Unveiling the Winning Contribution Patterns for Enhanced Financial Health

Chuck Grace,[1] Adam Metzler,[2] Yang Miao,[3] Longlong Feng,[4] and Alireza Fazeli[5]

Abstract

By middle age many investors have accumulated nest eggs of comparable size to their annual salary. In this case an additional percentage point of returns has the same mathematical impact on their wealth as a percentage point of savings rate. However, while savings rate falls under the direct control of the investors, investment returns only very weakly so. What is the experience of actual investing Canadians in facing this circumstance? Studying this topic is important because, despite extensive attention to the topic, savings rates in Canada have been slowly eroding for over 20-years. We examine how accumulation patterns impact investor outcomes by examining their investment transactions, over a three-year period ending in August 2022, using advanced data analytics in the form of machine learning to explore previously unknown patterns. This paper gives the resounding answer that investors are overwhelmingly likely to be better served by a focus on savings rather than on returns. We conclude that a consistent pattern of saving is a 'winning' strategy for wealth accumulation. Saving patterns were by far the most powerful determinant of lifetime utility. The simple act of opening an account and automated regular contributions is the most powerful technique that investors, policy makers, asset managers and advisors can deploy in the pursuit of wealth accumulation. Despite this conclusion we observed that most of the investors in this study did not appear to follow a strong saving strategy.

Recommended Citation

Grace, C., Metzler, A., Miao, Y., Feng, L., & Fazeli, A. (2024). Unveiling the winning contribution patterns for enhanced financial health. *Financial Services Review*, *32*(2), 1-28.

Introduction

By middle age many investors have accumulated nest eggs of comparable size to their annual salary, and many turn their attention to maximizing returns. At this point in their lives, an additional percentage point of returns has the same mathematical impact on their wealth as a percentage point of savings rate. What is the experience of actual investing Canadians in facing this circumstance? While savings rates fall under the direct control of the investors, investment returns remain at the mercy of unpredictable markets. This paper gives the resounding answer that investors are

[1] Corresponding author (fwl@uwo.ca). Western University, London, Canada
[2] Wilfrid Laurier University, Waterloo, Canada
[3] Western University, London, Canada
[4] Western University, London, Canada
[5] Western University, London, Canada

overwhelmingly likely to be better served by a focus on savings than on returns.

The math behind savings and its economic impact have been well studied. The topic has also been explored in depth by well-known behaviouralists such as Kahneman and Thaler (2006). It has been extensively mentioned as a contributing factor in the context of financial wellness. And enormous resources have been dedicated to growing savings through investment strategies. But there is little empirical study of real-world savings transactions and how those transaction patterns impact wealth accumulation – especially in a Canadian context.

We examine unique dataset(s) of investor transactions to examine the relationship between investor behaviours, investment strategies, household savings, and investment outcomes. Ultimately our goal was to determine whether investment returns or savings rates drove wealth accumulation for the investors in our datasets.

We examine these real-world observed behaviours through advanced data analytics in the form of unsupervised machine learning. We examine trading over a 3-year period ending August 2022, providing us with the opportunity to observe patterns during rising markets, declining markets and the turbulent phases during transitions. The data encompasses hundreds of unique investment strategies including advised and 'do-it-yourself' portfolios.

The algorithms determined that investors could be clustered into one of three groups (for each dataset) that were determined by saving's behavior, investment returns and portfolio outcomes. A brief description of the resulting clusters is noted below in Table 1.

Table 1. Summary of the clusters derived by the unsupervised machine learning algorithms.

Dataset 1	Dataset 2
Cluster 1: A group of investors (13.8% of accounts) whose wealth followed a downward trajectory starting with the market correction in February 2020, and continuing the same trajectory thereafter.	Cluster 4: A group of investors (5%) whose wealth followed a downward trajectory following major withdrawals from their accounts
Cluster 2: A group of investors (44.1%) whose wealth generally followed the markets – rising and falling in sync with general market trends.	Cluster 5: A group of investors (34% of the dataset) whose wealth generally followed the markets – rising and falling with the markets.
Cluster 3: A group of investors (42.1% of the dataset) whose wealth had an upward trajectory (growing) throughout the period.	Cluster 6: A group of investors (61%) whose wealth had an upward trajectory (growing) throughout the period.

It should be noted that the observed period encompassed much of the Covid-19 pandemic. In Canada, the federal government introduced the Canada Emergency Response Benefit (CERB) in March 2020 before transitioning to the Employment Insurance program and ending in May 2023. The CERB program provided financial support to employed and self-employed Canadians directly affected by COVID-19. The $500/week payments, in addition to the savings derived from working-at-home, coincided with a significant increase in household savings (see Figure 4). As a result, our study provided a unique opportunity to observe the impact of unusual cash inflows on household resiliency – as measured by wealth.

It is also worth noting that the examined period encompasses unusual market conditions marked by negative returns on fixed income investments and mixed returns on equities (see Appendix 1).

