

nekst>>

vol. 29 no. 4 | summer 2021 | nekst-online.nl

Special

The Optimal Wine Adventure

ASSET  Econometrics



Triangle

DNA - Quadratic
Traveling Salesman
Problem



The Teacher

Hein Fleuren



Report

Move for Charity



Something with a Light and a Tunnel

At the moment, you are probably studying for exams or resits. In that case, I wish you good luck with the upcoming mental breakdowns. Maybe, summer holidays have already started for you, but due to the COVID-19 measures, there is not a lot to do. So, if you are feeling a little bit down, do not worry. No matter where you are or what you are doing, our last edition of Nekst is here to cheer you up. This edition contains nothing but positive vibes and a hopeful outlook for the future.

To start with our specials. The first one describes a topic that fits perfectly with the wine lovers among us econometricians. If you want to know how wine and mathematics relate to each other, I advise you to go to page 7. Next, Nienke and Casper present an optimistic view for our summer by looking at what festivals are possible now that the vaccination program is speeding up. Anne's column about the new Dutch pension system also revolves around a brighter future. Furthermore, we report the amazing Move for Charity event with a small focus on Villa Pardoes. Finally, in The Teacher, Hein Fleuren talks about how optimization can help the world with its fight against global hunger.

So, you can see what I mean with positive vibes, right? We are almost there, and we have so many things to look forward to. Even writing this joyous preface puts me in a good mood. For me, it was a great pleasure to provide you with four editions of Nekst. I wish you an incredible summer!

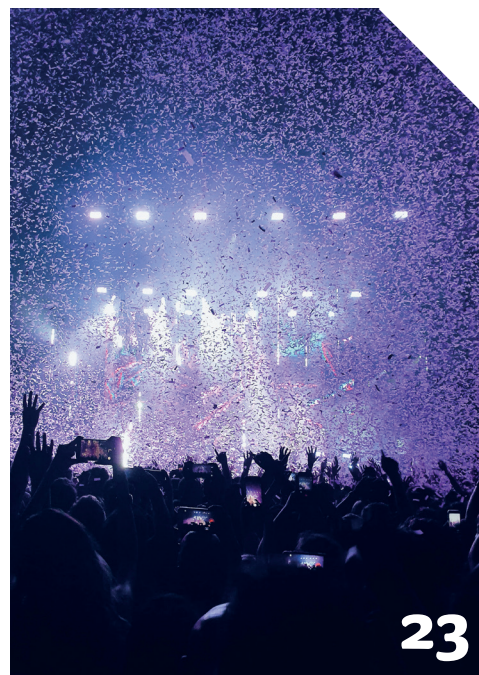
Yours sincerely,

Mylan Tran
Editor-In-Chief

Table of Contents



Move for Charity



Special: Festivals This Summer, Is It Possible?



Triangle

COLOPHON

Nekst is the quarterly magazine of Asset | Econometrics ©2021
Insertion of an article does not mean that the opinion of the board of Asset | Econometrics or the editorial staff is verbalized.

Correspondence

Asset | Econometrics
Tilburg University
Room E 1.10
P.O.Box 90153
5000 LE Tilburg
Telephone: 013 466 27 47
info@Asset-Econometrics.nl
www.Asset-Econometrics.nl
www.Nekst-Online.nl

Editorial staff

Tamara Dert
Casper Heemskerk
Troy de Juncker
Nienke Kempers
Nienke Keuning
Maureen Lacet
Michelle Luijten
Karol Morales Salazar
Jarno Ringhs
Aimee Smarius
Marijn Wolferink



The Teacher: Hein Fleuren



Practical Report



Ik hou van AMW



Application Poster

Advertisements

- cover Flow Traders
- 3 Career Platform Tilburg
- 37 Study Support

Articles

- 4 Dear Members
- 5 Move for Charity
- 7 Special: The Optimal Wine Adventure
- 10 Column: Kuno Huisman
- 11 Triangle
- 15 Ik hou van AMW
- 16 Building Blocks
- 17 The Teacher: Hein Fleuren
- 19 Application Poster
- 21 Application Poster Article
- 22 Column: Anne Balter
- 23 Special: Festivals This Summer
- 25 Practical Report
- 28 Committee Profile: QIG
- 29 Let's Talk
- 31 Quatsch
- 32 Puzzle
- 33 Thank You Letter
- 35 Graduates
- 38 Agenda

Editor-in-Chief
Mylan Tran

Lay-out editor
Dahli Koskamp

Nekst-Online
Patrick Floor
Luuk Sommers

Contributions
Anne Balter
Dominique Bavelaar
Selina Bötzel
Jeffrey Buijk
Luc Geurts
Armanda van Herwaarden
Job Hoven
Kuno Huisman
Frank de Meijer
Juil Schuurmans
Juliëtte van der Velden

Printing
Gildeprint

Circulation
950

ASSET  **Econometrics**



Career
Platform
Tilburg

Meet your future employer here!

1

Visit www.CareerPlatformTilburg.nl

2

Choose your **interest field**

3

Discover **endless opportunities**

Find your dream job

Dear Members,

Approximately a year ago, my year as chairman of Asset | Econometrics started, which means that this is already my last 'Dear Members'. This automatically means that my board year is already coming to an end, and that the academic year 2020-2021 is as well! I hope that you have passed all your exams, achieved the goals you had for this year, and that you are currently enjoying your much deserved summer break.

Before looking forward, I would like to start by quickly looking back at this year. During the year, we have tried to organize as much as possible together with all our committees. We are very proud of all the work every active member put in his or her committee and we would like to stress that without your enthusiasm and hard work, the events would not have been such a success. Of course, I can not mention all events we organized lately, so I would like to tell you about only two of them. On April 21, The Econometricians for Society committee organized a Run for Charity. Around 35 people participated and ran or walked for charity. All together, an incredible amount of 1000 euros was collected for Villa Pardoes! The second event is the always legendary Astrics Beer Cantus, which was organized by the Drinks & Activities committee. The cantus took place in a hybrid form, so you could join with a small group and sing together with the D&A cantus-zangers through Zoom. Over 80 people participated and made the night unforgettable.

Besides looking back, it is also time to look forward! At the time you are reading this edition of Nekst, it will be known for a while which four new board members will be installed as our successors during the Department Members Meeting on Tuesday August 31. We are very enthusiastic about Wout Temmink, Patrick Floor, Joris te Booij,

and Floris Somers that will join the 43rd board of Asset | Econometrics and have the fullest confidence that they will make the next year amazing. We would like to wish them all the best, and let's really hope that next year will look like a normal year and that you will have a lot of possibilities to get to know and meet the new board.

Besides wishing the next board all the best, I would also like to thank my fellow board members. This board year can of course not be described as a regular board year. During this hectic year, we went through a lot together. The year started on a positive note, with a very sunny and almost care-free summer, but then the feared second (and third) wave came. We as a board had to make tough decisions, discuss endlessly about the measures, but most of us all had to stay positive. I can wholeheartedly say that the six of you have made my year. Thank you for all the fun board activities, for always having my back, for all the incredible memories, and for all the laughs we had this year. Thank you for everything, Michelle, Stephan, Juliette, Tjum, Constan-tijn, and Bram.

Lastly, I want to end by saying thank you to every single one of you. Because besides my board, you also were the people that made this year amazing. So, thank you for being part of our beautiful association and contributing in so many ways. Even though our events were not as you were used to, a lot of you found your way to our alternatives. Asset | Econometrics would not exist without its members and I think this strange year has shown that even more than ever.

On behalf of the board,

Juul Schuurmans

Chairman Asset | Econometrics 2020-2021



Asset | Econometrics

A 'Drink' for Charity

On Wednesday April 21, I participated, together with Maaïke Wessels, in the Move For Charity event organized by the Econometricians for Society committee of Asset | Econometrics. It was an offline event, where we walked to raise money for charity, drank beer, and played games with fellow econometricians. It was a great mix of fun activities, while at the same time contributing to society and last but not least, burning the COVID-19 kilos.

For this activity, we had the choice between walking or running. Maaïke and I decided to go for a walk. I personally do not like running at all, so I was happy there was a choice. Before the activity started, we asked our family, friends, and roommates to sponsor us for the walk. Together we managed to raise a nice amount of money for charity. At 15.50 hours, it was our turn to start the walk. We had to gather at the 'Spoonpark' where we received a roadmap and a stamp card for the games. The roadmap explained where the four different stops were, and the stamp card was used to count the number of points we earned during the games. The duo with the highest number of points received a prize.

After around five minutes of walking, we arrived at the first location. Here we had to find a colored egg, which I found rather quickly. I had to make sure that during the afternoon the egg was kept safely. The

first stop also included a drinking game: one of us had to take an 'adje'. It sounded pretty easy, so Maaïke volunteered to do it. However, it was not the regular way of doing it. I had to stand on a ladder and pour the beer into her mouth. This did not go very smoothly, and the beer ended up everywhere. After that was done, we walked to the next location. When we arrived, one of us had to walk a 'parcours' with the aforementioned egg, without dropping the egg while it was balancing on a spoon. The first part of the course was blindfolded, and the other person had to give instructions. Maaïke managed to successfully walk across the course with the egg still intact. Once again, we finished the stop with a beer. We could earn extra points if we did a 'trechteradje', so we did.

The next stop had a quiz that included some general questions about the committee. The questions ranged from guessing the number of ECTS the committee members had to questions about previous activities. After the quiz, we had the option to roll a dice for extra points. Depending on the number we rolled, we had to take that number of beers. Unfortunately, I threw 5. This was the worst number possible, because if we threw 6, the two quiz masters would drink along with us. We finished the beers, and we walked towards the last stop, which took quite a long time. At the last stop, we had a beer tasting, where we had to guess which brand we were drinking. There were two different beers, which we both



Selina Bötzel

Bachelor EOR

Age: 21

guessed incorrectly. We ended the last stop with an 'adje shotgun'.

The day ended where it also started, at the 'Spoonpark'. We handed in the -still intact- egg, and after that, we had to run around the park. We also heard the total amount of money that was raised that day. Before we started the walk, we had to guess what we thought the amount would be, but it was substantially higher than we predicted. Looking back at it, the day was a success. The weather was nice, the activities were a lot of fun, and that all while raising money for charity.●



Moves for Charity

Villa Pardoes

written by **Nienke Keuning**

The Move for Charity event was very entertaining for all the participants. Since this was one of the rare offline activities at the moment, I for one really enjoyed being outside and meeting fellow econometricians while still adhering to the COVID-19 measures. Nevertheless, equally as exciting as the event itself, is the nice amount of money that was raised for charity.

Villa Pardoes is the beautiful charity that the Econometricians for Society (EFS) committee chose to raise money for. It is a wonderful holiday accommodation where children who are very ill can spend a week with their parents and siblings free of worries and full of lively

experiences. The park itself has lots of kid-friendly activities like horse riding and face painting. Also, the kids who stay there can visit the Efteling through a special entrance. As most of us have gone there during our childhood or still like going there, we know how much impact 'het sprookjesbos' can have on little kids. While the children can have a week of fun, the parents who stay there can share stories and bond, which makes it valuable for kids and parents alike. With this accommodation, they are able to give at least 600 children per year an unforgettable experience.

The committee members are mainly focused on helping children and teenagers with their charity work. A previous cause was a game evening for lonely teens. Organizing small things for kids



and teens that have a positive effect on their mental health is very important to the EFS committee. Villa Pardoes was the right choice since a family member of one of the committee members got to go there. Also, they wanted to do something locally, so it would feel more close by.

Thus, Villa Pardoes was the perfect charity for this event, but how much money did it reap? Altogether, the participants raised an amount of €956,50 which was announced enthusiastically by one of the committee members. The board members of Asset | Econometrics then decided they would round this up to €1000. In conclusion, this event could not have gone any better! ●



The Optimal Wine Adventure

Wine or Beer tonight? This question is asked more frequently in our student society than we care to admit. The myth that in a Van Dale dictionary, it was once mentioned that students can never officially be alcoholics, is sadly not true [1]. We are very sorry to ruin your feel-good excuse, but we are also happy to announce a replacement that you will still be able to use even after your student time. It all starts with the right reply: wine of course!

written by **Tamara Dert** and **Maureen Lacet**

That wine makes people feel more sophisticated is a well-known fact, but in a sense, this turns out to actually be true according to Gordon M. Shepherd, Professor of Neuroscience at Yale School of Medicine [2]. "Wine tasting involves multiple sensory and motor systems, as well as central conceptual systems for cognition and memory - and systems for emotion and pleasure", Shepherd says. This is why the brain becomes more stimulated whilst tasting wine than it would, for example, when listening to music or even solving complex math problems, which involve fewer body systems working together. Talking about complex math problems, we still like those and as you know, they can be used everywhere to make life better. Drinking wine for instance, with all of those different sensory systems, is a lot more enjoyable if you have a wine in your hand that you actually like. So, put your beer aside and let's have a look at what you should be drinking.

Predict the wine

When going to the supermarket, it is overwhelming how many bottles there are to choose from. The etiquette can indicate the grape, the region, or sometimes just the chateau. Let's be honest, it is a jungle. So it is time to take a step back to the main question: how can a wine unambiguously be described? Wine has several chemical measures; acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, and alcohol. There is a data set available consisting of red or white wines and their chemical compounds from the Vinho Verde region in Portugal [3]. This data set also contains the variable quality. However, quality is a subjective variable. The rating is on a scale from 1 to 10 and is based on at least three evaluations from wine experts (what a job!). Since all of the wines in the data set are from the same region in Portugal, we need to keep in mind that the data will not generalize to

every wine. How can these wines be connected to math? The following complies with the lines of the article: wine & math, a model pairing by Lars Verspohl [4].

A linear model could reduce relationships between individual data columns to an equation, for example:

*Alcohol * Alcohol Weight + Acid * Acid Weight + Sugar * SugarWeight = Quality* (likelihood to be good).

This is, of course, a very simple example. But, it shows that it is possible! The inputs are weighted by a factor that tells the model how much impact the attribute has on the output. The next question is then, how to determine these weights? We cannot just say: well, let's put more weight on Alcohol, because I like alcohol. No, here is where machine learning is necessary. We need to teach our predictive model how to learn from the data.

We have a training set that contains all the model inputs and the model output variable and we train our model on this data. This means that we give our model this training set and label what we want our model to learn, in our case the quality of the wines. Then, we let our model detect patterns that help conclude our classification into high- (rating 7 or above) and low-quality wines (rating below 7). Next, we want to check how well our model performs on new data: our test data. We let the model predict the output variable based on the inputs and compare the accuracy of this prediction. If our model is accurate enough, then we can use our model to predict the quality of future wines based on these chemical measures. This process is called supervised learning and is basically learning by example for computers.

Next, we need to look at our data. The data set on the red wines contains 1600 wines

and their chemical properties and quality. We are mainly interested in how the physicochemical properties relate to each other and specifically how they relate to quality; we want to classify the wines into high- or low-quality wines (see Figure 1).

An interesting attribute to look at is, of course, alcohol (see Figure 2). Higher quality wines are likely to have more alcohol in this data set.

