Interactive Simulation and Visualization of Long-Term, ETF-based Investment Strategies

Martin Büßemeyer
Hasso Plattner Institute, Digital Engineering Faculty
University of Potsdam, Germany

Daniel Limberger
Hasso Plattner Institute, Digital Engineering Faculty
University of Potsdam, Germany

Willy Scheibel
Hasso Plattner Institute, Digital Engineering Faculty
University of Potsdam, Germany

Jürgen Döllner
Hasso Plattner Institute, Digital Engineering Faculty
University of Potsdam, Germany

ABSTRACT

Personal, long-term investment products, especially ones for retirement savings, require thorough understanding to use them profitably. Even simple savings plans based on exchange-traded funds (ETFs) are subject to many variables and uncertainties to be considered for expected and planned-upon returns. We present an interactive simulation of an ETF-based savings plan that combines forecasts, risk awareness, taxes and costs, inflation, and dynamic inflows and outflows into a single visualization. The visualization consists of four parts: a form-fill interface for configuration, a savings and payout simulation, a cash flow chart, and a savings chart. Based on a specific use case, we discuss how private investors can benefit from using our visualization after a short training period.

CCS CONCEPTS

• **Human-centered computing** → *Information visualization.*

KEYWORDS

interative visualization, financial visualization, 401(k) plan visualization, ETF savings plan visualization, retirement savings calculator

ACM Reference Format:

Martin Büßemeyer, Daniel Limberger, Willy Scheibel, and Jürgen Döllner. 2021. Interactive Simulation and Visualization of Long-Term, ETF-based Investment Strategies. In *The 14th International Symposium on Visual Information Communication and Interaction (VINCI '21), September 6–8, 2021, Potsdam, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3481549.3481568

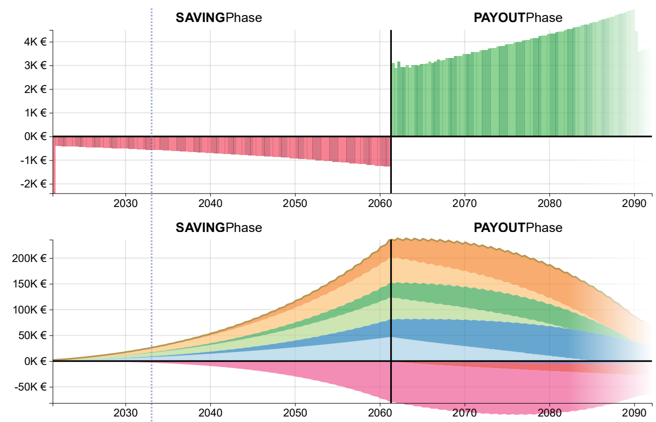

Figure 1: Our visualization of a plan with a dynamic € 100 monthly investment split into three accumulating ETFs. The cash flow chart (top) depicts the monthly investments and payouts. The depot chart (bottom) shows the simulated forecast of the savings, taxes (red), costs, and inflation (magenta) over a time-span of 40 (saving) plus 30 (payout) years. This shows, i.a., a dynamic payout of € 800 with the start of the pension is reasonably sustainable for at least 20 years.

1 INTRODUCTION

Pension entitlement is currently falling continuously due to the demographic change [4, 21]. To make up for a gap in personal retirement provision, one can resort to offers from financial institutions. For example, *Riester* in Germany [4, 19] or the more widely known *401(k)* plans in the USA [23]. The underlying strategies are often actively managed and produce unnecessary costs, while passive strategies, tend to win over active strategies in the long run [11]. This is one of the reasons why passive investment strategies based on long-term saving plans using exchange-traded funds (ETFs) have become quite popular. more attraction in other countries. Since most direct banks and brokers provide opportunities for regular, small investments in ETFs with reduced costs, it is fairly easy to minimize costs and administrative expenses in order to maximize savings. The ETF investment system itself, however, is rather complicated and requires stock market literacy [2]. Therein lies a fundamental conflict of interest, as forecasts from brokers or banks themselves are not designed to be transparent to the customer. In order to maximize incentives to create a depot everything is kept as simple as possible. This leads to over-simplified communication of costs, taxes, and idealized forecasts [6]. The main contributions of this work are

- a savings plan simulation that takes growth, dividends, taxation, and transaction costs into account, and
- a visualization using a cash flow chart and a savings chart accounting for inflation, taxes, costs, etc.