We conclude that a consistent pattern of saving – even in turbulent markets - is a 'winning' strategy for wealth accumulation. Saving patterns were by far the most powerful determinant of wealth accumulation. Investment performance played a role in wealth accumulation, but it was muted when compared to savings behaviour.

We note that systematic saving on a regular or automated schedule enhanced the outcomes. As did saving more frequently – for example, bi-weekly as opposed to quarterly or irregularly. The observed results were consistent with stochastic simulations implying that ‘the math’ works. We note that ‘keep it simple’ by automating savings can be an effective strategy for wealth accumulation that cuts through the noise and confusion to create a tangible impact on financial wellness.

Across all the clusters, demographics such as age, gender, geography, risk tolerance and income were statistically immaterial in predicting outcomes with respect to wealth accumulation.

We observe that advised investors had higher savings rates and lower withdrawal rates than the DIY investors although the size of the DIY dataset is significantly smaller and the investors significantly younger.

We observed that 56% of the observed investors in this study did not appear to optimize their savings behaviour over the period studied as compared to the active savers.

Despite the extensive study by multiple disciplines, savings rates in Canada are not improving. Canadian savings rates have been slowly eroding for over 20 years. They trended up during the 2019 pandemic but have subsequently reverted to pre-pandemic levels - levels described by some policy makers as dangerously low. Canadian savings rates are currently “middle of the pack” among G20 countries and forecasted to be the lowest amongst the G20 by 2025. In 2023, Canadian households are preoccupied with inflation and the impact of rising interest rates, putting pressure on household budgets and the potential for savings rates to decline further. On average, spending is outstripping incomes, household debt levels are rising, and a significant percentage of Canadians are worried about having sufficient retirement savings.

Perhaps the trends noted above could be reversed if more Canadians could be encouraged to follow simple, automated savings plans?

Literature Review and Background Context

The topic of ‘saving’ has been explored by several disciplines. Over the years, the topic has proven to be important to policy makers, economists, portfolio managers, actuaries, and behavioural scientists. The topic is also important to the financial services industry who look to ‘household saving’ to fuel a plethora of investment products and services. More recently it has also been viewed as intrinsic to the definition of financial wellness.

Quantitative Modelling: Savings or Returns

Both empirical evidence and the quantitative modelling sketched below, and expanded in Appendix 2, suggest that by middle age investors are likely to accumulate a large enough capital base that their investment decisions are as impacted by return considerations as by considerations of savings rate. It is at this stage, when investors are in their early 40s, that many many people decide to move their assets to a more full-service investment manager, like investors in our dataset. However, savings rates fall under the direct control of these investors while investment returns only weakly so.

According to a standard and simple discrete time model in which an investor begins at $t = 0$ with assets V_0, invests a constant fraction f of a constant income X each time period, and invests at a constant return rate μ reinvesting all proceeds, it can be shown that to close approximation the growth in a portfolio over a small k(= 2 or 3) years is given by: $V_{n+k} – V_n = fkX + \mu kV_n + ½ \mu^2k(k-1)V_n + ½ f\mu k(k-1)X$

The impact of μ will be small compared to the impact of f until such time as the portfolio grows to be about twice the income of the investor. This occurs at about the time $\ln(2)/\mu$ which is 14 for $\mu = 5\%$.

However, this mathematical sensitivity analysis does not consider the fact that the savings rates are more completely under the control of an investor than the return rate. The reality is that it is hard to move μ very much, and (in contrast to our simple model here), μ is random.

The goal of this paper is to see if there are natural groupings of investors – some of whom appear to be working on enhancing μ through market timing, and others concentrating more on savings, and to compare them over a real time-period to see which group can generate more retirement savings.

Life Cycle Hypothesis

Our research touches on the theoretical Life Cycle Hypothesis framework proposed by Modigliani et al. in 1954. The scholars proposed that an optimizing behaviour implies a smooth consumption and lifetime utility where individuals accumulate wealth through savings during their working years. The hypothesis posits that individuals will transition to retirement at a lifetime peak in income and wealth. Our research contributes to the research in this area by examining actual individual portfolios as they approach what should be the peak in their consumption curve.

Financial Wellness and Savings

Recently, savings have been linked to the concept of financial wellness (Vlaev et al., 2014, Kempson et al., 2017, Suh 2021, Metzler 2021). In their ground-breaking research, Kempson et al. (2017) identified three key behaviours that define financial wellness – spending restraint, active saving and borrowing for daily expenses. In previous research, the authors of this paper (Metzler et al., 2021) concluded that savings, spending, and debt play uniquely powerful roles in financial resilience. The authors determined that of the 200+ variables used in the clustering, three of the top nine were related to savings.