Next, we want to look if a model can be built that represents the relationship between the variables and quality. To start simple, we only look at an Ordinary Least Squares (OLS) model (see Figure 3) that predicts wine quality only on Alcohol. We would then get the following equation:

$$y = a \cdot x + b$$

y: quality output value

x: alcohol input value

b: intercept

a: slope of the line (by how much quality goes up if we add one percent of alcohol, ceteris paribus)

This is way too simple for our distribution, and the line can result in alcohol percentages corresponding to negative qualities, which is not possible. Therefore we do not want a model with a straight line, but we can look at, for example, a logistic regression (see Figure 4). For this, we simplify our data to a binary variable: 0 for low quality, 1 for high quality. A logistic regression expresses the relation between alcohol and quality in probabilities. For example, if a wine has an alcohol level of 12, it has an 86% probability of being good.

To make our full model, we need to include all predictors and not only alcohol. So far, we have used linear models that model correlation. However, we might be interested in non-linear relations too. For this, we need nonlinear models. Model selection is therefore very important and



Figure 1: The distribution of quality. Blue dots are high-quality wines, pink dots are low-quality wines [4].

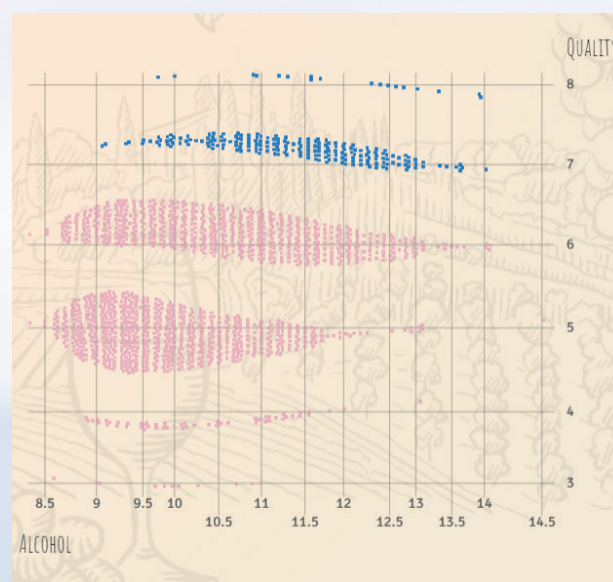


Figure 2: The relation between alcohol and quality. The higher the alcohol percentage, the better the wine [4].

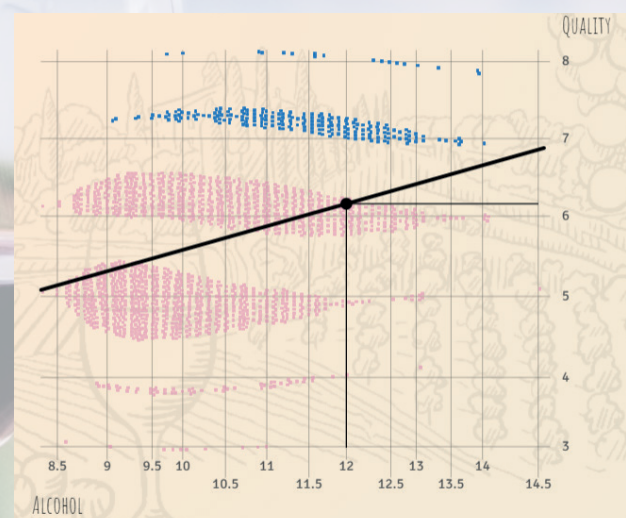


Figure 3: Ordinary Least Squares [4]

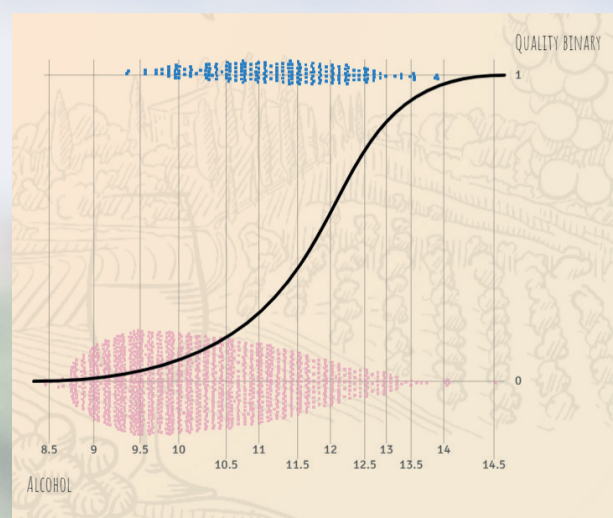


Figure 4: Logistic regression [4].

depends on what you are interested in. We could apply Random Forests to see which variables are more important in predicting quality. Principal Component Analysis can be implemented to see which variables explain most of the variation in the data set. But other machine learning techniques can be applied as well: think of Support Vector Machines or Neural Networks.

The eighteen-year wine plan

With the regression model, just discussed, we could theoretically predict what a wine we have not yet opened will be like and which chemical compositions lead to good wines. Both are interesting, but wine has to be drunk at the right time to really live up to

its potential. How to know when a wine is at its best is a well-researched subject (such as Gonen, Tavor, and Spiegel (2021) [9], I (Tamara) personally make use of vintage charts [6]). If you are in the mood for a rosé wine, then you are in luck. Rosés should not be drunk 'old' and are therefore often not that expensive. However, if you want a high-quality Bordeaux like from Chateau Pavie- St. Emilion (2001) at its peak, you will find some difficulty. You have to cough up around 300 euros a bottle [5]... If you do not have that money flying around, do not worry. With a bit of coordination and math, you can feel extravagant for a small price tag. Just last month, I bought a bottle of Bordeaux for eleven euros, also from

Chateau Pavie - St. Emilion with, according to wine experts, the same potential (both a 95 rating [6]). The only difference is the year. My bottle is from 2014 which means that it will take another ten to fifteen years before it attains the same perfect state of maturity.

By now, you have all guessed the solution: buy a 'young' wine, drink an 'old' one. So what does a perfectly balanced stash look like? If we want to keep drinking and having bottles at their peak of all ages, we get something looking like an ideal complex situation. We drink the old well-aged bottles (if newly bought, they are quite expensive) and buy just as many

young, cheaper bottles. However, unlike the most simple ideal complex situation, we want to drink different wines with different ageable characteristics. To deal with this issue you have to proportionally buy and drink wines of different rates of maturity [7]. You could, for instance, decide on a continuous six-tier system of three years in which you drink and consume 20 bottles a year (see Figure 5). You then need to buy nine bottles to drink in the first three years, five to be finished in four to six years, four bottles to be consumed in seven to nine years, one bottle to be opened in nine to twelve years, and one to save up till twelve to fifteen years. This way you constantly buy at a cheap price and throughout time, you will have a regular assortment of aged wine on hand.

Cheap Personalized Taste

Hopefully, you have been inspired. There is, however, a bit of a catch. It is doubtful that everyone has space for 133 wine bottles, agrees with the subjective term 'quality', and has the patience to wait eighteen years. Therefore, we would like to propose a third option: starting your own database (we all know that you like spreadsheets). When doing it right, you could create your own data set with details (of your choice) of the wines you have tasted within your price and patience range. Add your own opinion under 'quality' and you can regress your own taste. OLS might be too simple, but try another machine learning technique instead. This way you can find out exactly which wine characteristics you prefer (and which not). Then take into account that 90

percent of all wines are already at their best within the first year [8]. You will find that there are a lot of different 'flavors' to explore without having to keep a huge wine cellar. The wines of this category are often a bit more straightforward and cheaper white and red wines for which, it is useless to preserve them in hope of some great tertiary aroma. They are not all good nor all bad. You could describe them as a bit harsher, so you just have to find the ones you like. Do not forget: for the results to be statistically significant you have to taste quite a lot of wine, so let's go! And keep in mind: they might not all be superb but at least it stimulates your brain. ●

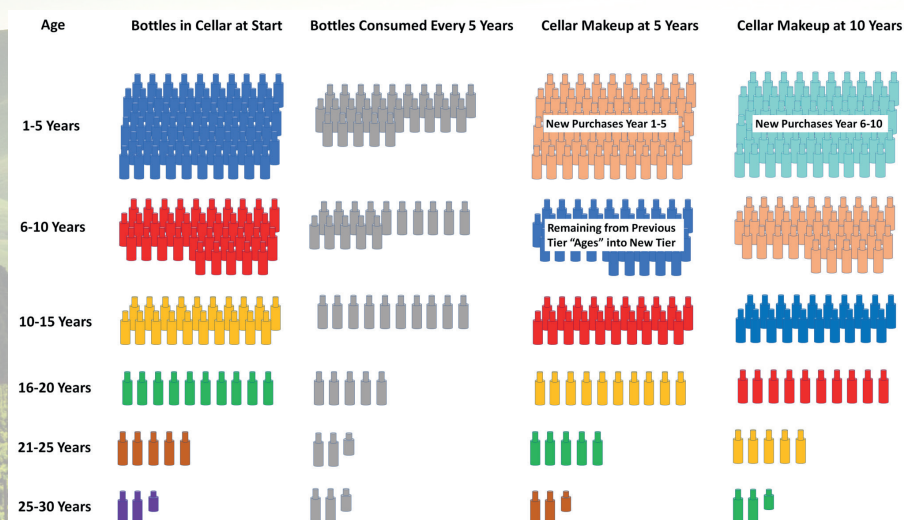


Figure 5: Minimum inventory level model, 3-year continues on in perpetuity (space needed for at least 133 bottles) [7].

Note 1:

Here at the start, a number of available bottles is assumed, this is not necessary for the process, but it does speed up the process of having wine ready at all maturities.

Note 2:

The bottleneck factor for this to work is that New Purchases proportionally have the right ageable qualities. Respectively 43% 1-5Y, 25% 6-10Y, 17% 10-15Y, 8% 16-20Y, 4% 21-25Y and 4% 25-30Y.

References

- [1] Van Dale en Jellinek: ook student kan alcoholist zijn | Punt. (avans.nl) (2018). Van Dale en Jellinek: ook student kan alcoholist zijn. Hoger Onderwijs Persbureau, May 11, 2021.
- [2] Wine tasting can work the brain more than math, according to neuroscience (nbcnews.com) (2018). Wine tasting can work the brain more than math, according to neuroscience. NBC news, May 11, 2021.
- [3] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis (2009). Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4), 547-553.
- [4] <https://pudding.cool/2021/03/wine-model/> (2021). Wine&Math a model pairing. The pudding, May 11, 2021.
- [5] 2001 Chateau Pavie, Saint-Emilion Grand Cru | prices, stores, tasting notes and market data (wine-searcher.com) (2021). 2001 Chateau Pavie, Saint-Emilion Grand Cru, France. Wine searcher, May 12, 2021.
- [6] Wine_Enthusiast_Vintage_Chart_2021.pdf (netdna-ssl.com) (2021). 2021 Vintage Chart. Wine enthusiast, May 16, 2021.
- [7] The Perfect Amount of Wines to Have in Your Cellar (Using Math) (thegrapepursuit.com) (2021). The Perfect Amount of Wines to Have in Your Cellar (Using Math). The grape pursuit. May 16, 2021.
- [8] Welke wijnen zijn geschikt om te ouderen? - Tervino Wijnblog(2019). Welke wijnen zijn geschikt om te ouderen? Tervino, May 17, 2021.
- [9] Gonen L.D., Tavor, T., Spiegel, U. (2021). The Positive Effect of Aging in the Case of Wine. Mathematics, 9(9), 1012.

Nekst Ten Years

In my previous three columns, I have described my past 30 years. The first ten years in Tilburg ended with my PhD defense around 20 years ago, followed by a decade at CQM of which more than half I combined with a role at our university. Finally, the last ten years described my working life: a combination of ASML and Tilburg University. Looking ten years ahead and making high-quality predictions is always difficult. Therefore, let me not give single-shot predictions but describe some possible scenarios. These I can then use to set some goals for myself!

From a work perspective, I expect that I will keep on doing the ASML/Tilburg University combination. Momentarily, we are finishing the paperwork for at least a five-year extension of the endowed chair I am appointed at. My role at ASML will most probably change quite a few times in the coming decade. Several reasons underpin my expectation to stay at ASML in the future. The most important one is that I expect more than enough possibilities to develop myself at ASML. ASML is currently the largest tech company in Europe (from a valuation perspective) and it is still growing! Last week (mid-April 2021) we announced our Q1 results and announced that we expect a 30% growth for 2021. There are three driving forces behind that growth: (1) short term chip shortages that you can read all about in the news these days driven by the speed-up of digitization as a result of COVID-19, (2) long term trends like 5G and AI, and (3) geopolitical trends where we see that on a global scale the US, Europe, and Asia want to increase their semiconductor capacities and capabilities such that they are more or less independent. The last two trends will not be over after 2022, so the future looks very bright for ASML.

I am currently the business controller for one of the two factories of ASML in the Netherlands. This means that I am responsible for finance in the management team of the so-called Twinscan Factory. In this factory, we make the DUV (deep ultraviolet) systems. In another factory, the newest systems, called EUV (extreme ultraviolet), are built. I have been in this role now for

two years and the reason that I switched from Decision Support to Business Control was personal development. This is the most operational role I have ever had in my career and it has been a good stretch assignment for me to develop on several aspects like convincing without power and operational reporting.

The most probable next step will be either a different business control role or a role in the newly formed Finance suborganization Business Insights & Control (Bi&C) that focuses more and more on data and analytics. Other business control roles are similar to my current occupation but focus on a different part of ASML, like sales or one of the business lines. The creation of the Bi&C organization is a major change and we need to make sure that we take full advantage of all possible data and decision science techniques by efficiently embedding these in our processes and tools.

At the university, I want to continue with most of the things that have been set in motion in the last five years (since my appointment on the endowed chair). Next to that, I want to focus more on the interaction between decision science and data science, also linking that to my work at ASML. Concretely, that means making the link between research and practice even more visible.

From a personal side, I hope to enjoy a lot more moments in the mountains with my family. Hiking, skiing, biking, but most of all relaxing and enjoying each other's presence! From a sporting perspective, I still want to make sure that our head of department takes me seriously. So, I want to climb Mont Ventoux on a bike within two hours, which means I need to improve my total time by 24 minutes. My goal for running seems 'easier' to achieve. For running a half marathon within 95 minutes, I only need to improve 200 seconds, so less than ten seconds per kilometer!

Now I have come to the end of my last column. I really enjoyed writing these columns and I hope it gives all of you some inspiration and food for thought! ●

Kuno Huisman

is director of business control Twinscan Factory at ASML. Next to that, he is a part-time full professor of decision making under uncertainty at Tilburg University where he teaches courses on professional skills and investment under uncertainty.



Recognizing DNA Patterns by Solving the Quadratic Traveling Salesman Problem

Although the recent popularity of data science is mainly due to trending advances in machine learning and big data, the world's most complex data set has been carried along by humanity for ages: its genes. It is this data set that explains how living organisms inherit features from their ancestors and plays a key role in thousands of complex biological processes. The ability to recognize patterns in an organism's genetic code contributes to the understanding of these processes. In this article, we consider a statistical model for this task and show how the 'best' model can be learned by solving a combinatorial optimization problem: the quadratic traveling salesman problem.