VINCI '21, September 6–8, 2021, Potsdam, Germany

ACM ISBN 978-1-4503-8647-0/21/09…$15.00
https://doi.org/10.1145/3481549.3481568

"The most straightforward (and generic) approach to enhancing the value of an unknown outcome is to communicate concrete consequences" [22]. Thus, self-efficacy with regard to financial decision-making improves the likelihood that individuals follow the retirement strategy [13]. "[A] recent study found that less than 25% of [American] individuals over the age of 50 succeeded at saving for retirement, often because they were unaware of fundamental economic concepts such as interest rate or inflation" [17].

We propose an interactive savings simulation and target non-experts to support decision-making on ETF-based saving plans (Figure 1). We enable interactive examination of various aspects of a savings plan by making all calculations in real time and provide immediate visual feedback. This allows users to (1) understand the interactions of parameters and risks during savings and payout phase, and to (2) identify scenarios that fit their own needs. Our open source prototype focuses on Germany's tax law (could be adapted to other countries) and is available at etf-vis.net.

2 RELATED WORK

"Financial literacy empowers people to craft their finances in a way that they are able to manage their everyday expenses, maintain an emergency fund, plan for children's education and prepare for their swift post-retirement years" [9]. This becomes more important since employers and the federal government forward the responsibility to the individual resident [13]. This literacy is understood in a broader sense and "[v]isualizations may be [...] [a] good approach for overcoming processing difficulties because they shift information processing to the perceptual system, improving understanding of the concept presented and allowing decision-makers to quickly learn from trends and patterns in the data" [13]. For example, visualization is used to get an overview on whole markets and publicly available finance data [20], or underlying mechanics and consequences [8]. Visualization has also become available for a more personal use [10] and was used to assess and communicate risks for savings [17]. for growth. We were surprised to see how few tools for forecasting and visualizations are currently available for end users. Those available, however, usually kept relevant factors like costs, inflation, and taxes out of consideration [3, 7]. Similar issues are present in user driven visualizations that generalize across country borders and thus cost and taxation models [18] though they are more transparent in their simulations.

There are several statistical models used for predicting stock markets. One prominent model is ARIMA. Though research has shown that ARIMA only works in some cases for forecasting short-term periods [1, 5, 14]. Moreover, we are aiming for long-term forecasting periods. We can see that the market evolved in an exponential fashion during the last decades [12]. Thus, simpler regression models seem sufficient for our long-term forecasting.

3 ETF DATA & SIMULATION

Our simulation provides access to several ETFs, uses historical data for forecasting, and takes costs and German taxes into account.

Source Data. We use historical data from Alphavantage to forecast prices. Their API provides the average monthly closing value for the last five years of many well-known ETFs. Similar services are available from Yahoo Finances and xignite.

German Costs and Taxation Model. German citizen have a yearly exemption amount that can be used on investment gains to save taxes. The tax amount is calculated by multiplying the gain by the partial exemption rate of 0.7. This prevents double taxation that would occur since the ETF owner has already paid taxes. The tax rate is composed of corporate income tax and the solidarity contribution, which add up to 0.26375%. In Germany, the FIFO principle is prescribed for the sale of shares: the oldest shares are sold first. Since these are likely to have the highest profit, taxes due at the beginning of the payout period will be highest and decrease over time. Accumulating and distributing ETFs differ in the way dividends are distributed and taxed. Compared to distributing ETFs, accumulating ETFs reinvest the distribution automatically. Instead of a dividend, the investor receives additional shares and no taxes are due. In 2018, the *Vorabpauschale* (advance flat taxes) was introduced as part of an investment tax reform. The taxes that need to be paid for a year's profit are limited by the expected profit. The expected profit is calculated by multiplying the amount invested by the *base interest rate* and is discounted proportionately for each month of the year that has already passed. Finally, the money distributed for the year is deducted from the resulting amount. Thus, there are effectively no taxes for distributing ETFs. Other costs arise from brokerage fees, usually a mix of fixed amounts and percentage costs.