A financially resilient individual can withstand financial setbacks such as sudden loss of income or unanticipated expenses. Low levels of financial resilience are a strong predictor of financial stress and can lead to more serious health problems[6]. The events of 2020 and the economic impact of COVID-19 gave increased urgency to the topic of financial resilience. The financial cost on all levels of government for financial countermeasures to COVID-19 are becoming apparent and a stronger understanding of the prevalence and type of financial fragility in Canadian households will allow more targeted policy interventions. There is clear value in helping financially stressed individuals understand the root of their financial challenges and then providing them with advice (tailored to their specific circumstances) on the steps they may be able to take to change their circumstances and alleviate their financial stress.

Savings Interventions

Numerous incentives have been explored by governments and industry to enhance savings. They include policy interventions.

Policy makers and government agencies have all explored household savings rates as a driver of 'healthy' economies and 'healthy' households (FCAC, 2021; Gale et al., 2005, Justera et al., 1999, Baldwin, 2022). Governments around the world regularly incentivize households to 'save more' – often with a focus on pensions and retirement. In Canada, retirement savings plans, tax free savings accounts and registered education savings plans are popular examples of government sponsored programs with assets under administration measured in the trillions of dollars[7]. Policy makers will also point to savings rates when exploring topics such as poverty and interventions for disadvantaged or vulnerable groups (Cruz et al., 2016, Hall, 2021).

Demographic Drivers

In our analysis we include specific demographic features (see Table 2 below) in order to observe wealth accumulation vis a vis potential demographic drivers. Researchers have linked savings rates and resiliency to several demographic factors including age (Maynard et al., 2008, Baldwin, 2022), income (Dynan et al., 2004, Turner & Luea, 2009, MacGee 2022, Cruz 2016), household composition (Cobb-Clark et al., 2016), and gender (Fisher, 2010). These factors are often combined under 'life cycle model' (Feiveson et al., 2019). However, Metzler et al. (2021) noted that while these demographic traits can be linked to financial resilience, the data does not support a causal relationship.

[6] Manulife, 2016 Financial Wellness Index

[7] Statistics Canada, www150.statcan.gc.ca/, Table 11-10-0016-01, released 2020-12-22, sourced June 2023

Table 2. Select Features Used in the Clustering Algorithms

	Description	Features (examples)
Demographic Features	General demographic information	Age, income, gender, marital status, residency
	Know your client information as prescribed by regulations	Investment knowledge, net worth, risk tolerance, investment horizon
Behavioural Features	Derived features that can be used as proxies for investor behaviour	Risk tolerance, automatic versus JIT trades (habit), portfolio churn, trading frequency
Financial and Transactional Features	Account detail: the account holding a portfolio on investments – for example RSPs or TFSA	Account type including RSP, TFSA, RESP etc.
	Portfolio detail: A basket of securities or holdings	Security ID, units, book value, market value
	Holdings: An individual security	Security ID, risk type, trading exchange
	Transactions: a transaction that changes the book or market value of the portfolio	Type of transaction, date, units, security, gross, net, currency
	Any fees or commissions derived from a transaction	Type of fee, date, units, dollar amount, currency
	Bookkeeping: the dealer's accounting or administrative view of the elements above	
Engineered Features/Ratios	Weekly market values, Min/Max Scaling, Trading sequences, Contributions/Withdrawals as % opening balance, Contributions/Withdrawals as % on income, Trades per account, Trades per month, Internal Rate of Return	

Behavioural Interventions

In our analysis we include behavioural features (Table 2) in order to observe wealth accumulation vis a vis potential behavioural drivers. Kahneman (2012) and Thaler et al. (2004) are widely known for theorizing that behavioural attributes drive savings success and that the concept of 'nudging' can be used to influence savings decisions. Goal setting (Soman et al., 2011), mental accounting (Shefrin et al., 2004), future self (Hershfield et al., 2011, Cheema et al., 2011), and risk aversion (Cagetti, 2003) have all been linked to savings behaviour. Dholakia (2016), in turn, noted that it may be more useful to focus on habits or traits, rather than behaviour, when attempting to predict sustainable saving activities. Further research (MacInnis et al., 2009, Hall, 2021) has noted that interventions meant to nudge decision-makers should be tailored to individual differences and the social forces that impact particular social groups. Newmeyer (2020) notes that the benefits of automated savings accrue at a higher rate for individuals with lower incomes and that this benefit depends on the presence of a personal savings orientation (Dholakia et al., 2016).

The Impact of Advice

In this paper, we examine both advised and 'do-it-yourself' investors to observe wealth accumulation vis a vis the influence of advice. The role of a financial advisor with respect to household consumption and utility has not been widely researched but there is emerging research that advisors enhance savings behaviour. Industry studies (Russell[8], Vanguard[9]) estimate advisors add 150 to 200 basis points (bps) to portfolio

[8] Russell Investments (www.russellinvestments.com), sourced June 29, 2023

[9] Vanguard (www.vanguard.com/pdf/ISGQVAA.pdf), sourced June 29, 2023

growth through coaching and investor discipline. Foerster et al. (2017) and Linnainmaa (2020) measured advisor value and identified a significantly positive relationship when adding an automatic savings plan and that non-advised investors did not take advantage of automated savings plans. Researchers at CIRANO (Montmarquette et al., 2016) determined that investors who work with advisors benefited from greater savings.