Transcription factor binding sites

An organism's genetic code can be seen as an ordered sequence of simple units called nucleotides that are stored in a large DNA molecule, see Figure 1. DNA patterns consist of four distinct types of these units according to their inherited base: guanine (G), adenine (A), cytosine (C), and thymine (T). The order of these nucleotides carries genetic information, similar to how the order of letters on a page carries information. The function of genes (i.e., pieces of DNA) is to provide information needed to make proteins in cells that are responsible for all kind of tasks, like building a new copy of the cell or repairing damage. In order to better understand these processes, it is desirable to have a good understanding of which DNA patterns are responsible for a certain cellular process. Pattern recognition in DNA is one of the main problems in bioinformatics, the interdisciplinary field that combines biology, computer science, and statistics. One important cellular process in gene regulation is transcription. Transcription is the process of copying a segment of DNA to a smaller similar molecule, called RNA. This process is initiated by a so-called transcription factor, which is a protein that binds to a specific DNA sequence. The place at which this transcription factor binds to the DNA molecule is called a binding site. Given a specific transcription factor, a key problem in bioinformatics is to recognize the binding sites of this factor. These binding sites are typically small ordered sequences of nucleotides G, A, C, and T. Whereas a transcription factor exhibits sequence-specific DNA binding properties, there is almost always variability in the sequence that they bind to. In other words, they bind to more than one sequence although a consensus sequence can often be established. For example, for a specific transcription factor called HNF4 α , Table 1 shows eight different types of DNA binding sites, although there exist many more such sequences.

Recognizing whether a given transcription factor like HNF4 α binds to a specific DNA sequence can be done using experi-

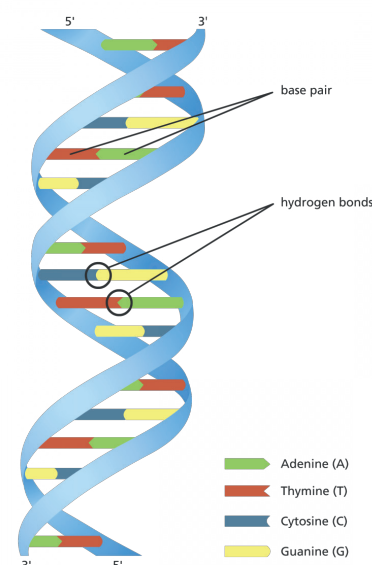


Figure 1: Simplified overview of a DNA molecule consisting of two strands (polynucleotides) in a double helix form, each consisting of four types of nucleotides (A, T, C and G).

mental techniques. This is not only costly, but also very time-consuming due to the extremely many possible DNA binding sites, even for short sequences. Since there is evidence that there are lots of dependencies among the positions in the binding sites, an alternative is to make use of statistical models, see [1, 4]. The idea behind these models is to train them on a set of known transcription factor binding sites, such that they can predict for new DNA sequences whether it will be a binding site for the transcription factor or not.

Permuted Markov Models

Suppose we have a dataset D of DNA sequences x of length L over a discrete finite alphabet Σ . In the case of DNA

1	A	G	T	T	C	A	A	G	G	A	T	C	A
2	A	G	T	C	C	A	G	A	G	G	T	C	A
3	A	G	G	G	T	A	A	A	G	G	T	T	G
4	A	G	T	T	C	A	A	A	G	T	T	C	A
5	T	G	G	G	C	A	A	A	G	G	T	C	A
6	G	G	C	A	A	G	G	T	T	C	A	T	A
7	G	G	T	C	C	A	A	A	G	G	G	C	G
8	G	G	G	T	T	A	A	A	G	G	T	T	G

Table 1: Eight known binding sites for transcription factor HNF4 α . Each binding site consists of 13 nucleotides.

sequences, this alphabet is the set of all nucleotides, i.e., $\Sigma = \{G, A, C, T\}$. The set D consists of *known* binding sites of a certain transcription factor obtained by experimental techniques. The main objective is to learn a model underlying the sequences in D that can be used to recognize new DNA sequences having similar patterns.

A powerful technique for this task is the use of (hidden) Markov models [1]. In such a model, a binding sequence is presented as a Markov chain that gives the probability of each nucleotide occurring at a particular position depending on the nucleotides at preceding positions. Let X_i for $i = 1, \dots, L$ denote the discrete random variable associated with position i in the sequence. Each random variable X_i can take values from Σ . This situation can be viewed as a Markov chain having 4^L states, where the outcome of X_i depends on the outcome of the preceding random variables X_1, \dots, X_{i-1} . The order of the Markov model determines the number of preceding positions that have a direct effect on X_i . In a Markov model of order one, we have $P(X_i = x_i | X_{i-1} = x_{i-1}, \dots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1})$ for all $i = 2, \dots, L$. In a Markov model of order two, we have $P(X_i = x_i | X_{i-1} = x_{i-1}, \dots, X_1 = x_1) = P(X_i = x_i | X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2})$ for all $i = 3, \dots, L$.

The underlying assumption of the Markov model described above is that the dependencies among the variables are local, i.e., the nucleotide at position i depends on the nucleotides on the positions adjacent to i . This does however not fit with the biological observation of transcription factor binding sites. The binding between a DNA molecule and a transcription factor is essentially a 3-dimensional geometrical matching process that may involve cooperation between nucleotides at non-adjacent positions of the primary DNA. For example, for the transcription factor HNF4 α simple statistical analysis reveals a significant dependence between position 4 and 8.

To overcome this, we define a similar Markov model on a permuted sequence of the initial positions. Such a model is called a Permuted Markov Model [4]. Let $\pi : \{1, \dots, L\} \rightarrow \{1, \dots, L\}$

be a permutation on the set of positions. Then $X_{\pi(i)}$ is the random variable that is associated with the nucleotide at position $\pi(i)$ in the original sequence. In a permuted Markov model of order k , we assume that $X_{\pi(i)}$ depends on $X_{\pi(i-1)}, \dots, X_{\pi(i-k)}$. In this way, we capture dependencies among adjacent positions in the permuted sequence, which correspond to possibly non-adjacent positions in the original sequence. Indeed, the identity permutation $\pi(i) = i$ provides the original Markov model. Experience has shown that permuted Markov models of order one are too restrictive to model transcription factor binding sites [4], so we focus on permuted Markov models of order two from now on.

Selecting the best model

Given a permutation π , we can estimate the transition probabilities using the frequency information of the sequences in D , as will be shown below. Hence, each π gives rise to a model and our task is to find the permutation π that provides the best model. As a measure for the fit of a model we use its log-likelihood, which is the logarithm of the total likelihood of the model under permutation π . Informally, this likelihood can be interpreted as the total probability of observing the sequences in D if the second order permuted Markov model based on permutation π would be the true model for generating the transition factor binding site patterns. Since taking the logarithm does not change the monotonicity of the probabilities, the ‘best’ model is the one that has the largest log-likelihood. Let $x \in D$ be a given DNA sequence, where x_i denotes the nucleotide at position i in x . To simplify notation, let the event $X_{\pi(i)} = x_{\pi(i)}$ be written in shorthand as $x_{\pi(i)}$. For example, the probability $P(X_{\pi(i)} = x_{\pi(i)})$ is abbreviated as $P(x_{\pi(i)})$. Now, the log-likelihood of the sequence x given π is

$$\log P(x|\pi) = \log P(x_{\pi(1)}) + \log P(x_{\pi(2)}|x_{\pi(1)}) + \sum_{l=3}^L \log P(x_{\pi(l)}|x_{\pi(l-1)}, x_{\pi(l-2)}). \quad (1)$$

Moreover, the log-likelihood of an independent and identically distributed dataset D of sequences given permutation π can be written as:

$$\log P(D|\pi) = \sum_{x \in D} \log P(x|\pi). \quad (2)$$

As already mentioned, the (conditional) probabilities in (1) can be estimated using the frequencies in D . Let $N = |D|$ and define for all positions $i = 1, \dots, L$ and nucleotides $y \in \Sigma$ the absolute frequency of observing nucleotide y at position i , i.e.,

$$N_i(y) := |\{x \in D : x_i = y\}|.$$

In a similar fashion, we can define the absolute frequencies of observing two nucleotides y_1, y_2 at position i and j , respectively, or three nucleotides y_1, y_2, y_3 at positions i, j and k , respectively. That is,

$$N_{i,j}(y_1, y_2) := |\{x \in D : x_i = y_1 \wedge x_j = y_2\}|,$$

$$N_{i,j,k}(y_1, y_2, y_3) := |\{x \in D : x_i = y_1 \wedge x_j = y_2 \wedge x_k = y_3\}|.$$

Using these frequencies, the probabilities $P(X_i = y_1)$, $P(X_i = y_1 | X_j = y_2)$, and $P(X_i = y_1 | X_j = y_2, X_k = y_3)$ for all positions $i, j, k \in \{1, \dots, L\}$ and all nucleotides $y_1, y_2, y_3 \in \Sigma$ can be estimated as follows:

$$P(X_i = y_1) = \frac{N_i(y_1)}{N}$$

$$P(X_i = y_1 | X_j = y_2) = \frac{N_{i,j}(y_1, y_2)}{N_j(y_2)}$$

$$P(X_i = y_1 | X_j = y_2, X_k = y_3) = \frac{N_{i,j,k}(y_1, y_2, y_3)}{N_{j,k}(y_2, y_3)}.$$

By plugging in these estimated probabilities into (1) and taking the sum over all sequences in the dataset D , see (2), we obtain a fixed value for $\log P(D|\pi)$. Our goal is to maximize $\log P(D|\pi)$ over the set of all permutations π . As the number of permutations grows exponentially large, it is impractical to compute the log-likelihood of all permutations and select the best one. Thus, we are in need for more clever optimization methods. In the next section, we will see that the problem of finding the optimal permutation can be reduced to an instance of the quadratic traveling salesman problem, as suggested by Fischer et al. [2].

The quadratic traveling salesman problem

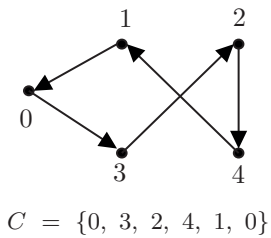
The task of selecting the permutation having the maximum log-likelihood in a permuted Markov model is a combinatorial optimization problem. Indeed, each permutation π leads to a fixed objective value and the number of possible permutations is finite. The order of the underlying Markov model determines the cost structure of these permutations, and hence the flavor of the involved combinatorial optimization problem. For permuted Markov models of order one, the log-likelihood is the sum of L terms where each term depends on (at most) two adjacent positions in the permutation. This problem can be reduced to the well-known traveling salesman problem. The log-likelihood of a permuted Markov model of order two, however, is a sum of terms where each term depends on (at most) three positions traversed in succession in the permutation. This higher-order combinatorial problem can be modelled as a quadratic traveling salesman problem, as introduced by Jäger and Molitor [3].

The quadratic traveling salesman problem (QTSP) is defined

on a complete directed graph $G = (V, A)$ with vertex set V and arc set $A := \{(i, j) : i, j \in V, i \neq j\}$. Given is a cost function $q : V \times V \times V \rightarrow \mathbb{R}$ that assigns an interaction cost to any three vertices traversed in succession. The cost $q(i, j, k)$ can be viewed as the quadratic cost between adjacent arcs (i, j) and (j, k) , where $q(i, j, k) = 0$ if $|\{i, j, k\}| < 3$. A directed Hamiltonian cycle in G is a sequence of vertices starting and ending at the same vertex such that every vertex is on the cycle exactly once. The cost of a directed Hamiltonian cycle $C = \{v_1, v_2, \dots, v_n, v_1\}$ with respect to q is given as

$$q(C) = \sum_{i=1}^{n-2} q(v_i, v_{i+1}, v_{i+2}) + q(v_{n-1}, v_n, v_1) + q(v_n, v_1, v_2).$$

The goal of the QTSP is to find a directed Hamiltonian cycle C in G that minimizes $q(C)$. The QTSP is known to be \mathcal{NP} -hard [3], which means that it is generally very hard to solve. We now show how to model the selection of the optimal permuted Markov model of order two as an instance of the QTSP. Let $V = \bar{V} \cup \{0\}$ be the set of vertices where $\bar{V} := \{1, \dots, L\}$ corresponds to the set of positions in the DNA sequence and 0 is an artificial vertex. A directed Hamiltonian cycle can now be interpreted as a permutation of the positions in $\{1, \dots, L\}$ in the sense that the vertex that is visited directly after the artificial vertex in the cycle is considered as $\pi(1)$, the vertex that is visited next in the cycle is considered as $\pi(2)$, etc. For example, suppose $L = 4$ and consider the following directed Hamiltonian cycle C in G :



This directed Hamiltonian cycle corresponds to the permutation π given by:

k	1	2	3	4
$\pi(k)$	3	2	4	1

One can easily verify that any permutation $\pi : \{1, \dots, L\} \rightarrow \{1, \dots, L\}$ belongs to exactly one directed Hamiltonian cycle in G and vice versa. For each permutation π , let C_π denote the corresponding directed Hamiltonian cycle in G . It rests to specify the construction of the quadratic cost function q , which is the key in the derivation. This construction should be such that $q(C_\pi) = -\log(D|\pi)$ for all permutations π .

Observe that we include a minus-sign, since we aim to find the permutation having the maximum value of $\log(D|x)$, which is equivalent to the minimum value of $-\log(D|\pi)$. Combining (1) and (2), the total cost of a directed Hamiltonian cycle C_π should become:

$$q(C_\pi) = - \sum_{x \in D} \log P(x_{\pi(1)}) - \sum_{x \in D} \log P(x_{\pi(2)}|x_{\pi(1)}) \\ - \sum_{l=3}^L \sum_{x \in D} \log P(x_{\pi(l)}|x_{\pi(l-1)}, x_{\pi(l-2)}).$$

In order to establish this, we distinguish three cases for q . The term $-\sum_{x \in D} \log P(x_{\pi(1)})$ in the expression above is modeled by the quadratic costs $q(i, 0, k)$, for all $i \in V$ and $k = \pi(1)$. Similarly, the second term in the expression above is modeled by $q(0, j, k)$ where $j = \pi(1)$ and $k = \pi(2)$. Finally, the remaining costs are modeled by $q(i, j, k)$ where $i = \pi(l-2)$, $j = \pi(l-1)$ and $k = \pi(l)$. Hence, we introduce the following quadratic cost function q :

$$q(i, j, k) := \begin{cases} -\sum_{x \in D} \log P(x_k) & \text{if } j = 0 \\ -\sum_{x \in D} \log P(x_k|x_j) & \text{if } i = 0 \\ -\sum_{x \in D} \log P(x_k|x_j, x_i) & \text{if } i, j, k \in \bar{V} \\ 0 & \text{otherwise.} \end{cases}$$

By construction, the directed Hamiltonian cycle C_π that minimizes this instance of the QTSP provides the permutation π leading to the maximum log-likelihood.