Growth Model. Statistical models like MA and ARIMA are used for short-term forecasts. Since we have to forecast far into the future, these models are not suitable. Instead, models capturing general price development of recent decades seem to be sufficient. Even though, prices have continued to rise exponentially [12], we decided to be pessimistic and assume linear model inspired by [15] in our simulation to avoid overly optimistic forecasts. This can be easily replaced by an exponential model, if desired, or adjusted directly by means of a confidence score on the price development.

Savings. The savings phase starts with an optional starting capital and continues with a static or dynamic monthly amount. New shares are calculated by dividing the investment amount without costs by the current price. The *Vorabpauschale* and payout of money take place in December. If a distributing strategy is used, taxes on returns are considered. The payout is then reinvested without additional costs and the *Vorabpauschale* for this year is calculated.

Algorithm 1: Pseudo code of the calculation for a saving month.

```
distributions ← 0
costs, investment ← calculateCosts(investment)
newShares ← investment/currentPrice
if currentMonth == December then
    if isDistributionModel then
        distributions ← distribution
        taxes, distribution, taxFree ←
          applyTaxes(distribution, taxFree)
    newShares ← newShares + distribution/currentPrice
    expectedGain ← calculateExpectedGain()
    vorabP ← clamp(gain, 0, expectedGain − distributions)
    taxes ← taxes + applyTaxes(max(0, vorabP − taxFree))
inflation ← totalValue − totalValue ∗ (1 − 0.02)^yearsPassed
return newShares, costs, taxes, inflation
```

Algorithm 2: Pseudo code of the calculation for a payout month.

```
receivedPayout ← 0
if currentMonth == December then
    if isDistributionModel then
        distributions ← distribution
        taxes, distribution ← applyTaxes(distribution)
        payout ← payout − distribution
    else
        newShares ← distribution/currentPrice
for oldestInvestment inv: until payout == 0 do
    gain ← calculateGain(inv)
    newTaxes, newPayout, taxFree ←
      applyTaxes(gain, taxFree)
    payout ← payout − inv
    receivedPayout ← receivedPayout + (inv − newTaxes)
    taxes ← taxes + newTaxes
if isDistributionModel then
    expectedGain ← calculateExpectedGain()
    vorabP ← clamp(gain, 0, expectedGain − distributions)
    taxes ← taxes + applyTaxes(max(0, vorabP − taxFree))
costs, receivedPayout ← calculateCosts(payout)
inflation ← totalValue − totalValue ∗ (1 − 0.02)^yearsPassed
return newShares, receivedPayout, costs, taxes, inflation
```

Finally, the effectively lost amount of money caused by inflation is calculated. This process is shown in Algorithm 1.

Payouts. During the payout phase, a specified monthly payout is withdrawn from the depot. For the distributing variant, payouts in December are not reinvested directly, but used for the monthly payout first. Next, an iteration over the past investments is done according to the FIFO principle in order to withdraw the required amount. For each past investment, the gain is calculated and the occurring taxes (including the *Vorabpauschale* in December) are deducted from the amount. Finally, the occurring broker costs and the inflation are calculated. The process is shown in Algorithm 2.

4 VISUALIZATION TECHNIQUE

The visualization features two charts, the cash flow and detail view, which are synced via a time-axis (x-axis) The switch from saving to payout phase is indicated by a vertical line. The end of the payout phase fades out to indicate inaccuracies in life expectancy. The actual data at a certain time is printed on the right side when hovering over that point in time. The hovering is synced across

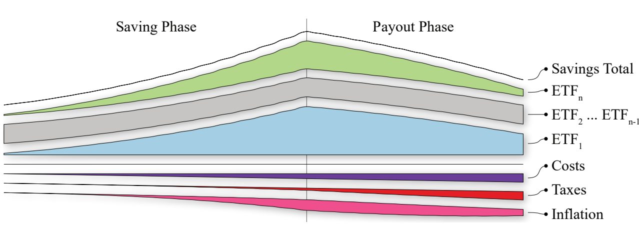

Figure 2: An explosion view depicting the composition of our depot view: The upper part consists of *n* ETFs with their values stacked on top of each other. An additional line on top indicates the total saving. The negative space consists of costs followed by taxes and inflation. The x-axis splits positive and negative space and serves as semantic separator.