Investment Strategies and Wealth Accumulation

In our analysis we include several financial features (Table 2) to observe wealth accumulation vis a vis potential investment strategy drivers. In finance and actuarial sciences, researchers have tended to focus on investment risk and return as the primary drivers for wealth accumulation. Established techniques such as diversification (Markowitz, 1991), asset pricing (Merton, 1973), lifetime ruin (Bayraktar, 2010), portfolio optimization (Markowitz, 2010) and target driven portfolios (Blake et al., 2013) are all focused on maximizing returns while minimizing risk – once a basket of savings has been accumulated.

Canadian investors generally follow one of two trading strategies with their Registered Retirement Savings Plans (the focus on this paper) – either active or passive. Active trading refers to the periodic trading in specific securities, typically to deliver alpha (unusual returns) or minimize risk (volatility). The antithesis of an active trading strategy would be a passive trading strategy where investors largely ‘buy and hold’ investments for the duration of their investment horizon. Do-it-yourself (DIY) investors can trade as often as they wish while advised investors are presumably influenced by their advisor’s recommendations and availability.

We have not included annual portfolio rebalancing under the definition of active trading as it represents a realignment to the investor’s risk tolerance rather than an attempt to ‘time the market’ or generate alpha.

The Bottom Line

Despite the interventions noted above, savings trajectories in Canada appear to be moving in the wrong direction. In Canada, savings rates trended up during the 2019 pandemic but have subsequently reverted to pre-pandemic levels - levels described by some policy makers as dangerously low (Baldwin, 2022, MacGee, 2022) (see Figure 1 and 2). As well, while incomes in Canada appear to peak in the 45 to 54 age group, savings rates peak in the 35 to 44 age group – well before the peak predicted in the Life Cycle Hypothesis (see Figure 3).

An empirical study could provide valuable insights into the impact of savings rates, as well as elucidate strategies individuals could adopt to improve savings.

Figure 1. Household Savings Rates as Percentage of Disposable Income[10]