Solving the QTSP

Various algorithms for solving the QTSP have been proposed. We finalize this article by considering a binary formulation for the problem, which forms the backbone of many of these solution techniques. Let us define a binary variable x_{ij} for all $(i, j) \in A$ such that

$$x_{ij} = \begin{cases} 1 & \text{if } (i, j) \text{ is contained in the Hamiltonian cycle} \\ 0 & \text{otherwise.} \end{cases}$$

Now we can formulate the QTSP as follows:

$$\min \sum_{i \in V} \sum_{j \in V} \sum_{k \in V} q_{ijk} x_{ij} x_{jk} \quad (3)$$

$$\text{s.t.} \quad \sum_{j \in V} x_{ij} = \sum_{j \in V} x_{ji} = 1 \quad \forall i \in V \quad (4)$$

$$\sum_{i, j \in S} x_{ij} \leq |S| - 1 \quad \forall S \subseteq V, S \neq V, \emptyset \quad (5)$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in A. \quad (6)$$

The constraints (4) manage that the in- and out-degree of each vertex is equal to one. In order to prevent the possibility

of subtours, i.e., the arcs (i, j) with $x_{ij} = 1$ form a cycle of size smaller than n , we add the constraints (5). These constraints manage that we can select at most $|S| - 1$ arcs for any proper subset S of V . The constraints (4)-(6) together imply that x is indeed the characteristic vector of a directed Hamiltonian cycle. The objective function (3) incurs a cost q_{ijk} if and only if arcs (i, j) and (j, k) are both present in the Hamiltonian cycle.

The QTSP solution techniques in the literature mostly rely on branch-and-bound and branch-and-cut algorithms. Fischer et al. [2] show that this latter approach is currently the best performing algorithm for the QTSP, being able to solve instances for binding site sequences up to $L = 200$. ●

References

- [1] K. Ellrott, C. Yang, F.M. Sladek, T. Jiang. Identifying transcription factor binding sites through Markov chain optimization. *Bioinformatics*, 18:100–109, 2002.
- [2] A. Fischer, F. Fischer, G. Jäger, J. Keilwagen, P. Molitor, I. Grosse. Computational recognition of RNA splice sites by exact algorithms for the quadratic traveling salesman problem. *Computation*, 3:285–298, 2015.
- [3] G. Jäger, P. Molitor. Algorithms and experimental study for the traveling salesman problem of second order. *Proceedings of the 2nd international conference on Combinatorial Optimization and Applications*, 211–224, 2008.
- [4] X. Zhao, H. Huang, T.P. Speed. Finding short DNA motifs using permuted Markov models. *Journal of Computational Biology*, 12:894–906, 2005.



Frank de Meijer
PhD Candidate

Another Memorable Weekday

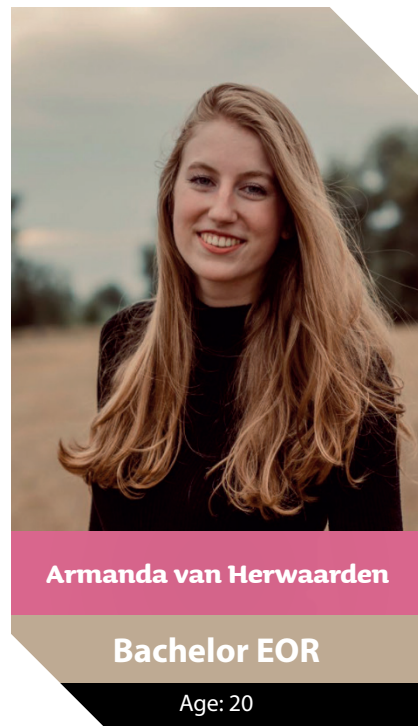
On Tuesday, April 20, the Active Members Weekend Committee of Asset | Econometrics organized the online event 'Ik hou van AMW'. Normally, this committee organizes a weekend away at a remote location which is top secret. It is the perfect opportunity to get to know your fellow econometricians a little bit better in a relaxed setting. This weekend usually takes place in April, but it had to be canceled due to the measures related to the coronavirus. Even though it was too bad that it could not be carried through, the AMW committee came up with a new nice idea; 'Ik hou van AMW'. Although this was not the same as a festive weekend away with all active Astrics members, this event was a good alternative! The concept is based on the Dutch television program 'Ik hou van Holland' which is hosted by Linda de Mol. The purpose of this program is to win the battle against the other team by winning several games. The show tests the participants on their general knowledge. I immediately subscribed after seeing the announcement for this event in the Whatsapp group, so I could be part of this happening!

'Ik hou van AMW' started at a quarter past seven in the evening. It was possible to participate in the event as an individual or in pairs. If you had chosen the option to register individually, the committee linked you to someone else who did not have a partner either. Then the AMW committee matched you with another duo which completed your team for the evening. The purpose was to compete in quartets against each other in different game rounds to take a chance on winning a big prize. Besides that, they provided us participants with a well-stocked goodie bag containing some specialty beers and delicious snacks the day before the event. You

had to control yourself not to eat anything beforehand! Moreover, this bag included a customized deck of cards with the AMW logo. The back of the cards was pink which made the cards very cute. Also, I liked the self-made label of the Corona beer.

In total, there were eight game rounds and ten teams. I registered as a duo with Marie Versteeg and we were linked to the duo Bob Suijkerbuijk and Pierre Verhulst. We were dedicated to win and claim the big prize! The game of our first round was solving the '4 words, 3 rubrics' kind of puzzles. In this game, you get twelve words in total. Four words are linked to one specific category and you have to try to name all three categories in order to win this game. In case you do not guess all the rubrics in time, the opposing team has the opportunity to guess. Next, we played 'Kaaswerpen'. This game is designed to test your topography. The game master mentions a city and the participants have to pin it on the map. Everyone who pins it right within a certain region gets a point. This was a fun game, but at the same time, a bit difficult. You often know the approximate location of a city, but not the exact coordinates. Furthermore, we played 'guess the lists' which is an exciting game, especially under time pressure. The game master brings up a category and you have to say how many things of this specific category your team can name within the given time. The other team has the possibility to indicate whether they can mention more things. You lose the game if you do not reach the number you specified at the beginning of the game.

Then, we played a music round. Two songs have been mixed up and it is your task to guess both the artist and the name of the two songs. We did not only play this game, but also the photo round. The game master showed a picture that is extremely



zoomed in. And so, gradually, the picture is zoomed out. You win this game if you are the first one who recognizes what is displayed on this photograph. Lastly, we played a game with estimation questions. The game master asked a question and you had to give your answer as a team. The opposing team could state whether they thought the answer was higher or lower than the answer provided by the other team. The team whose answer was closest to the true value won this game. An important thing that I have learned during this game is how many steps the big white staircase in the Cube has. Personally, I expected this number to be higher...

All in all, it was a successful event which delivered my team a joint third place! ●

Building Blocks Escape Room

On Tuesday March 9, the online escape room with Building Blocks took place. Due to the COVID-19 measures, we were not able to visit their office in person, unfortunately. However, Building Blocks had organized a great event with a fun escape room and a lovely box with beers and snacks for us to enjoy from our own homes.

"This task was for us to liberate a scientist who had been kidnapped."

The afternoon started with a tour of their office, which is located right next to the train station of Tilburg (Plant). On the fifth floor, we were given a glimpse of their work spots and the cafeteria before reaching the office where the company presentation would be given. The presentation was very insightful and it was a great way of showing how Building Blocks operates, what type of clients they have, and what sort of projects they are doing. But what stood out to me the most was the structure of Building Blocks. They are a relatively small company of around 40 to 50 people and use a flat organizational structure, meaning that every employee is equal and there is a lot of room for discussion among colleagues.

On to the most important (or most fun) part of the activity: the escape room! Beforehand, a package had been delivered to each of us with the necessary documents to find the solution of the escape room. We were

divided into groups of four students with one 'supervisor' from Building Blocks and were given one hour to solve the task at hand. This task was for us to liberate a scientist who had been kidnapped. We had to do this by finding the password which consisted of four two-digit numbers. The numbers could be found by discovering the clues that linked the documents. After each group was put into their respective breakout room, my group and I frantically started by looking for hints. A tactic that worked surprisingly well as we found the first two numbers with relative ease and in a short amount of time. However, the remaining two proved to be more difficult. After a couple minutes of

"They are a relatively small company of around 40 to 50 people and use a flat organizational structure, meaning that every employee is equal and there is a lot of room for discussion among colleagues."

intense deliberation (and a penalty minute for asking our supervisor for a hint), we were able to find the final numbers. We finished the whole riddle in 40 minutes!

Now, all we could do is wait. We spent the last twenty minutes in the breakout room discussing how well we did, as well as asking a couple of questions about working at Building Blocks before heading back to the main room,



where the results would be announced. Unfortunately, we did not break out of the escape room the fastest as the winning team had already solved it in half an hour. Despite not winning, we were quite proud of ourselves as we came in second. After the results, the networking drink started and everybody had the opportunity of talking with different Building Blocks colleagues. This was a relaxed yet informative way of getting to know the company better in addition to hearing more about the personal experiences of the people working there. Overall, the Building Blocks session was a lot of fun and an enjoyable way of learning about the company. I would like to thank Building Blocks for organizing this event! ●

Hein Fleuren: Gardener, Entrepreneur, and an Honorary Member

written by **Jarno Ringhs** and **Mylan Tran**



When we asked professor Fleuren for an interview, he replied enthusiastically. However, he was also very curious for a particular reason. He wondered why this was the first time he received an invitation, since he had already been working at the university for 20 years. So, there we were, totally speechless, after our first question. Luckily, Hein was very excited to do the interview. About *Nekst* he said: "I definitely read it! It is a luxurious magazine. I especially like to read about where students end up after graduating and about my colleagues. I, for one, did not know that Cristian Dobre was a referee! I also enjoy reading about what students think about their studies." This was the start of an exciting interview and afterwards we were also wondering why we had not interviewed him earlier: Hein Fleuren is truly one of the most interesting professors we have.

A very busy student life

Hein was born in Oss, Noord-Brabant. After finishing high school, he decided to go to the University of Twente. "I wanted to become an engineer since my older brother was an engineer. TU Eindhoven was a more logical choice since it was much closer, but I really liked Twente because of its unique campus!", Hein explained. In Enschede, he studied Mechanical Engineering (focusing on Operations Research), but later also added the Master Mathematics to his curriculum. Within Mathematics, he majored in Information Technology. Afterwards, he obtained his doctorate on Vehicle Routing and Scheduling. Hein said: "I love the combination between innovation and prac-

tice. I want to see if methods actually work in practice under different circumstances."

As if doing two Masters and a PhD track was not challenging enough, Hein also experienced a highly fulfilling student life. He did a lot of sports such as running and table tennis. Next to that, he had a very active 'flat'. Together with his roommates, he went mountain climbing and cycling. They also participated many times in the well-known Batavierenrace! Hein added: "We even did the 'Pheidippides loop' in Utrecht. This is a relay race in which you run a whole marathon with your teammates." As you can see, Hein was quite the sports guy, but he also liked to join vinyl records sales. To conclude the story of his student time, Hein also told us that he used to be a bartender. Moreover, if you were to build a time machine and go back in time, you would probably find him in the Vestingbar of the university, where he spent many hours with his friends.

Professional career

Hein started his business life at CQM in Eindhoven. After ten years, he came to Tilburg. First, he worked for CentER Applied Research in combination with a one-day professorship. Then, he decided to become an entrepreneur next to being a professor. "In these years, I did many projects for TNT Express which I really liked. Nonetheless, in my opinion, 'optimization' was not doing so well during this time. It was too focused on cost reduction and kind of old-fashioned", Hein explained. It was then that he decided to start his own company: BlueRock Logistics. With its main office in Den Bosch, BlueRock is nowadays still growing. "BlueRock is what I have worked on for the past eight years, but a few years ago, I wanted more.

I wanted to see if the field of optimization could contribute more to the world. I do want to emphasize that I still enjoy working together with large corporations on beautiful optimization problems. However, I felt the need to use optimization for other causes too such as the Zero Hunger Lab."

World Food Programme and Zero Hunger Lab

We asked Hein how the World Food Programme and the Zero Hunger Lab are related. He explained: "I had already worked for the World Food Programme (WFP) before, but to make an actual impact, money is needed. However, all of the WFP's money goes to food security. So, professor Dick den Hertog, who is also connected with the WFP, and I started brainstorming new ideas. We thought it was peculiar that the WFP is able to assist 100 million people per year, but the actual hunger statistics state that there are 700 million people who need help. So, how can we help those other 600 million? That is when we came up with the idea of the Zero Hunger Lab. Together with Perry Heijne, co-founder of the 'lab', we went to the Ministry of Foreign Affairs to pitch our story. They found it fascinating that we wanted to use data science to fight world hunger and decided to give us a large subsidy of 2.5 million euros. When TISEM also jumped in, it meant that we had enough resources to start the lab. So, I think you could say that the Zero Hunger Lab arose from the World Food Programme. Now after 2.5 years, we are still working with several PhD students to improve our solutions. As a result, nowadays, I work two days a week with the Zero Hunger Lab, one day as department staff, and the other two days for BlueRock."

Franz Edelman Award

This year, professor Fleuren and professor Den Hertog amongst others won the Franz Edelman Award for their work with the WFP. Koen Peters, one of Hein's PhD students, was the lead of their team. The Franz Edelman is the most prestigious award in the world that attests to the contributions of operations research and analytics in both the profit and non-profit sectors. Dick and Hein are probably the only two people that have won this prize two times! Hein won it in 2012 for his work with TNT Express. Dick won it in 2013 for his work on the flood standards to protect the Netherlands. "It is an incredible experience to win such a big award again. Our work for WFP has a totally different dimension because we are helping undernourished people. Instead of increasing the profit of a corporation or reducing the emission, we are actually impacting human lives!", Hein exclaimed. He continued by expressing his pride for Tilburg University: "By winning this award, the university has won three times now. If we include the work of Goos Kant as well, we have reached the finals four times. This puts the university in a great spotlight."

His own course

Next to his work for the Zero Hunger Lab and BlueRock, Hein is of course also a teacher. He teaches the course Professional Business Analytics Skills, which is part of the Business Analytics and Operations Research Master. He has set this course up with Kuno Huisman and Goos Kant, because he felt that at the time something was missing in the curriculum of the Econometrics and Operations Research program. He said: "I have seen a lot of young people come in and I also have had the privilege to guide a lot of these people. However, Kuno, Goos, and I came to the conclusion that there was barely anything in the program that gave the students something to hold on to for the practical work later on in their career. So, we developed a course, which Goos had followed during the Bachelor, into a course for the Master program. We added a part about acquisition and extended it mainly with presentation skills. Currently, the fifth edition of the course is in progress and I enjoy it very much!"

The Impact Team

During the pandemic the book "The new common: How the Covid-19 pandemic is transforming society" got released. Hein was one of the four editors of the book. Hein said the following about the book: "I love to do as many things as possible in my life and next to the Zero Hunger Lab, I have also been involved with the Impact Team of Tilburg University. This team wants to do multidisciplinary research to eventually contribute to society. During the start of the pandemic, the idea came up to ask several scientists from different disciplines about their view on how this crisis will change the world. Together with Margriet Sitskoorn, Ton Wiltthagen, and Emile Aarts, we moved from words to deeds and we published this book. Recently, it has been published as open access and in the meantime it has been downloaded more than 20,000 times!"