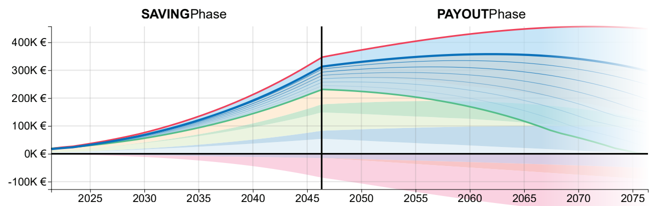

Figure 3: The detail view when the confidence display option is activated. The standard stacked areas are more transparent and the confidence area is placed over it. The area displays the possible range of the total depot value. The area borders are enhanced by a green (maximal confidence) and red (minimal confidence) line. The used confidence interval is 0%-100%. The main blue line displays the confidence of the current configuration (30%). Other lines are added artificially, displaying confidences in 10% steps upwards.

both charts and is displayed via a vertical dashed line. Next, the main user interface contains all parameterization options and is placed on the left. The whole prototype is depicted in Figure 4.

Cash Flow Overview. The cash flow chart visualizes the actual account movements using a bar chart. The red bars indicate the investment, whereas the green bars represent the payout. The payout already includes all fees and taxes. Each bar aggregates a certain configurable time period, e.g., a month or a quarterly period.

Detail View. The detail view displays the depot value development and all associated negative effects: costs, taxes, and inflation. A stacked area chart (Figure 2) with a negative and positive part is used. The positive part displays the value development of the ETFs. Thereby, each ETF consists of two stacks. The lighter color shows the investment and the darker one is for the revenue. An additional line emphasizes the total value of the depot. The negative part consists of the accumulated costs (purple) and the accumulated taxes (red). Additionally, the inflation is added as a visual cue. On the bottom of this chart, all actual values are listed in text form. They display the difference in the respective value to the previous parameter configuration in respect to the hovered point in time. Moreover, a confidence area can be optionally added over the existing stacked areas. In this case, the borders of the confidence area are the total depot value that would incur when the minimum and maximum confidence would be used. The middle line displays the currently selected confidence and is identical to the total value line when the confidence area is not displayed (Figure 3).

Parameterization. Parameters are divided in the categories money, time, cost, and visualization. In the money options the starting capital, monthly investment, and payout can be set. Investment and payout amounts can be set to dynamically increase on a yearly basis. The cost section gives control to the costs used for the saving and payout phase. Furthermore, the time options specify the length of both phases by means of the current age, life expectation, and years left until retirement. The visualization options include ETF selection and confidence configuration used for forecasting.

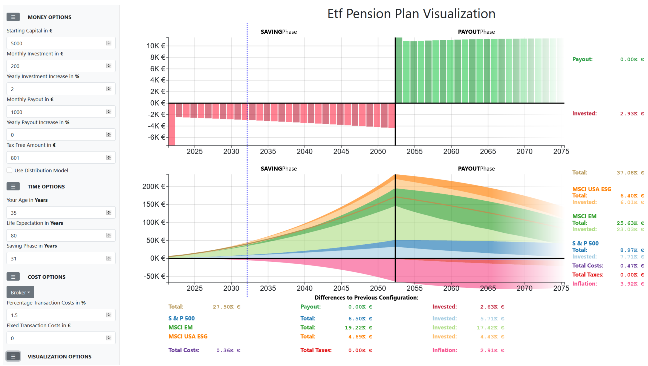

Figure 4: A complete overview of the existing prototype. A parameter pane divided into money, time, cost, and visualization options is on the left. The cash flow chart (top) displays the yearly investments and payouts. The depot chart (center) displays the forecast including costs, taxes, and inflation. Details w.r.t. to the year pointed at are displayed on right next to each chart. An agglomeration of all values is displayed in a tabular fashion (bottom).

5 USE CASE: FAMILY RETIREMENT SAVINGS

In this section, we (1) discuss the results of our prototypical simulation and visualization and (2) evaluate it by applying it to a certain use case and show the usefulness of the visualization.