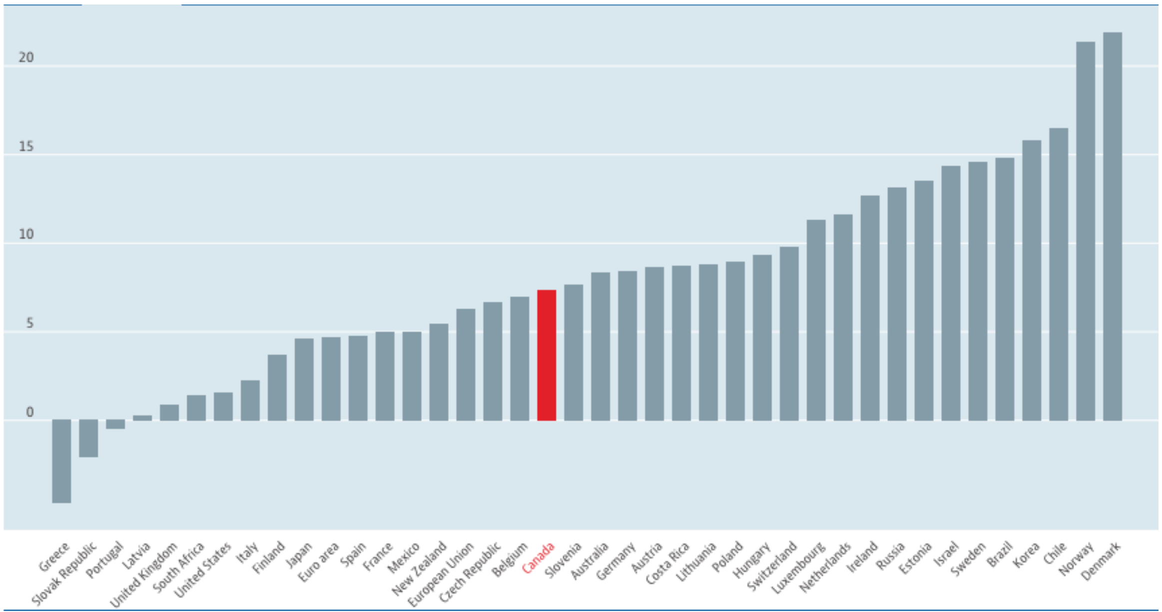
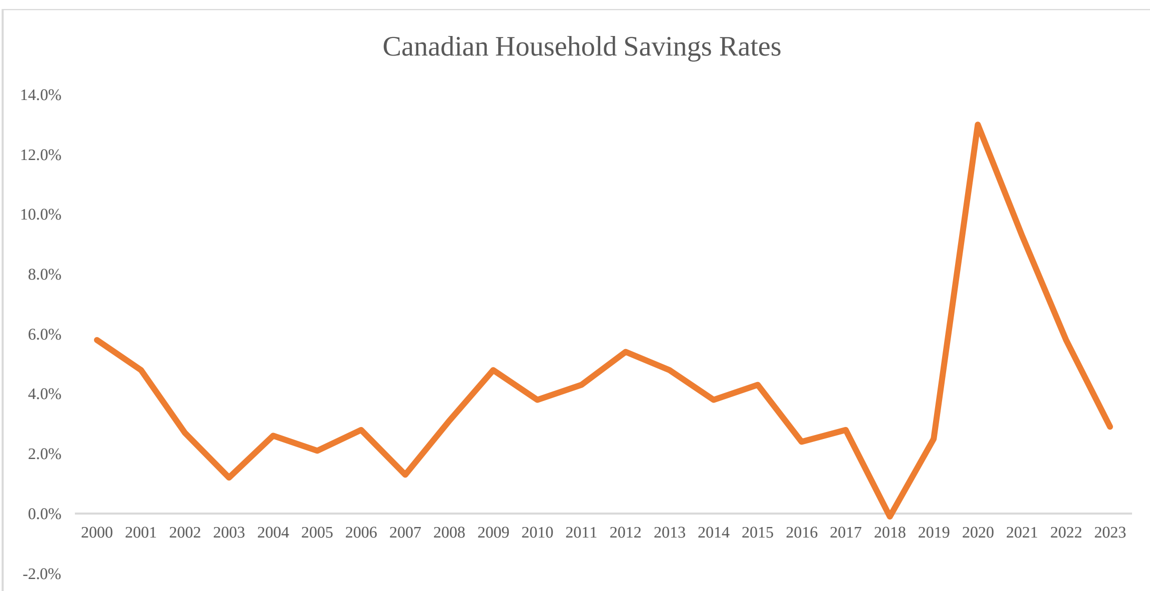
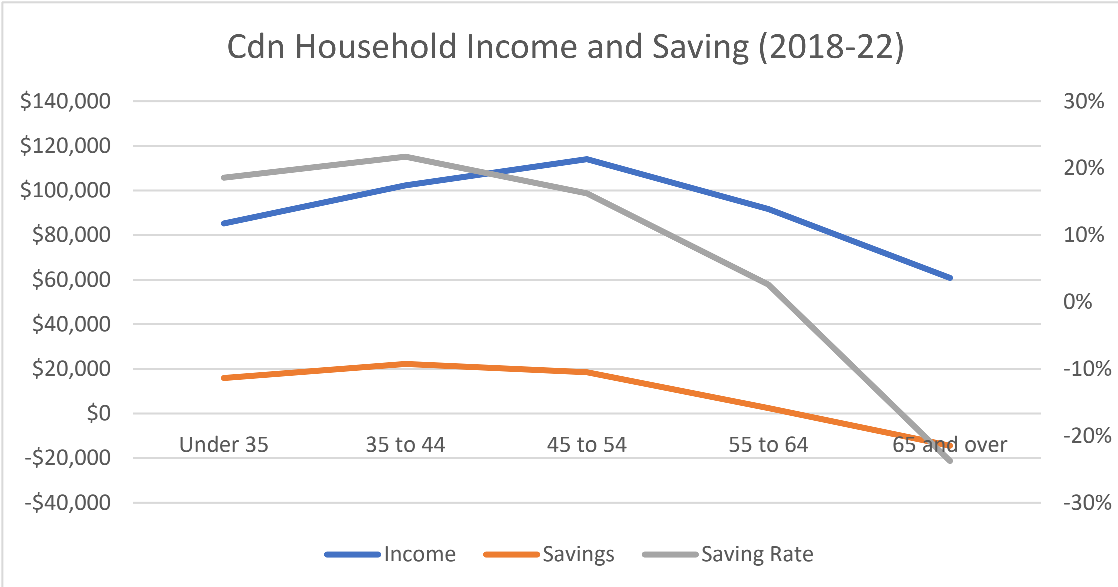


Figure 2. Canadian Household Savings Rates as Percentage of Disposable Income[11]



[10] OECD (2023), Saving rate (indicator). doi: 10.1787/ff2e64d4-en (Accessed on 29 June 2023)
[11] https://doi.org/10.25318/3610011201-eng, sourced June 29, 2023

Figure 3. Canadian Household Savings Rates as a Percentage of Disposable Income, by Age Group (2018/22)[12]



Data Description and Analytical Methodology

Data/Sample

The datasets include anonymized data on investors under two distinct circumstances.

- Dataset 1 encompasses 7,400 investors who work closely with financial advisors.
- Dataset 2 encompasses 477 investors who do not work with financial advisors – sometimes referred to as DIY investors.

We examined savings behaviours for the period August 2019 to August 2022, providing us with the opportunity to observe behaviour during rising markets, declining markets and the turbulent phases that transition the two.