His spare time

Despite his busy schedule and long working days, Hein finds enough time during the week to practice his hobbies. He is very passionate about them: "I am fascinated by gardening and I love to grow things such as bell peppers for example. In addition, I also like to grow more exotic things, like a Brazilian strawberry tree. Next to gardening, I spend my time as an enthusiastic amateur photographer. I like making pictures and selling them. From the money I make from these pictures, I try to buy the lenses for my camera. It is not about the money for me, but I truly enjoy it when people like one of my photos enough to hang it in their living room or to publish it in a magazine. Furthermore, I am a huge lover of music. The music I listen to the most is classical music, but I also enjoy more modern music." When asked about his favorite composer, he hesitated a little bit before answering the question. "I am not sure if I want to reply, because my wife would probably not like the answer. I like Tchaikovsky and Rachmaninov, because they are able to vary between huge chunks of noise and the most beautiful melodies. I find this super impressive and I also used to go to concerts very often, but my wife is not very fond of this music. So, when I am listening, I turn the volume up a bit", he said jokingly. ●



dr. Hein Fleuren

Bert & Ernie Questions

Beer or Wine?

"Beer"

Cinema or Theatre?

"Cinema"

Pancakes or Fries?

"Fries"

America or Asia?

"America"

Tilburg or Oss

"Tilburg"

Get Ready For Your First Job!

On this poster, you can find a lot of tips and tricks for your applications process. These tips were gathered by talking to old board members of Asset | Econometrics and were gathered specifically for econometrics students. If you are interested in how this poster was made, you can read the article we wrote on the next page. Besides this poster, make sure that you use other information sources as well to prepare for your job interview! All the best with preparing for the job market.

Busy with studying?

Start thinking about your career as well!

Find out what you like most in your studies and orientate yourself on the goals you want to reach. Later on, this will help you to find the most suitable job for you!

Do more than just studying! Besides the fun, this will also help developing your soft skills. There are a lot of different ways to do this, for example:

Become active at your favorite association. You can join committees or even do a board year.

Attend company events, Business courses, and workshops. This will also help to build up your network.

Find a student job that is actually interesting for an econometrics student, do not settle for a supermarket job.

Become visible for recruiters: create a LinkedIn profile and include everything you do on there.

Found a company you like?

Get ready for your application!

Be well prepared for your job interview:

Have an answer ready to frequently asked questions. You can find some examples on the bottom right of this poster.

Try to know a lot about the company you are applying to. Think about business values, structure, and actualities around the company/sector. Use Google for this (duh) or ask someone that works there.

Find out who you are talking to. Look them up on LinkedIn.

Prepare questions to ask yourself about the application process, the company and your conversation partner.

During your Econometrics study

Near the end of your study

Preparation for

Feeling ready to contribute to society?

You should think about what company and function fits you most. Try to find an answer to the following questions:

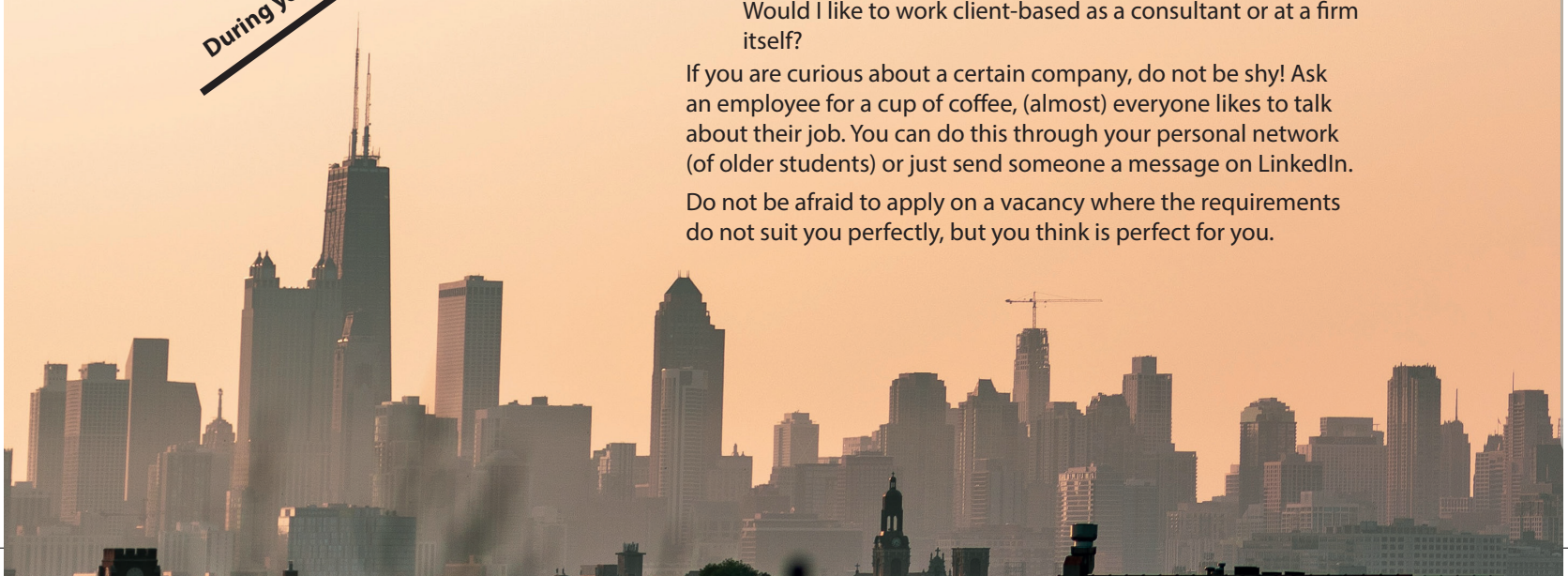
In which sector do I want to work?

Do I want to work at a big or small firm?

Would I like to work client-based as a consultant or at a firm itself?

If you are curious about a certain company, do not be shy! Ask an employee for a cup of coffee, (almost) everyone likes to talk about their job. You can do this through your personal network (of older students) or just send someone a message on LinkedIn.

Do not be afraid to apply on a vacancy where the requirements do not suit you perfectly, but you think is perfect for you.



You are hired! What now?!

The real learning will begin after you are hired. Have faith in yourself!

Keep on investing in your personal development.

Your feeling with a company is very important. Keep on looking if you are not happy with your current job!

After your application

During your application

your application

How to survive your application conversation?

Ask questions back! It should be a two-way conversation.

Try to answer questions with practical examples.

When you are an econometrician, there is no doubt about your knowledge. They want to know more about you as a person. Try to emphasize this.

Always be honest, it is no problem to not know the answer to some questions.

Be confident, they should be happy to have you.

Write your perfect Curriculum Vitae!

Everything that is on your CV should be on there for a reason. You should be able to explain this.

Do not give yourself a score for skills. You can use terms like beginner/intermediate.

Tone it down with the creative graphics, you want to come across professional.

Add references of, for example, your thesis. (Ask permission!)

Write your perfect motivation letter!

Make your letter approximately one A4 long.

Write your letter specifically for the company. Try to find what kind of person they are looking for and incorporate this.

Make your letter personal and stand out.

Mention your strenghts and try to give examples.

You will not be hired because of your letter but you can be rejected because of it. Make sure that you have the basics in order, like spelling and formatting.

Some specific points to prepare for your application!

What are your strengths and weaknesses? (Have examples ready!)

Tell me about yourself. (Make it personal and enthusiastic!)

What attracted you to our company?

Where do you see yourself in five years? (Please do not say 'at another firm')

What is your ambition within this company?

Could you tell me about your thesis?

The Making of: The Application Poster

It was during our first Advisory Council session of the academic year of 2020-2021, that one of the former board members came up with the idea which was the basis for this article and the poster on the previous pages. He suggested that it would be nice to do something for recently graduated econometrics students that could use some tips for their applications. We as a board really liked this idea, and assigned two people to work it out! We, Juliëtte van der Velden and Juul Schuurmans, were assigned to the job. We both found it a really interesting project, and decided that we would like to spend more time on it. In this article, we will explain the entire process of how we made the poster: from gathering all of the information, until the end product.

During this Advisory Council session, the idea was born to approach former board members and ask them for help. These former board members already have a lot of working experience now, have often applied multiple times at different firms, and sometimes even sat at the other side of the application process. We decided that we wanted to start with an open vision and gather as much information as we could. We contacted all the former board members from the academic years 1999-2000 till 2011-2012 and held an open conversation with most of them about their experiences.

During these conversations, our main target was to gather tips that are specifically relevant for econometrics students. Together with the former board members, we mainly talked about the following three subjects: Tips in general, tips for your motivation letter and CV, and tips for during the conversation rounds. About these three subjects, we asked them what were things that stood out positively or negatively, and if there were any things they would have liked to know when they were searching for a job themselves. After gathering all the information, and sorting the information per subject, we thought about how we wanted to translate all of this into something for our members.

While looking at all the tips, we found out that we could divide them all into different moments in time during the application process. Some general tips were more relevant for a student who was getting ready for the job market, while another tip was very relevant for the student who was currently preparing for their interview. We therefore came up with the idea of making a timeline poster and putting it in Nekst! At the beginning of the timeline, there is the student who is still orienting and does not know where he/she wants to work yet, and at the end of the timeline you can find tips for the last phases of the application process. On the previous pages, you can see the result of our design idea.

We would like to thank all the former board members who contributed to the



making of this poster! Without your great tips, this of course would not have been possible. In all honesty, we both really liked all the conversations we had and we were able to learn a lot about the application process. Besides this, we both also really liked to see a lot of former board members, learn where they ended up, and to get to know them a little better. For all the people that are applying for a job, we wish you all the best and we hope that you can use some of the tips! ●

"At the beginning of the timeline, there is the student who is still orienting and does not know where he/she wants to work yet, and at the end of the timeline you can find tips for the last phases of the application process."

The Vivid World of Pensions

This is already the last column of the academic year 2020-2021. Summer is coming. For some of you, this means you will soon receive your diploma and start your (first) job. For others, entering the job market still takes some time. Maybe you will start working for a bank, a pension fund, an insurer, or for one of the regulators or consultancy firms, respectively controlling or advising the previously mentioned companies, or you might do so in some years from now. You might know that some interesting developments are happening in the Dutch pension sector. These will kick off in 2023 with the transition process and are planned to be fully running from 2027 onwards. This reflects that the speed, impact, and challenges are actually enormous, potentially surprising those of you who think that pensions are dull.

The Dutch pension system is ranked as the world's best pension system (Mercer, 2020). To keep up with the challenges of increasing life expectancies and low interest rates, there is a need for more sustainable systems. In the current system, the majority builds up pension entitlements. The new system will, however, be based on individual accounts with collective risk sharing. This transition implies that the total pot that pension funds own today needs to be divided among the individuals. How to and other related questions are being scrutinized by researchers of Netspar among other parties, such as the government and politics including regulators and ministries, but also pension funds and insurance companies. As independent researchers, we translate societal problems into mathematical questions and do the reverse for its solutions.

Pension reform is not only high on the agenda in the Netherlands but is taking place in other countries too. Ranked number two among the world's best pension systems, is Denmark. By the way, only since 2018 were we able to 'win' from them. In Denmark, pension products included guaranteed minimal returns up to 2007. Since then, these have gradually been replaced by market-sensitive products that do not feature a guaranteed return anymore. At first sight, you might think that alleviating guarantees would decrease pension payments. However, without having the obligation to invest most of the pension wealth in bonds or swaps in order to ensure the promised return, the pension fund can now actually optimally invest in a diversified

portfolio. This would generate more returns though at a higher risk. On average, the expected pension payments would therefore increase. In addition to the increased financial risk, there is now also the presence of macro longevity risk, which is the risk that we will live longer than expected. This risk is moved from the pension provider to the receiver. One interesting question to investigate is thus what the value of the implicit longevity hedge of the guaranteed product was and whether a fair compensation has been given to those who switched to the market product.

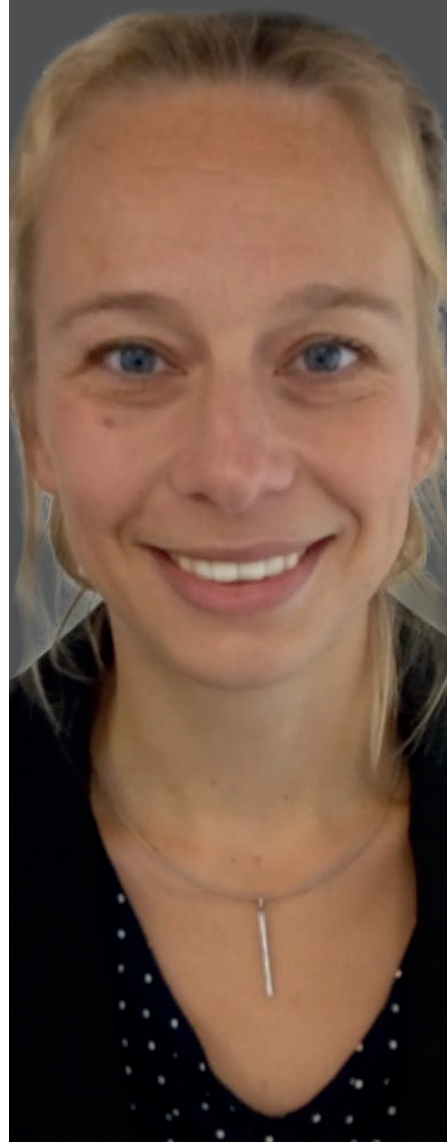
Based on the fact that most systems focus on a somewhat more individualistic perspective, personal preferences will play an increasing role in the design of pension products. Pension providers need to determine the investment strategy on behalf of the pension holders. How should they cluster these preferences? First, the new contract describes the importance of measuring one's risk aversion. What is yours? To do a small experiment, how much money would you like to receive with full certainty such that you are indifferent to receiving €50.000 with 50% and €100.000 with 50%? If you could get, as an alternative, €100.000 with 100% certainty, then you would obviously go for the latter option. You should figure out at which value you would like to switch to the risky option when decreasing the certain alternative. If you switch at the expected payoff of the risky event, i.e., €75.000, and we fit the certainty equivalent to a utility function with constant relative risk aversion, then you are, what we call, risk neutral. If your certainty equivalent is between €75.000 and € 54.000 your risk aversion level is in the range 0 to 10. This is the common range reported in experiments and in the literature. The next question is, what would be the associated optimal investment profile and what would be a fair one-size-fits-all strategy? Also for insurance companies, there is the question of how many different strategies they should offer. Moreover, by the experiment above you might realize that you do not even know your own risk aversion level. This indicates the presence of the ambiguity involved and thus highlights the demand for robust decisions.

I suppose that you will graduate within the transition period if you did not already do so. With the skills you developed here in Tilburg, I see a bright future in this vivid world of pensions!