Use Case: Family Retirement Savings. Let's imagine a young family with one child. We assume a starting capital of € 1000, a monthly investment of € 100 and payout of € 1000. In order to compensate the inflation both, the investment and payout increases by 1% per year. Additionally, the tax-free amount for a married couple is currently € 1602 and the accumulating version of the ETFs shall be used. Moreover, the couple is around 25 years old, will retire in 43 years and has a life expectation of 80 years. Their investment is equally spread across two standard ETFs the S & P 500 and MSCI EM. Furthermore, the family is assumed to already have an account at the bank ING which has relative high investment costs with € 4.99 fixed costs and 0.25% percentage costs. The confidence is set to 100% to minimize expectation. The simulation predicts a depot value of € 161000 with overall costs of € 5200, € 0 taxes, and € 60000 effectively lost to inflation at the end of the saving phase. Their depot lasts until 2084 which is one year earlier before the expected lifetime. Until that date, € 8000 costs and € 7000 taxes will be paid. To ensure that the savings last until the expected lifetime, a cheaper broker with only 1% percentage costs can be chosen to minimize the costs. This will result in € 3150 costs, € 8120 taxes and a leftover depot amount of € 15700 at the expected lifetime (still not enough though). In this case, the visualization enables long-term planning and shows the impact of individual factors. At best, it could lead to better future planning and decision-making. For example, changing the bank and increase monthly investment dynamically to minimize costs and allow for a buffer respectively.

Discussion. The simulation has several advantages and a few disadvantages. First, it relies on a few recent years for forecasting. This is hard to come by since ETFs have become popular only in recent years. Next, the simulation is deliberately using a linear model instead of an exponential model to prevent overestimation. Moreover, the simulation is based on the current situation. Thus, it is based on the current taxation laws and broker costs. Both will probably change in the future, which would invalidate the simulation. The visualization is good at displaying the negative factors costs, taxes and inflation, and the ratio between those negative factors and the depot value. It covers both, the depot development and the cash flow simultaneously. A drawback might be the unconventional layout of the inflation, which has been mentioned by several individuals who got in touch with the visualization. The other negative values are already discounted from the depot value, except the inflation. We decided to add a little hint by providing a visualization option that discounts the inflation from the total line. In its presented state, the simulation and visualization are missing further explanations for the parameters. For now, the user is expected to have basic knowledge on ETFs and what factors are involved in a long-term savings plan or retirement strategy. However, a certain learning and exploration phase is expected. The user is expected to discover and understand the parameters and their effect on the simulation and visualization by means of iterative exploration of the problem and solution space. A similar approach has been done by [18]. There, a lot of different visualizations are offered, each, however, stands for itself and does not allow for direct visual transfer between one another. We think that the visualization is a meaningful resource for individuals who intend to start ETF saving plans and want to understand the impact of a certain strategies and figure out the specifics for their plan. Furthermore, the visualization could be applied to other use cases, e.g., the passive income for one's children, which would be handed over to them when they are grown up.

6 CONCLUSIONS

We implemented a prototypical simulation and visualization for an ETF retirement plan strategy, which is based on the German cost and taxation model. The prototype (etf-vis.net) is build using Typescript, React, Bootstrap, and D3, and source code (github.com/varg-dev/etf-vis) are publicly available. It and can be used for different use cases next to the retirement planning such as planning a passive income for one's child. It might be able to fill the gap of personal visualizations which can "support people with limited visualization literacy and analytics experience" [10] and, following, the supports understanding and actions to enact on one's own retirement plans [16]. As a result, we extend available literature towards what is needed to create effective, interactive visualization for the communication and literacy of upcoming financial products [13].

In order to evaluate the effectiveness of the visualization, a user study with different groups regarding age, economic stand and financial literacy could be done. The visualization could be extended to explain the underlying simulation. Moreover, the simulation could be extended with further components, such as a simulation of a financial crash. That way, a user could interactively explore the consequences on the retirement savings. We believe that the type of visualization presented is a useful addition to the tools already available and facilitates the planning and adjustment of long-term savings goals and strategies.

ACKNOWLEDGMENTS

This work is part of the "Software-DNA" project, which is funded by the European Regional Development Fund (ERDF or EFRE in German) and the State of Brandenburg (ILB).

REFERENCES

[1] Mahantesh C Angadi and Amogh P Kulkarni. 2015. Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R. *International Journal of Advanced Research in Computer Science* 6, 6 (2015).