Both datasets encompassed Registered Retirement Savings Plans (RSPs) to help control for risk tolerance, time horizon and structural constraints (see Appendix 3). The datasets included data points down to the daily transaction level.

Dataset 1 was provided by a registered investment dealer that has provided investment products to Canadian retail investors for over 30 years. The dealer hitherto has approximately 300 advisors who work with approximately 23,000 clients across Canada, with over $10 billion Canadian dollars (CAD) in assets under administration. Clients typically have multiple accounts each with different purposes. For example, a client may have accounts for: (i) retirement savings; (ii) children's education savings; and (iii) other savings. The data are comprised of 7,400 RSP accounts with associated Know Your Client (KYC) information, trade, and transaction details from 2 August 2019 to 5 August 2022 (see Table 2). The dealer provides a variety of financial products and services designed to support independent advisors. Furthermore, the dealer's focus is to provide positive outcomes to clients and advisors, and not to push certain financial products.

Dataset 1 investors work with a registered investment representative or financial advisor. More specifically, the advisors work under an investment dealer governed by the Investment Industry Regulatory Organization of Canada

[12] Statistics Canada Table: 36-10-0587-01 (formerly CANSIM 378-0152), sourced January 2024

(IIROC)[13]. Under the IIROC regime, advisors provide a broad range of services and can recommend investment solutions from thousands of investment choices[14]. Investment dealers are obligated to assess their client's risk tolerance when onboarding. The assessments generally take the form of questionnaire that gathers information on the client (Know Your Client or KYC) and scores the risk tolerance. An effective KYC protocol collects two types of information: (1) objective demographic data (e.g., identity), and (2) subjective information on the client's investment needs, financial objectives, investment knowledge, appetite for risk and other financial circumstances. In previous research, researchers (e.g., Thompson et al., 2021) noted that advisors are diligent at ensuring recommended portfolios match the client's stated risk tolerance. This determination allowed us to control for risk tolerance in our analysis.

Dataset 2 was provided by a registered investment dealer that has provided investment products to Canadian retail investors for over 9 years. The dealer operates under what is known as a "Robo Advisor" model where investors open and trade on an account online with minimal advice or service from the dealer. The dealer has approximately 12,000 clients across Canada, with over $.79 billion Canadian dollars (CAD) in assets under administration. Clients typically have multiple accounts each with different purposes. For example, a client may have accounts for: (i) retirement savings; (ii) children's education savings; and (iii) other savings. The data are comprised of 477 RSP accounts with associated KYC information, trade, and transaction details from 2 August 2019 to 5 August 2022. Dataset 2 represented younger investors with significantly smaller opening portfolio balances.

Both originating datasets were edited by the data donors prior to our receipt to ensure all client identifiers were anonymized consistent with Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) and standard research ethics protocols. Even after anonymization practices, there is the possibility that clients could be identified using machine learning algorithms (Rocher et al., 2019). Therefore, no individuals will be identified or referenced in this paper and any subset of the data cannot be shared with readers.

Machine Learning Algorithms in Finance

Machine learning algorithms have been widely used in financial applications. In this paper we are particularly interested in the use of clustering methods for financial trades and transactions. In our study, we deployed machine learning to uncover patterns that are otherwise difficult to discern given the complexity of the data. Our datasets encompassed over 200 discrete variables, some of which changed daily over the 36 months of observation. We deployed two machine learning techniques: Dynamic Time Warping and K-Means clustering using PyCharm, tslearn and dtaidistance in Python.

Our approach allowed us to systematically organize portfolios into clusters that demonstrate distinct investment behaviors. The figures below (Figures 4 and 5) represent the trajectories of three distinct client groups within each of Dataset 1 and Dataset 2. This visual representation helped validate our data-driven grouping and helps to highlight the unique trends within each cluster.

Further detailed explanations of our methodologies are included in Appendix 4.

[13] On January 1, 2023, IIROC merged with the Mutual Fund Dealers Association and the combined regulatory was renamed CIRO or the Canadian Investment Regulatory of Canada.

[14] Product choices for IIROC licensed representatives can include, for example, bonds, debentures, mortgage-backed securities, stocks, warrants, options, futures, mutual funds, exchange traded funds, labour sponsored funds, commodities, trusts, and hedge funds.

Figure 4. Dataset 1 Cluster Visual, Portfolio Values Scaled to 1.0, Aug 2019 to Aug 2022