Enjoy your summer. ●

Anne Balter

is an assistant professor at the Department of Econometrics and OR and a senior research fellow at Netspar. Her research interests include mathematical finance, in particular robust portfolio problems, pension economics, and real options.



Festivals This Summer: Is It Possible?

Already more than seven million people have been vaccinated throughout the Netherlands. Every minute, 186 more get vaccinated [1]. Also, more and more people receive the awaited white envelope invitation. While the vaccination numbers are increasing every day, questions arise on how this vaccination schedule is set up and how it has been adjusted throughout time. Also, since these rising numbers create an optimistic outlook for the future, what will students be able to do this summer? Where should we go on holiday and when will we get to go all out at a festival again?

written by **Casper Heemsker** and **Nienke Keuning**

At the moment of writing, those born before the year 1964 have received their invitations to get vaccinated. Vaccinations are expected to be done around September this year as seen in Figure 1. Those aged between 18 and 29 will qualify to get their first vaccinations around the end of June this year [2]. In January, this was still set for May and early June [3]. A possible reason for this setback is the stop of the AstraZeneca vaccine in March. This might have decreased the willingness of the people to get vaccinated [4]. Other factors that influence the vaccination timeline include approval of the vaccines, distribution of the vaccines, and new developments and advice from the government [5]. However, since the end of June is not so far away, the odds are definitely in our favor.

Statistically, a change in corona infections since the start of testing is definitely visible. The groups that are vaccinated are decreasingly seen in the hospital with corona symptoms. As seen in Figures 2 and 3, the number of positive corona tests for people aged above 90 made a decline. Of course, this is no surprise. For those between the ages of 80 and 89, this decrease was even higher [7]. Soon, it is expected to have similar graphs for the age groups which are yet to be vaccinated. This means that, hopefully, by the time that the summer comes around, not only seniors are able to enjoy a nice vacation.

Everybody wants a carefree summer where they can live without restrictions again. However, at this point, it is only speculating what is possible this summer. The holidays are getting closer. Normally, this is the time when everybody is looking forward to a trip to the Spanish coast or all kinds of festivals spread out in the summer months. Nowadays, the expectations are

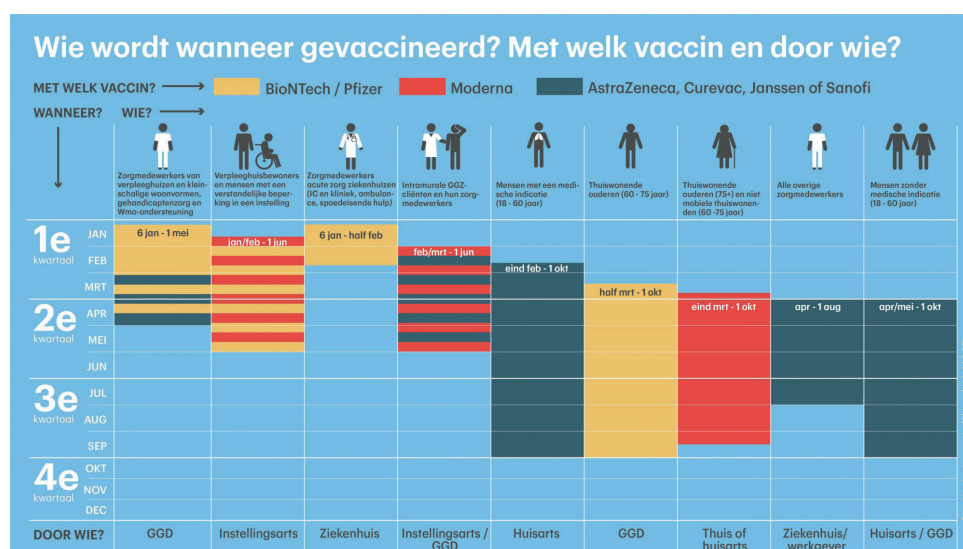


Figure 1: Dutch vaccination strategy [6].

Aantal positieve testen van mensen ouder dan 90 jaar door de tijd heen

— 90+ - - - Alle leeftijden

Dagtotalen per 100.000 inwoners

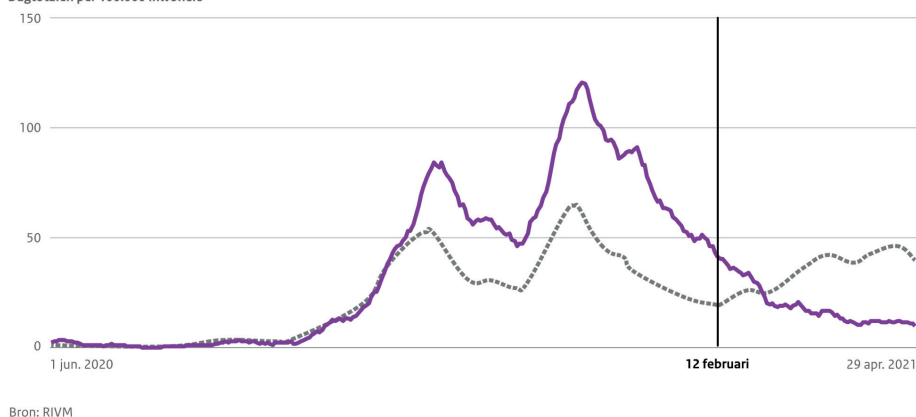


Figure 2: Amount of positive tests of people aged older than 90 throughout time [7].

Aantal positieve testen van mensen tussen 80 en 89 jaar door de tijd heen

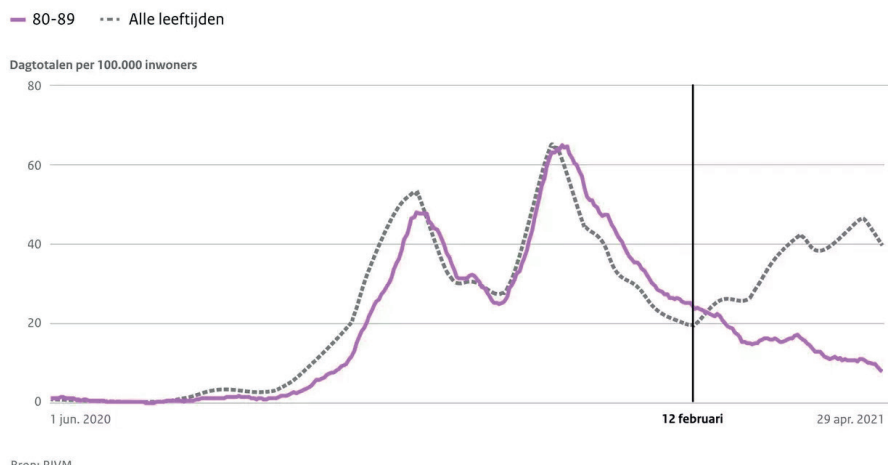


Figure 3: Amount of positive tests of people aged between 80 and 89 throughout time [7].

somewhat lower in general. But, there is some light at the end of the tunnel. Familiar festivals of the past years are already organizing new concepts for this year. The Dutch government really tries to help the event sector. Since the beginning of March, there have been multiple field labs to investigate the behavior of the people during all kinds of events. On top of that, the Dutch government has organized pilots. Pilots have been organized to see how the visitors of the event respond to testing before an event. Not all conclusions from these field labs and pilots are there yet but the first conclusions look promising [9]. In line with the field labs, the Dutch government has given a guarantee fund for organizations of events. According

to the Dutch government, festivals and other outside events are reasonable from the first of July. In case of cancellation because of the virus, the organization will get all of their costs reimbursed that they already had to pay in advance [8]. With this guarantee fund and the promising results from the vaccination strategy, the first festivals are already planned for July.

The most important question left will be: what are festivals going to look like? One simple answer is not possible at this point. The festivals in the Netherlands are diverse and so is the approach for this year. For example, the traditional festival Lowlands tries to organize everything as normally as possible [10]. Lowlands has announced

their line-up already. Over 50 artists will perform this year and the organization aims for more than 50000 visitors. Festivals will probably include a testing or vaccination policy. Every visitor needs to show a negative test or proof that he/she is vaccinated. After entering the festival area, the restrictions will probably be gone and everyone can dance and move freely.

At this point, it may seem very unrealistic: dancing with thousands of other people in a very strict area while enjoying the music of live artists. However, the foresight is hopeful. Many festivals have already canceled or postponed their edition for this year. Some big names in the festival sector like Pinkpop, Best Kept Secret, and WOO HAH! are already canceled. International events like Sziget and Tomorrowland are also canceled or only available for a small number of visitors. Nevertheless, most of the festivals are still on the calendar. Lowlands, Paaspop, Dutch Valley, and Down the Rabbit Hole are some familiar examples of the festivals that are still planned in August and September. Only the future will tell if the festivals will take place, but the perspectives are great.

So, this is what we expect to happen in the next few months. Of course, this is mostly speculation. We are very interested to see what actually happens and how things can change in just a few months. We are really looking forward to the summer, where we hopefully can enjoy the warm days with each other without the restrictions. ●

References

- [1] <https://coronadashboard.rijksoverheid.nl/landelijk/vaccinaties> (2021). Coronadashboard: COVID-19-vaccinaties. Rijksoverheid, May 16 2021.
- [2] <https://www.rijksoverheid.nl/onderwerpen/coronavirus-vaccinatie/volgorde-van-vaccinatie-tegen-het-coronavirus/volgorde-vaccinatie-voor-mensen-die-niet-in-de-zorg-werken> (n.d.). Volgorde vaccinatie voor mensen die niet in de zorg werken. Rijksoverheid.
- [3] <https://www.rijksoverheid.nl/documenten/publicaties/2021/01/12/vaccinatiestrategie-roadmap> (2021). Vaccinatiestrategie: planning (roadmap versie 12 januari 2021). Rijksoverheid, January 12 2021.
- [4] <https://www.rtlnieuws.nl/nieuws/politiek/artikel/5222469/de-jonge-prognose-vaccinaties-was-te-hoog-maar-we-prikken-zo-snel> (2021). We lopen achter op schema: minder mensen dan verwacht maken prikafspraak. RTL Nieuws, March 29 2021.
- [5] <https://www.rijksoverheid.nl/onderwerpen/coronavirus-vaccinatie/volgorde-van-vaccinatie-tegen-het-coronavirus/factoren-voor-de-vaccinatie-planning> (n.d.). Planning van vaccinatie tegen het coronavirus in Nederland. Rijksoverheid.
- [6] <https://www.rtlnieuws.nl/economie/artikel/5207455/coronavirus-inenten-vaccinatieschema-strategie-de-jonge-ema> (2021). Waarom het nationale vaccinatieprogramma wekelijks kan veranderen. RTL Nieuws, January 8 2021.
- [7] <https://coronadashboard.rijksoverheid.nl/artikelen/zo-zien-we-6-miljoen-prikken-terug-in-de-positieve-testen-en-ziekenhuisopnames> (2021). Coronadashboard: Zo zien we 6 miljoen prikken terug in de positieve testen en ziekenhuisopnames. Rijksoverheid, May 7 2021.
- [8] <https://www.nrc.nl/nieuws/2021/03/19/vanaf-1-juli-weer-naar-festivals-volgens-het-kabinet-moet-dat-kunnen-a4036481> (2021). Vanaf 1 juli weer naar festivals? Volgens het kabinet moet dat kunnen. NRC, March 19 2021.
- [9] <https://nos.nl/artikel/2378608-buitenevenement-als-voetbalwedstrijd-kan-veilig-met-helpt-publiek-zegt-fieldlab> (2021). Buitenevenement als voetbalwedstrijd kan veilig met helpt publiek, zegt Fieldlab. NOS, April 28 2021.
- [10] <https://www.npo3fm.nl/nieuws/3fm-nieuws/a533aed4-519b-43eb-a522-aab740a5a22e/hoe-gaat-lowlands-2021-eruit-zien> (2021). HOE GAAT LOWLANDS 2021 ERUIT ZIEN? 3FM, May 12 2021.

Optimizing the Personalized Product Recommendations within E-Commerce

During my Master thesis, I was given the assignment to create a model which recommends the best fitting products to an individual customer. At the start of the thesis, I was provided with a paper [1], which mainly included text mining models. Although this might seem a strange model to predict the best fitting products to a customer, there are actually many similarities between a piece of text and the shopping history on a website. Therefore, I was very excited to start my thesis and to try and find the optimal fit for every customer.

Predicting the next word in a sentence

In order to start the assignment of predicting the optimal products for a customer, my thesis started with a model [2] that predicts which words are close to each other, based on multiple text files. The idea behind this model is that words which are close to each other within the same text, are likely to be connected with each other. On the other side, words which do not appear close to each other are considered to be different from each other. In this model, we want to create pairs of words which have positive value between each other and pairs of words which have a negative value. This has been done by defining center words that get assigned a positive word, which is close to this word in the text, and multiple negative words. These positive word combinations are created by picking a word in a text and choosing n words before and after this word.

Take for example the sentence 'The Nekst is a quarterly magazine, which belongs to Asset Econometrics.' When we want to describe the word 'magazine' in this sentence and choose $n = 4$, then the words 'Nekst', 'is', 'a', 'quarterly', 'which', 'belongs', 'to' and 'Asset' all form a positive context pair with the word 'magazine'. These context pairs are now filled with negative words from the remainder of the text, that are sampled based on the following probability: $P_n(w) = \frac{1}{Z}(n_w)^p$. In this probability function P_n , n_w describes the amount of times a word is used within the text. This means that words that appear often in the text are more likely to receive a negative pairing. This is mainly useful to avoid that common words, such as 'is' and 'the' are too often considered as positive words. The power $p \in \{0, 1\}$ is used to scale the probabilities of common word. That is, when $p = 0$ all words would be treated equally and when $p = 1$, words that are used often are punished very badly. Finally, we also have the constant Z that is used to make sure that the probabilities of all words sum up to 1.

When we take the previous sentence, which is written within a practical report, as an example, we could have a context pairing for the word 'magazine' that has 'Nekst' as its positive pairing and 'shopping', 'the', and 'zoos' as its negative word pairings. When we sum over all different context pairings i in the text, with center word v^c , positive pairing w^p and negative pairings w^{n_j} , we can apply formula (1), which maximizes the probability of positive words appearing together and minimizes the probability of negative words appearing together.

$$\arg \max_{v,w} \left\{ \sum_i \left(Pr(w_i^p | v_i^c) \prod_{j \in n_{neg}} Pr(-w_i^{n_j} | v_i^c) \right) \right\}. \quad (1)$$

From formula (1), we now have two different matrices, namely center matrix $V \in \mathbb{R}^{L \times w}$ and context matrix $W \in \mathbb{R}^{L \times w}$. In these matrices, w are the amount of words in the text file, while L is equal to the amount of latent dimensions we use in our model. Within a latent dimension, the characteristics of the word are described. For example, the first latent dimension could be describing whether words are nouns. In this case, words that are nouns can all have a high value for this latent dimension, while words that are not nouns should have a low value. By calculating $W^T V$, we get a $w \times w$ matrix in which words that are likely to appear next to each other in the same text have a high value, while words that are not likely to appear together have a low value.