[2] Adnan Balloch, Anamaria Nicolae, and Dennis Philip. 2014. Stock Market Literacy, Trust, and Participation. *Review of Finance* 19, 5 (2014), 1925–1963. https://doi.org/10.1093/rof/rfu040

[3] BlackRock. [n.d.]. *Der ETF-Sparplanrechner | iShares.* https://www.ishares.com/de/privatanleger/de/anlegen/sparen-mit-etfs/etf-sparplanrechner

[4] Axel H Börsch-Supan, Michela Coppola, and Anette Reil-Held. 2012. *Riester pensions in Germany: design, dynamics, targetting success and crowding-in.* Technical Report. National Bureau of Economic Research.

[5] Jiajia Cheng, Hu nan Deng, Guang Sun, Peng Guo, and J. Zhang. 2020. Application of ARIMA Model in Financial Time Series in Stocks. In *ICAIS*.

[6] Johann U. de Villiers and Elze-Mari Roux. 2019. Reframing the Retirement Saving Challenge: Getting to a Sustainable Lifestyle Level. *Journal of Financial Counseling and Planning* 30, 2 (2019), 277–288. https://doi.org/10.1891/1052-3073.30.2.277

[7] extraETF. [n.d.]. *ETF-Sparplanrechner | ETF-Sparpläne einfach berechnen.* https://de.extraetf.com/finanzrechner/etf-sparplanrechner

[8] Aaron Fry, Jennifer Wilson, and Carol Overby. 2013. Teaching the design of narrative visualization for financial literacy. *Art, Design & Communication in Higher Education* 12, 2 (2013), 159–177. https://doi.org/10.1386/adch.12.2.159_1

[9] Kirti Goyal and Satish Kumar. 2021. Financial literacy: A systematic review and bibliometric analysis. *International Journal of Consumer Studies* 45, 1 (2021), 80–105. https://doi.org/10.1111/ijcs.12605

[10] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2015. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (2015), 420–433. https://doi.org/10.1109/TVCG.2014.2359887

[11] Gerd Kommer. 2018. *Souverän investieren mit Indexfonds und ETFs: Wie Privatanleger das Spiel gegen die Finanzbranche gewinnen.* Campus Verlag.

[12] Markus Leippold, Lujing Su, and Alexandre Ziegler. 2016. How index futures and ETFs affect stock return correlations. *SSRN* (2016). https://doi.org/10.2139/ssrn.2620955

[13] Annamaria Lusardi, Anya Samek, Arie Kapteyn, Lewis Glinert, Angela Hung, and Aileen Heinberg. 2017. Visual tools and narratives: new ways to improve financial literacy. *Journal of Pension Economics and Finance* 16, 3 (2017), 297—-323. https://doi.org/10.1017/S1474747215000323

[14] A. Mohamedashik, K. Senthamaraikannan, and D. Balamurugan. 2019. Statistical Techniques For Stock Price Prediction In India.

[15] Lucas Nunno. 2014. Stock Market Price Prediction Using Linear and Polynomial Regression Models.

[16] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. 2014. The Persuasive Power of Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2211–2220. https://doi.org/10.1109/TVCG.2014.2346419

[17] Stephen Rudolph, Anya Savikhin, and David S. Ebert. 2009. FinVis: Applied visual analytics for personal financial planning. In *2009 IEEE Symposium on Visual Analytics Science and Technology.* 195–202. https://doi.org/10.1109/VAST.2009.5333920

[18] Tyler, Portfolio Charts. [n.d.]. *Methodology – Portfolio Charts.* https://portfoliocharts.com/methodology/

[19] Norman Hans-Peter Ubber. [n.d.]. Exchange Traded Fund Savings Plans as Old-age Provision Tool in Germany. In *Brno, November 28, 2019 European Scientifc Conference of Doctoral Students.* 149.

[20] Martin Wattenberg. 1999. Visualizing the Stock Market. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania) *(CHI EA '99).* Association for Computing Machinery, New York, NY, USA, 188–189. https://doi.org/10.1145/632716.632834

[21] Sonja Wendel, Meimei Dai, Benedict Dellaert, and Bas Donkers. 2016. Consumer retirement planning over the life cycle. (2016).

[22] Josh Wiener and Tabitha Doescher. 2008. A Framework for Promoting Retirement Savings. *Journal of Consumer Affairs* 42, 2 (2008), 137–164. https://doi.org/10.1111/j.1745-6606.2008.00102.x

[23] Edward N Wolff. 2011. *The transformation of the American pension system: Was it beneficial for workers?* WE Upjohn Institute.