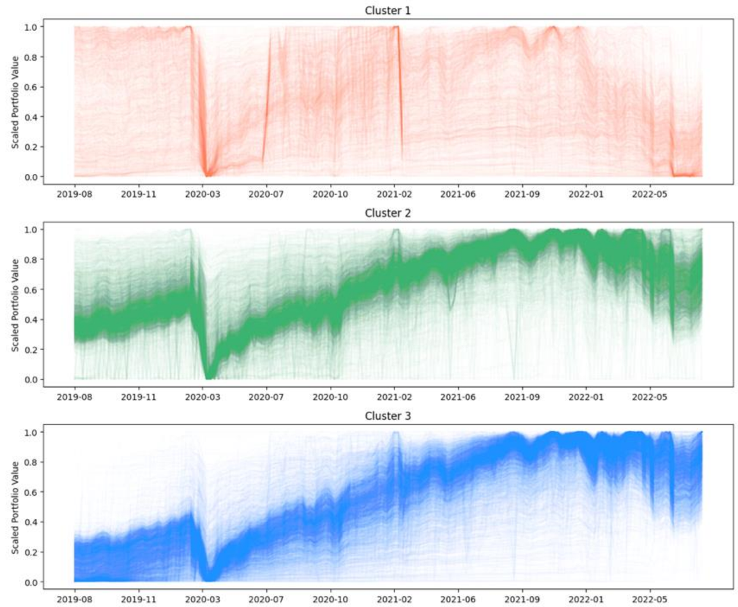
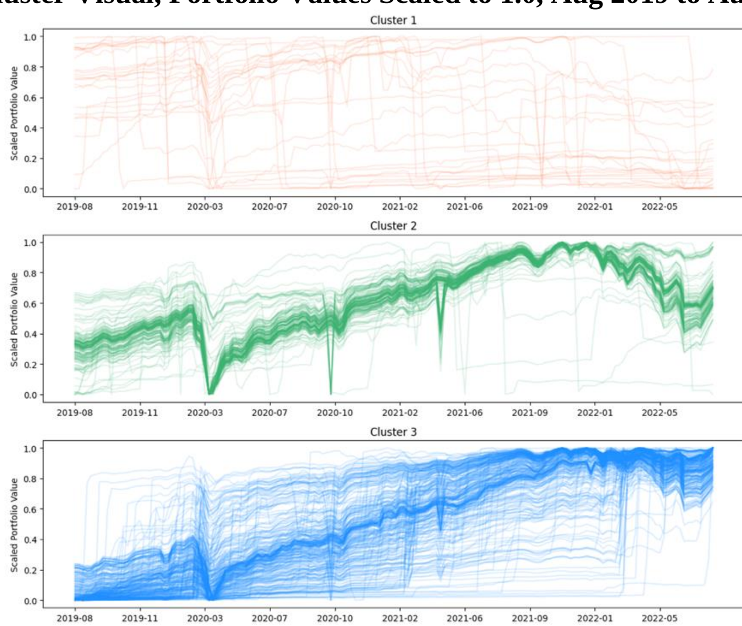


Figure 5. Dataset 2 Cluster Visual, Portfolio Values Scaled to 1.0, Aug 2019 to Aug 2022



Variables

Variables or features covered a broad range of data elements but can be summarized as demographic, behavioural, financial, and engineered. In our clustering, we focused on 74 discrete data elements and several derived ratios or engineered features as described in Appendix 4 and summarized in Table 2.

Results

Our clustering identified three unique groups per dataset (Table 1). Within each dataset, the clusters were very similar in terms of KYC data (age, income, gender, risk tolerance etc. but the clusters were differentiated by their saving behaviour (Table 4). In particular, the clustering was driven by four dominant features.

1. Net contributions,

2. Net contributions as a % of opening balance,
3. Net contributions as a % of income and
4. Contribution frequency.

Median net contribution rates (CR) ranged from a low of -59% to a high of 235%, compared to their opening balance.

Our analysis demonstrated that an active savings strategy was the most effective strategy for building wealth (utility) over the period examined. An active savings strategy was, on average, 5X more powerful at building wealth than relying on investment returns (see Table 3, Appendix 5 and Appendix 6).

Table 3. Portfolio Growth Attribution[15], Internal Rates of Returns (IRR) versus Contribution Rates (CR), by Cluster

	IRR (annualized)[16]	CR (annualized)	Overall Growth (Closing balance/Opening balance)
Dataset 1, Cluster 1	0.7%	-11.9%	-11.6%
Dataset 1, Cluster 2	2.3%	5.3%	14.5%
Dataset 1, Cluster 3	2.6%	28.9%	51.6%
Dataset 2, Cluster 4	1.1%	-121.2%	-50.5%
Dataset 2, Cluster 5	2.5%	3.3%	15.8%
Dataset 2, Cluster 6	1.7%	58.8%	158.7%

The six clusters experienced dramatically different outcomes with respect to wealth accumulation. In Dataset 1, Cluster 1 had (on average) a 12% decrease in wealth over the period while Cluster 2 had a 15% increase and Cluster 3 a 52% increase. In Dataset 2, Cluster 1 had (on average) a 51% decrease in wealth over the period while Cluster 2 had a 16% increase and Cluster 3 a 159% increase.

Clusters 3 and 6 demonstrated savings patterns that could be considered consistent with the Life Cycle Model predictions for this stage in their lives. i.e. CR was high and consistent with their peak income years. However, Clusters 1, 2, 4 and 5 exhibited patterns that were inconsistent with the Life Cycle Model. This observation is important because 56% of the investors observed in this study were from those four clusters. i.e. a majority of investors do not appear to be optimizing their savings rates during their peak income years perhaps contributing to the noted erosion in national savings rates.