Applying a text mining model to predict e-commerce data

Now that we have predicted which words occur often together in the same sentence, the question arises how this can be used to predict which product a customer is most likely to purchase. As described by Gabel, Guhl, and Klapper [1], we could treat the shopping history on a website in a similar way as the words in a text. Within this shopping history, we have different customers that have purchased one or multiple products. When we compare this situation with the text file we have described

before, we can therefore compare the shopping history with the text and the different customers as sentences within this text.

Where we before created a positive context pairing when words were close to each other in the same text, we can now create this positive context pairing for products that have been purchased by the same customer. When we assume that a customer purchases a product for its own benefit and not as a present for someone else, we can argue that products purchased by the same customer are likely to be purchased by a similar type of customer as well. To give an example, we can think of a sports clothing shop in which one customer purchases football boots and shin pads while another customer has purchased a tennis racket and a headband. When a new customer comes in and is looking at football boots, we are more likely to recommend this customers shin pads than we would recommend a new tennis racket. By creating positive and negative context pairing in a similar way as for the text mining model and by replacing words w with products p in formula (1), we can now create $L \times p$ matrices W and V , containing the likelihood for every product pairing to be purchased by the same customer.

But, the difference between predicting the next word within a text file and predicting the next purchase of a customer, is that most customers have already purchased multiple products. This does, however, mean that we now need to know which product is most likely to be purchased next for a group of products. In order to cope with this problem, we can use the Skip-Gram model, which is described by Rong [3]. The idea behind the Skip-Gram model, which is illustrated in Figure 1, is that we create for every product p which has been purchased by this customer c a vector b_p which is 1 for product p and 0 for all other products. This vector b will now be multiplied with center product matrix V , resulting in a $1 \times L$ hidden layer vector h . After this hidden layer vector h has been calculated, we will multiply it with the context matrix W , resulting into an output vector y , in which similar products to the input product have a high value. Now, we want to calculate this output vector y for every product in the input of this customer c , resulting in vectors y_1, \dots, y_n in case this customer has purchased n products. Finally, we can now calculate the output vector y_c for customer c using formula (2), where b_c is the purchase history of customer c .

$$y_c = \sum_{p \in b_c} \frac{1}{n} \cdot y_p \quad (2)$$

The products with the highest values in y_c are the products that will be recommended to this customer.

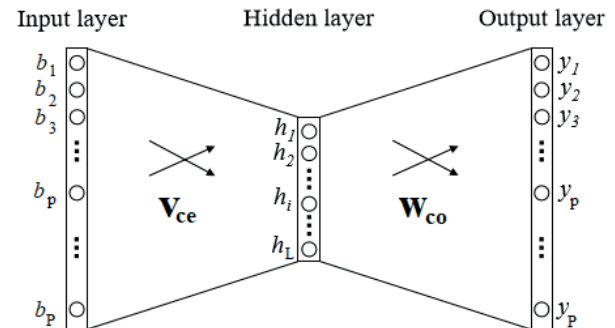


Figure 1: The Skip-Gram Prediction Model

Applying a clustering algorithm to discover which products are similar to each other

Up until now we have discussed a model that can be used to recommend products to a customer, based on their purchase history. Although this will of course be useful for customers that are familiar with the website, we also have to take into account that there will be new customers that are visiting a website for the first time, from which we now have information about. For these customers we can not make any reasonable recommendations, but we can create recommendations which are attached to other products. This way, when a customer would select a certain product, we can easily recommend additional products, which are linked to this product.

When we want to recommend multiple products based on one product page which has been visited, a clustering algorithm can be applied. By clustering the different products that are sold on this website, we can determine which products have similar characteristics to each other and therefore are more likely to be either substitutes or complements to this product. A clustering algorithm that is fitting for these circumstance is the t-Distributed Stochastic Neighbor Embedding Algorithm [4] or shortly abbreviated t-SNE. When we want to analyze a product, this product can have multiple different layers containing information about this product. When we take for example a carton of milk, certain characteristics we can already determine from this product are the fact that this is a drink, the color is white, it is packaged in a carton, and it is a dairy product. Next to these factual observations, we can also have observations like: the products are mainly purchased by female customers, customers who are around 30 years old, and customers who are shopping in the morning. Every product can therefore have many different layers of information, which makes it difficult to compare it with different products.

A characteristic of the t-SNE clustering, is that when using it, we can visualize high-dimensional data into a two or three dimensional map, which makes it easier to compare products with each other.

In our model, we can apply t-SNE clustering on the center matrix V , which has L layers of information about every product. By summarizing these L layers of information into a center matrix \hat{V} containing two layers, we can more easily compare the similarities of products, which makes it also possible to visualize the products in one graph. Without going to much into the mathematical details, t-SNE aims to create proportional probabilities using a Euclidean distance function. When we have P different products, this means that we can create P different vectors: $\{v_1, v_2, \dots, v_P\}$ containing the L different layers for every product. For these P different vectors, we can calculate the Euclidean distance $d(v_i, v_j) = ||v_i - v_j||$ between every pair of product vectors. When we apply t-SNE clustering, the Euclidean distance of the two-dimensional vectors $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_P\}$ is proportional to the original vectors.

Based on the data that has been used during the thesis, it is possible to see what the results of this t-SNE clustering are. In Figure 2, we can see the illustration of 450 different products that are sold on a certain website within e-commerce. In this picture, we can see approximately twenty different product groups based on the data. In the top right of this picture, we can see different types of days-out. We can see for example that group 1 mainly contains theme parks focused on younger children, group 3 contains theme parks outside of the Netherlands, and group 7 contains theme parks within the Netherlands. Group 2, 4, and 5 all contain tickets to different zoos, which are spread over different locations of the Netherlands. On the other side of the picture, we can see different types of textile products, which appear to be the total opposites of day-out tickets. In group 20, we recognize different types of bed linen, while in group 21 we can recognize towels and washcloths.

Internship

I did my research on this topic during an internship at SphereMall. During this internship, I was given the opportunity to work with data of a client within e-commerce who sells many different types of products to their customers. The value that SphereMall mainly wants to provide to their clients is an analysis of how customers are behaving on their websites and extend this with models which can guide these customers more easily towards finding the product they are looking for. ●

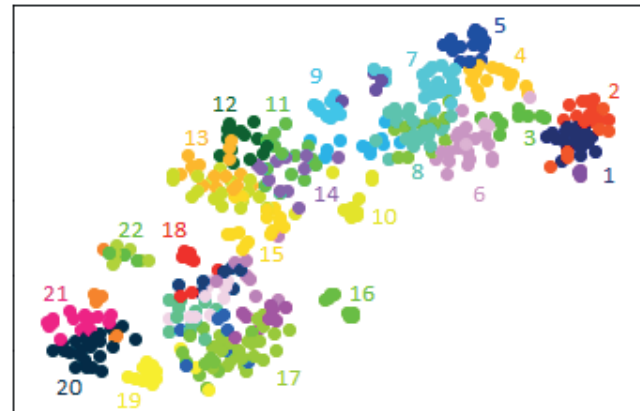


Figure 2: t-SNE Results of E-Commerce Products

References

- [1] Gabel, S., Guhl, D. Klapper, D. (2019), "P2V-MAP: Mapping Market Structures for Large Retail Assortments." *Journal of Marketing Research*, 56(4), 557–580.
- [2] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013), "Distributed Representations of Words and Phrases and their Compositionality." In *NIPS*.
- [3] Rong, X. (2014), "Word2vec Parameter Learning Explained." arXiv preprint arXiv:1411.2738, 2014.
- [4] Van der Maaten, L. (2014), "Accelerating t-SNE using Tree-Based Algorithms." *Journal of Machine Learning Research*, 15, 3221 - 3245.



Jeffrey Buijk
Graduated Master BAOR (2021)

Keeping Up With The Quantitative Investment Group

The Quantitative Investment Group (QIG) is a small group of econometricians that use quantitative modeling to make predictions about individual stock prices. We use these predictions to invest almost automatically in the stock market every month. Despite that our predictions are not entirely accurate, the concept of having this goal is amazing. To accomplish this goal, we apply simple known models like a Fama-French Three-factor model but also more complex Machine Learning algorithms like a random forest. These models give an outcome of best stocks and according to this result, we form our new portfolio for the coming four weeks.

In the QIG, we have several research groups that program a certain project. This project could be developing a new strategy, improving an old strategy, or automating standard procedures like reporting. Each month, everyone discusses their progress with the rest of the group. I think joining the QIG is the ideal way for econometricians to gain experience by programming econometric models, valuing companies, and being active in the financial market.

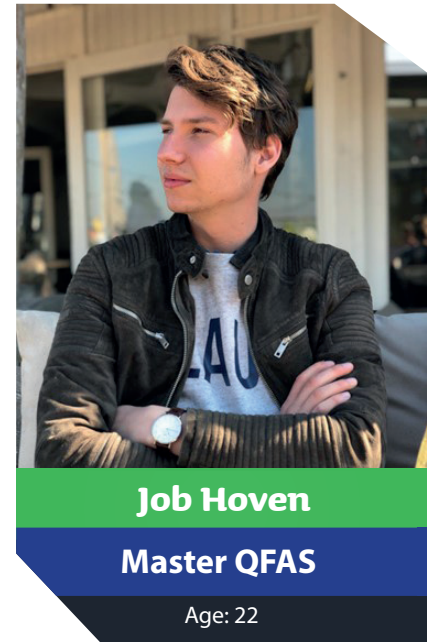
From time to time, we visit companies or attend masterclasses to provide insights to our members. Last month, we attended an E-houseday provided by Northpool B.V., a commodity trading company with a strong focus on energy products. This is

a young high potential trading company that is looking for econometricians. They buy energy contracts to make a profit with energy price fluctuations. To this end, they use predictive modeling on several aspects like the weather to decide if they are going to buy or sell these contracts.

During this E-houseday, they introduced themselves as a company and provided a fun trading game. In this trading game, we were separated into different groups and played a simple simulation of a commodity trader within their company. We got some information and needed to decide how much Watt to buy or sell for the next hour. The goal of the game was to make a high profit. Now I think back, I happen to remember that the QIG had a high profit in general compared to other people who played the game according to Northpool B.V.. This was nice to hear!

After the game, a trader of the company told us about his experience within the firm and provided us the chance to ask questions. Noticeably, the company seems to exist almost exclusively out of people with a more scientific or mathematical background than a financial one. The trader said that he also had a scientific background but that he was very interested in the company either way and decided to work there. I agree with him, it is extremely cool what they do!

In addition to this E-houseday, there were also some other interesting happenings during our last quarter. Each quarter, we look at the overall performance of our portfolio for the last period. We have had



Job Hoven

Master QFAS

Age: 22

an extremely good period. This might be due to overfitting but it is nice to see these results. As chairman, I am responsible for creating and improving this quarterly report together with implementing and discussing policy changes. Currently, we are discussing if we want to keep short strategies in our portfolio, for example. It is nice to see these active discussions within the group. This indicates the involvement of our current members and keeps the meetings interesting.

Last January, I took over the role of chairman from Jelle Thijssen who had done it for a year and a half. Before January, I assisted him as treasurer and I liked working together with him. It was nice that we were both very passionate about the investment group. I wish him good luck in his future career! At the moment, I have already made some adaptations to the group like the quarterly report, so I am curious to see how that unfolds in the future! ●

"I think joining the QIG is the ideal way for econometricians to gain experience by programming econometric models, valuing companies, and being active in the financial market."

Let's Talk

After almost a year and a half of dealing with COVID-19 and a heavy lockdown of a few months, the corona measures are finally starting to loosen up again. As the school year is almost over, it looks like there might be more possible during the summer break. Since we all cannot wait to get back to our favorite pubs with some friends, let's hope the measures loosen up some more in the nearby future. We sent this questionnaire to our members of Asset | Econometrics asking about their upcoming plans for the summer. We also wondered if they will go to on-campus lectures again when possible and if they are going to get vaccinated. The results of the survey with some fun comments are shown in this Let's Talk! After a long and difficult year at the university, we hope that you all did well on your final exams. Enjoy your well-deserved holiday!

written by **Troy de Juncker**

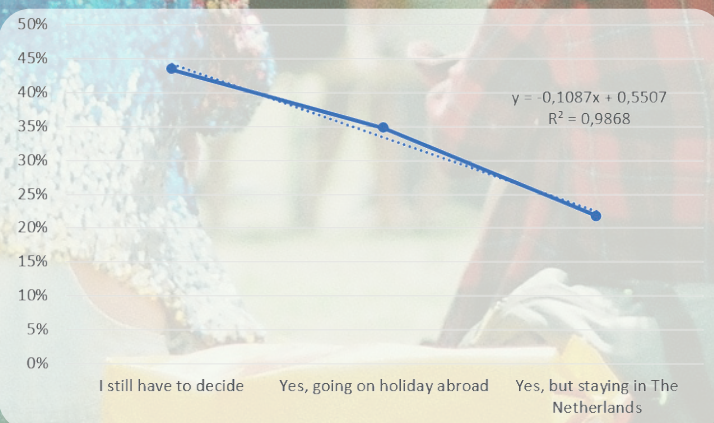
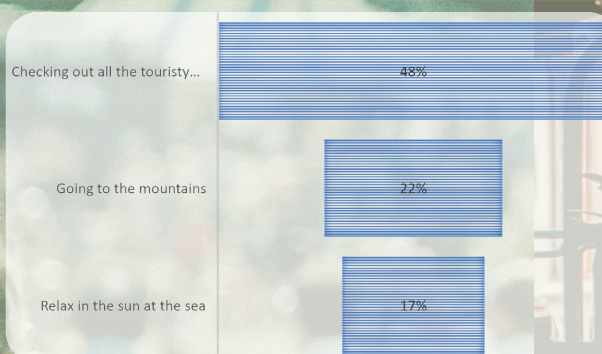
What is your ideal holiday?

Marieke

A combination of all answers. I love a good mountain hike, but a beach day in between will make it complete.

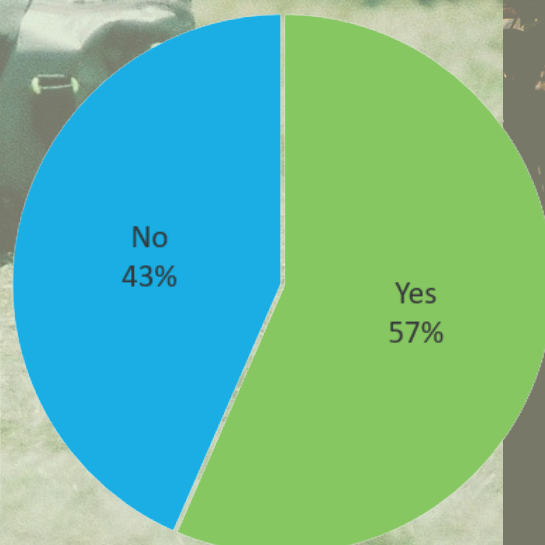
Emma

A combination of relaxing at the beach and checking out all the touristy hotspots. However, it should go together with drinking some beers in the evening!



