of data, it is difficult to know how to use the tools to set prices, evaluate the effectiveness of an advertising campaign, or anything like that.

More broadly, AI is a very important new technology for the economy. AI can be thought of as software that attempts to actively learn about the environment, and in particular learn which decisions or actions to take in different circumstances. In doing this, the AI needs to draw inferences from past data. This is a problem of causal inference. I've been working on incorporating insights from the statistics and econometrics literatures on causal inference into AI algorithms.

There's a lot of confusion about the interpretability of models in computer science and in ML - but some aspects of those problems can be considered through the lens of how economists think about regression models when some important variables are unobserved. One way to make a model more superficially interpretable is to make it simpler. For example, in a regression context, to include a small set of variables. An ML technique known as regularized regression intentionally simplifies a regression by setting many coefficients to zero. However, in social sciences, we have long understood that leaving out a covariate can make some kinds of interpretation more complicated. If the left-out covariate is correlated with those that remain and correlated with the outcome of interest, then regression coefficients will incorporate a mixture of the relationship between the outcome and included covariates, and the relationship between the outcome and omitted variables. So the simple regression may appear to be more interpretable (because it is simple), but what is actually happening is misinterpretation. This is important to social scientists because in social-science data, it is typical that many attributes of individuals or locations are positively correlated - parents' education, parents' income, child's education, and so on.

Another notion of interpretability that is very familiar to social scientists is causality - if a model estimates a causal effect, the framework of causal inference tells the researcher directly how to interpret the model. The model tells you the causal effect of an intervention; for example, the impact of raising prices on consumer demand. Causal models (or more precisely, the part of a more complex model that represents a causal effect) are by definition interpretable, because the framework specifies the interpretation precisely and mathematically.

Across many different applications of ML, we will encounter many other considerations. One area that has attracted a lot of attention is bias and fairness. How can fairness constraints be defined? What type of fairness is desired? Is it possible to achieve multiple notions of fairness at the same time? How can we ensure that ML and AI respect desired notions of fairness?

Another area for future research concerns how to make ML models less prone to manipulability. For example, if certain behavioral patterns help a mobile-phone user get a loan, the consumer might start visiting different areas of a city. Or if resources are allocated to homes that look poor via satellite imagery, homes or villages might modify aerial appearances to make their homes look poorer.

Is ML breaking down academic boundaries?

There will be a substantial increase in interdisciplinary research. I have a lot of co-authors now in lots of different areas - management science and engineering, computer science, civil engineering and transportation - because one person can't be an expert in all aspects of a problem, ranging from computational issues, to domain and industry knowledge, to data engineering. In addition, as digitization spreads, all disciplines will gain a much greater ability to intervene in the environment in a way that facilitates measurement and causal inference - when people get most of their information digitally across areas like health, education, shopping, and travel, there will be opportunities to experiment with that information provision to learn how to make it more efficient.

We will also see more interdisciplinary majors; Duke and MIT both recently announced joint degrees between computer science and economics. The curricula will evolve from a truly engineering base to being more problem-solving. That will increase the demand for economists generally, but also change the way we teach and research.

AI applications will succeed or fail based on how well they work in context. The factors that impact this kind of success are often a combination of detailed domain knowledge and social science. Thus, a combination of social science, statistics, computer science, and engineering will be important.

Will ML make the world a better place?

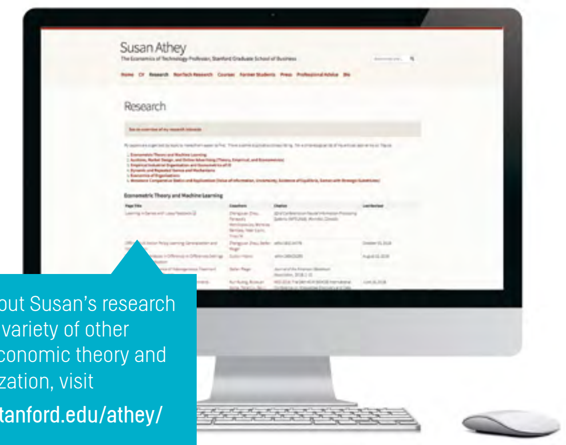
We will see a lot more research into ML's societal impacts. ML does improve the world in many ways, helping people find information more efficiently, helping them find the best products for themselves, and helping them learn more efficiently. At the same time, large-scale regulatory problems will need to be solved. Regulating autonomous vehicles and drones is a key example – such technologies have the potential to create enormous efficiency. There will also be many disruptive changes. I talk to many firms that are about to lay off a lot of workers. We do need to worry about this. We're going to have a lot of research about the impact of AI and ML on the economy, and how to help workers with transition.

We also need to set out a research agenda that is very close to the literatures on ML and AI, but that focuses on making AI safe and effective for humans. This agenda includes questions about fairness, unintended consequences, equilibrium behavior when AI interacts with human systems, robustness, stability, and many other issues. This agenda will inform business organizations, governments and regulators. Economists and other social scientists are very well positioned to contribute to this literature.

How would you summarize the new challenges for economists?

Economics will be profoundly transformed by ML. We will build more robust and better optimized statistical models, and modify algorithms to have other desirable properties, ranging from protection against over-fitting and valid confidence intervals, to fairness or non-manipulability. A variety of new research areas will offer better measurement, new methods, and different substantive questions. We will grapple with how to reorganize the research process, which will have increased fixed costs and larger-scale research labs. We will change our curriculum and play an important role in supplying the future workforce with empirical and data-science skills. And we will have a whole host of new policy problems to study.

Many of the most profound changes are well underway. There are exciting and vibrant research areas emerging, and dozens of applied papers making use of the methods. But this does not remove the need to worry about causality and other traditional concerns for economists. As ML automates some of the routine tasks of data analysis, it is even more important for economists to maintain their expertise at the art of credible and impactful empirical work.



To read more about Susan's research on ML, and on a variety of other topics in microeconomic theory and industrial organization, visit

https://people.stanford.edu/athey/research

TNIT News

TOULOUSE NETWORK FOR INFORMATION TECHNOLOGY

- *Scientific Director:* Jacques Crémer
- *Editorial contributions:* James Nash
- *Graphics:* Olivier Colombe
- *Illustrations:* I-Stock

TNIT
Manufacture des Tabacs
21 allée de Brienne
31015 Toulouse Cedex 6 France
+33(0)5 61 12 85 89

www.tse-fr.eu/digital

tnit@tse-fr.eu

The Toulouse Network for Information Technology (TNIT) is a research network funded by Microsoft, and coordinated by TSE. It aims at stimulating world-class research in the economics of information technology, intellectual property, software security, liability, and related topics.

Issue 19 February 2019

Toulouse School of Economics | Digital Center

Microsoft