Do you already have holiday plans?

Since it possible again, have you had lunch/dinner at a terrace yet?



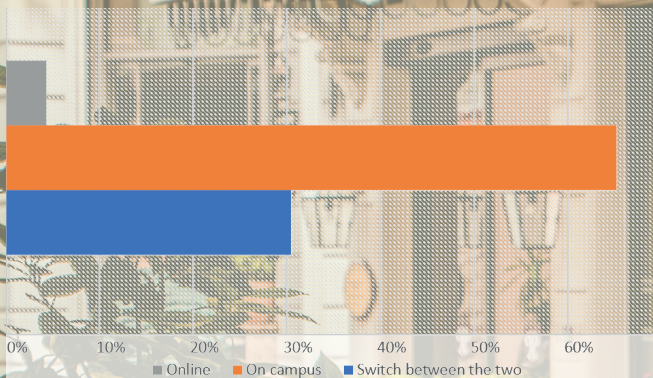
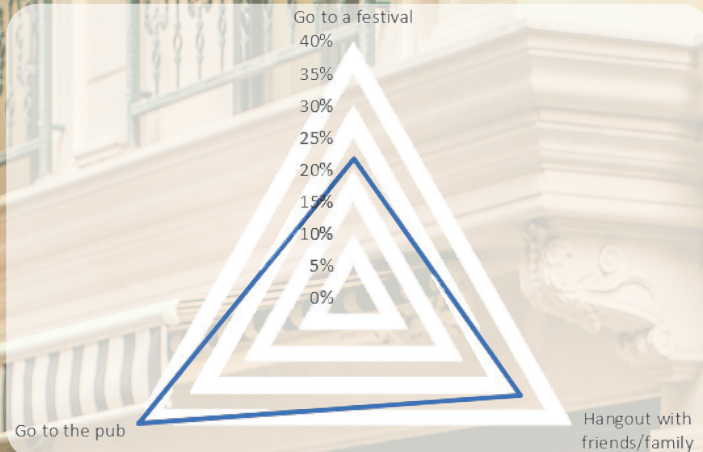
What is the first thing on your to do list when all the measures are gone?

Manon

I already went shopping and hung out with friends and family. So, the next thing to do is going to festival!

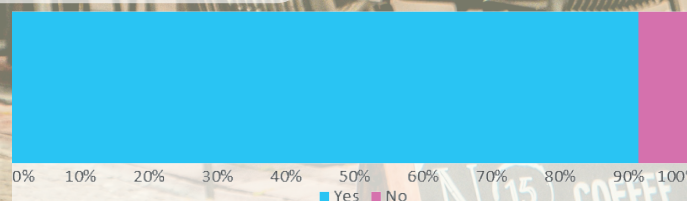
Britte

One answer: Winter sports!

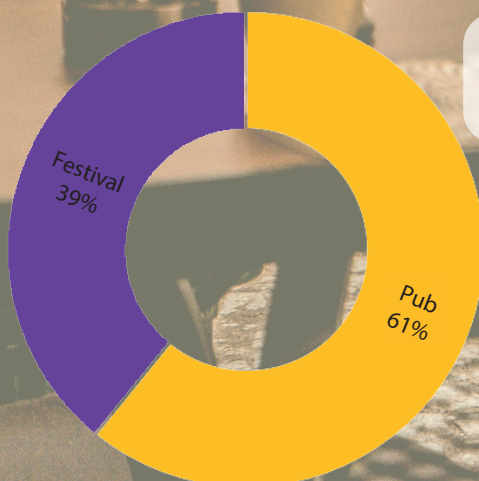


When it is possible, are you going to follow all lectures on campus or do you prefer them online?

Are you going to get your vaccination?



Would you rather go to a festival or go for a night out at the pub?



Quatsch!



Quatsch?

Over the past few months, the editorial staff of Nekst received many quotes that relate to the study of Econometrics and to the activities organized by Asset | Econometrics. Hereby, we present to you a selection of some striking and funny quotes! Please send in your quotes at: www.Asset-Econometrics.nl/more/nekst/Quatsch

Bram

"Hij kan echt niet voetballen"

Juul

"Bram je zit in heren 9"

Joris

"Oh dit is wel heel moeilijk rekenen, wat is 5 keer 8 ..."

We moeten 36 seconden omzetten naar minuten

Britte

"36 seconden, is dat niet gewoon één minuut?"

Jax is de hond van Emma

Bob

"Heb je er wel eens over nagedacht dat Jax gewoon nog maagd is?"

Tjum

"Steen papier schaar, hoe werkt dat?"

Melissa: "Welke commissie doe jij Rico?"

Ricardo: "AMD"

Stephan: "Rico je zit in promo..."

Juliëtte

"Ik snap het woord percentage niet meer"

Dear Nekst committee,

What a weird year it has been. Creating four Nekst magazines without even seeing each other in real life was not always an easy job. We were only able to write together on Zoom, but in my opinion, each Nekst contained well-written and interesting articles. We strived for the right balance between informal and formal articles while keeping up with the 'hot' topics. As of now, I am quite sure that you are fed up with me texting you about Nekst, and whether you have finished your articles already. Nevertheless, I hope that you have enjoyed this committee! Since this the last edition, I would like to thank each and every one of you individually in this letter.

*To start with **Michelle**, our coordinator. You were motivated to make Nekst a great success from the start. You were always enthusiastic and made sure that we all kept our spirits high. Thank you for making sure each Nekst contained three columns, a practical report, and a triangle.*

*Next up, we have **Casper**. I really appreciated your input and motivation. You showed your amazing writing skills in every type of article you wrote. No matter if it was a meet the crew, the teacher, or an interview with Herbert Hamers. Dominique even said that you are a talented writer and I completely agree.*

*Then, I would like to thank **Jarno** for joining the Nekst committee again, after already being in it for half a year. We could always count on your humorous comments and pleasant collaboration. I really enjoyed writing three articles together with you: a company interview, a special about investment strategies, and the teacher about Hein Fleuren!*

*Now, it is time to put our MVP in the spotlight: **Tamara**. You were the creative brain behind Nekst this year. From great ideas for Let's Talk, a Where is Waldo, a company interview, to proposals for specials (and writing them, of course). I loved how you would text me out of nowhere with fascinating topics to write about.*

*Moreover, I would like to thank **Nienke Keuning**. You did not interview one teacher, but two: Cristian Dobre and Martin Salm! Furthermore, you also wrote a well-founded and highly relevant special about the American Voting system. I want to thank you for your enthusiasm and for delivering great work.*

*Cheers to **Aimee** who was the only first-year student in the committee. Every Nekst edition, you checked multiple articles without me asking. I would even go as far as calling you my personal article checker. Also, you did a great job writing your first company interview with Ortec Finance.*

*Thanks to **Nienke Kempers** for loving our association so much that she joined a committee without actually being an econometrician anymore. For me, you will always be an econometrician by heart, and I thank you for checking all the articles when I needed you to.*

To **Marijn**: thank you for your work for Nekst. You joined Jarno and me for the interview of Dassault Systemes with Christel Opheij. You can be proud of the final result of this three-page interview.

The Nekst committee contains many people, but for the last two editions also **Mauveen** joined us. Many thanks for being enthusiastic and being a nice addition to our committee! You immediately joined the task force to organize a great committee event and you wrote an interesting special in our latest edition.

Then, I like to thank **Karel**. You wanted to be an active member so bad that you called the university to get in touch with the board. Eventually, you joined our committee while still being in your homeland Ecuador!

This Nekst committee sets itself apart from other years by having, among other things, a highly-motivated Nekst-Online staff. Thank you, **Luuk** and **Patrick**, for giving our plain-looking website an entire makeover. The site is now faster and more attractive to look at. It is also easy to find the articles you want to read. I really appreciate all of your effort in changing and promoting Nekst-Online.nl.

I would also like to congratulate you, **Luuk**, for doing a great job of not listening to yourself. You said at the beginning of the year that you would only join the Nekst if you did not have to write the articles. Somehow along the way, you convinced yourself to do the opposite and wrote not one but two articles including a great special about chess!

Then, I would like to thank our puzzle master, **Patrick**. You were in charge of the puzzles of all four Nekst editions. Instead of googling boring and common riddles, you created all puzzles on your own. For each puzzle, you even included an elaborate explanation of how to solve it. If econometrics does not work out for you, you might consider starting your own puzzle company?

Last but not least, I want to express my gratitude for our designer: **Dahli**! Thank you for putting up with me this year and listening to my nagging about everything Nekst related. It is just that you were so crazy to take this job upon you. At the moment, I am designing this letter, because you are in the hospital. And it is safe to say that you would probably have done a much better job than I am doing right now. Hopefully, you will get better soon (and help me out!). I really enjoyed working together with you in the late hours to make the magazine as perfect as possible this year. You particularly surprised me with your creativity and eye for detail. Sometimes, we (read: you especially) even put in too many hours to align photos, stripes, and dots, but the final results speak for themselves. Thank you for being my partner in crime this year, the magazine would not even exist without you!

Love,

Your favorite Editor-in-Chief
Mylan

Asset | Econometrics congratulates...

Name **Steffie van Poppel**

Title Simulating the volatility of the predicted
experience rate mortality factors: A case study

MSc EME

Supervisors Dr. M. Salm, Prof.dr. B. Melenberg

Name **Sjoerd Trommelen**

Title The use of Long-Short Term Memory networks
in pairs trading strategies

MSc QFAS

Supervisors Dr. P. Cizek, Prof.dr. B. Melenberg

Name **Marco Alberti**

Title The dynamic causal effect of a monetary shock
A smoothed local projection IV estimation with
relative selection criterion

MSc EME

Supervisors Dr. O. Boldea, Dr. D. Kojevnikov

Name **Lennart Sinjorgo**

Title Multicoloring of highly symmetric graphs

MSc BAOR

Supervisors Dr. R. Sotirov, Dr.ir.ing. M.J.P. Peeters

Name **Jin-Yong Ruan**

Title Time Varying Factor Modeling and Asset Allocation
applied to an insurance balance sheet

MSc QFAS

Supervisors Dr. N.F.F. Schweizer, Prof.dr. B. Melenberg

Name **Peter te Marvelde**

Title Identifying the impact of government debt on
economic growth in developing countries

MSc EME

Supervisors Dr. P. Cizek, Dr. S. Sadikoglu

Name **Dafni Lappa**

Title Risk averse procurement portfolio optimization:
effect of updated information and correlations

MSc BAOR

Supervisors Dr. Y.Merzifonluoglu Uzgoren, Dr. J.C. Wagenaar

Name **Maciej Nowara**

Title Pricing equity derivatives with neural network-based
calibration — a comparative study

MSc QFAS

Supervisors Dr. N.F.F. Schweizer, R. Fu MSc.

Name **Jeroen Kouwenberg**

Title Valuing credit risky loans with stochastic interest
rates

MSc QFAS

Supervisors Dr. R. van den Akker, Dr.ir. G.W.P. Charlier

Name **Renata van der Knaap**
Title Cyclic railway timetabling using an iterative SAT approach with a feedback mechanism
MSc BAOR
Supervisors Dr. R. Sotirov, Dr. Y. Merzifonluoglu

Name **Wiebe Kuper**
Title Stochastic jump diffusion processes and their application to asset price modeling
MSc QFAS
Supervisors Prof.dr. B. Melenberg, Dr. F.C. Drost

Name **John van den Hurk**
Title Transfer Learning For Price Prediction In Real Estate
MSc BAOR
Supervisors Prof.dr.ir. H.A.M. Daniels, Dr. J.C. Vera Lizcano

Name **Karsten van Assel**
Title Hybrid Optimization of Replicating Portfolios for Stochastic Liabilities: Improving stress test performance of optimal cash-flow replicating portfolios
MSc QFAS
Supervisors Dr. F.C. Drost, Dr. P. Cizek

Name **Emina Tandirevic**
Title The impact of (local) differential privacy on model performance for binary classification problems
MSc EME
Supervisors Dr. R. van den Akker, Prof.dr. B.J.M. Werker

Name **Annikie Bastiaansen**
Title Using panel data models to analyze mental health during the COVID-19 crisis among the general population in the Netherlands
MSc EME
Supervisors Prof.dr. A.H.O. van Soest, Dr. B.M. Siflinger

Name **Joris Pirée**
Title The Implications and Effects of Lowering the Carbon Footprint of a Passively and Actively Managed European Equity Portfolio: a Quantitative Approach
MSc QFAS
Supervisors Prof.dr. B. Melenberg, Dr. F.C. Drost

Name **Bianca Denekamp**
Title Optimization of battery swap routing for an action of battery swap routing for an e-scooter sharing company
MSc BAOR
Supervisors Dr. J.C. Wagenaar, Dr.ir.ing. M.J.P. Peeters








...on obtaining their Master's degree

Need Study Support?

www.Asset-Econometrics.nl/Study

You can find old exams and more on our website!

Guidelines

-  Introduction Analysis and Probability Theory
-  Mathematical Analysis 1
-  Mathematical Analysis 2
-  Probability and Statistics
-  Statistics for Econometrics
-  Games and Economic Behavior
-  Games and Cooperative Behavior

Give us your feedback!

We try to contribute to the quality of the education within the Econometrics department in several ways. For example, we take part in soundboard meetings with the Academic Director(s) and the Program Coordinator. Do you have feedback on one of the Bachelor or Master courses? Contact one of the board members or fill in the Course Evaluation form on our website!

Write a Guideline!

We are always looking for students willing to write a new guideline. For every decent self-written guideline that is published you will be financially compensated up to €150,-. Do you want to publish a guideline? Contact us via Education@Asset-Econometrics.nl!

ASSET



Econometrics

Calendar

September

- 3-5. Active Members Weekend
- 8. TUC
- 15-16. Banking Days
- 19. QIG Masterclass
- 20. Monthly Afternoon
- 24. Sports Activity
- 28. Game Afternoon
- 29. Inhouse Day

October

- 1-2. Freshmen Night
- 5. Hackathon
- 6. InDesign Training
- 7. LEST
- 11. EOR Business Dinner
- 13. AMM
- 26. OG Activity
- 27. Board Information Session
- 29. Sports Activity

November

- 2. Econometrics in Practice Day
- 4. COdE
- 9. Monthly Afternoon
- 11-15. Europe Trip
- 17-19. Finance Expedition
- 23. Cycling Dinner AMM
- 24. Econometricians for Society
- 26. Active Members Day
- 29. Inhouse Day
- 30. Freshmen activity

Register and find more information about our events at
www.Asset-Econometrics.nl/events

NOTHING BEATS TRADING

GRADUATE TRADER

SUMMER TRADING INTERN

We are looking for talented individuals with excellent mathematical and analytical skills combined with an interest in global financial markets.

Our Traders manage and optimize our daily positions, formulate innovative trading strategies whilst also developing tools. As a Summer Intern you will learn about our trading strategies, trading system and experience life as a Trader at Flow.

Reach out to us at:

careers.europe@flowtraders.com
flowtraders.com/careers

Flow Traders is an international leading principal trading firm.

FLOW ■ TRADERS