It is worth noting that in Canada, withdrawals from an RSP before retirement are subject to onerous tax implications. The negative savings exhibited in Clusters 1 and 4 may therefore suggest that these investors were under unusual financial stress over this period.

We found little evidence to suggest active trading (portfolio churn) resulted in superior returns (neither for the advised or the DIY investors) or unusual growth in wealth.

Demographic features, risk tolerance, trading strategies, portfolio mix, and behaviour features

[15] ANOVA testing of the IRR and CR calculation yielded F-stat values of 109.6 and 393.709 respectively and p-values of 0.000 indicating significant differences in the average value of the clusters.

[16] By way of comparison, over the period examined, Canadian equity markets were up 5.7% and Canadian bond markets were down 4.2%. A balanced portfolio (60EQ/40FI) portfolio would have had a return of approximately 3.0%. (see Appendix 1)

had minor to insignificant roles in driving the clustering.

We noted that saving is a universal strategy - the results were the same regardless of age groups, genders, risk tolerances and income levels.

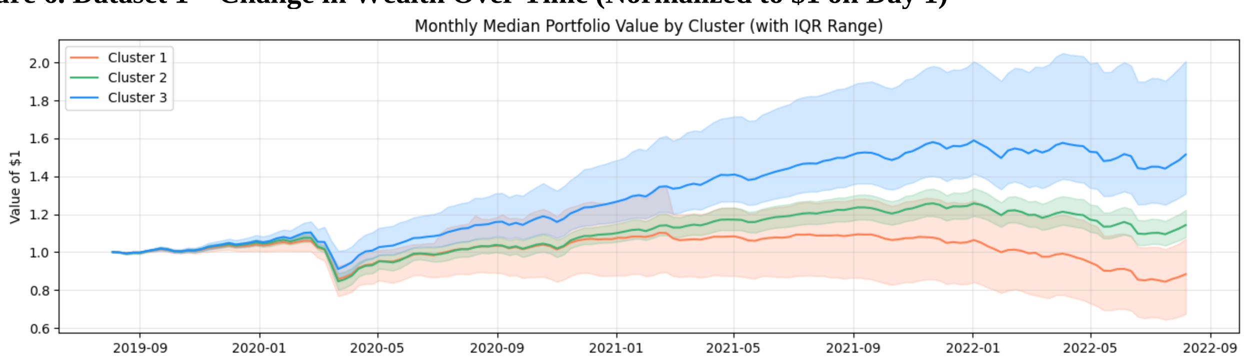
Table 4. Median Points for Select Features by Cluster

	Dataset 1			Dataset 2		
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
n (accounts)	1,020	3,266	3,114	24	164	289
Share of total accounts	14%	44%	42%	5%	34%	61%
Age	61	54	58	44	43	40
Income	$ 70,000	$ 75,000	$ 85,000	$ 89,000	$ 88,500	$ 80,000
Investment Knowledge (% Good)	54%	45%	48%	N/A	N/A	N/A
Gender (% male)	52%	54%	52%	N/A	N/A	N/A
Marital Status (% married)	75%	75%	78%	N/A	N/A	N/A
Employment Status (% working)	88%	98%	94%	N/A	N/A	N/A
Opening Portfolio Balance	$ 99,020	$ 96,913	$ 73,236	$ 27,361	$ 16,922	$ 6,875
Risk Tolerance (V@R, BPS)	235	229	241	N/A	N/A	N/A
Total trades per acct/per month	0.73	0.66	1.37	0.48	0.12	0.65
% systematic trades	56%	64%	58%	N/A	N/A	N/A
Ave. account contributions (total)	$ 11,406	$ 9,266	$ 29,167	$ 1,630	$ 2,380	$ 16,385
Ave. account withdrawals (total)	$ 13,397	$ 3,471	$ 4,509	$ 17,964	$ 941	$ 209
Ave. net contributions (total)	$ (1,991)	$ 5,795	$ 24,658	$(16,334)	$ 1,439	$ 16,176
Investment Returns (CAGR)	0.7%	2.3%	2.6%	1.1%	2.5%	1.7%
Net contributions as a % of income	-2.7%	7.2%	26.0%	-18.4%	1.6%	20.2%
Net contributions as a % of opening balance	-2.0%	6.0%	33.7%	-59.7%	8.5%	235.5%

All six clusters had investment returns that were consistent with their risk tolerance and the general market conditions at the time (see Appendix 1). Median IRRs ranged from a low of 0.7% to a high of 2.6%. By way of comparison, over the same period, medium term Canadian government bonds had a return of approximately -4.2%, Canadian equities 5.7% and U.S. equities 10.8%. A balanced portfolio (60% fixed income, 40% equities) would have had a return of approximately 3.0% before fees.

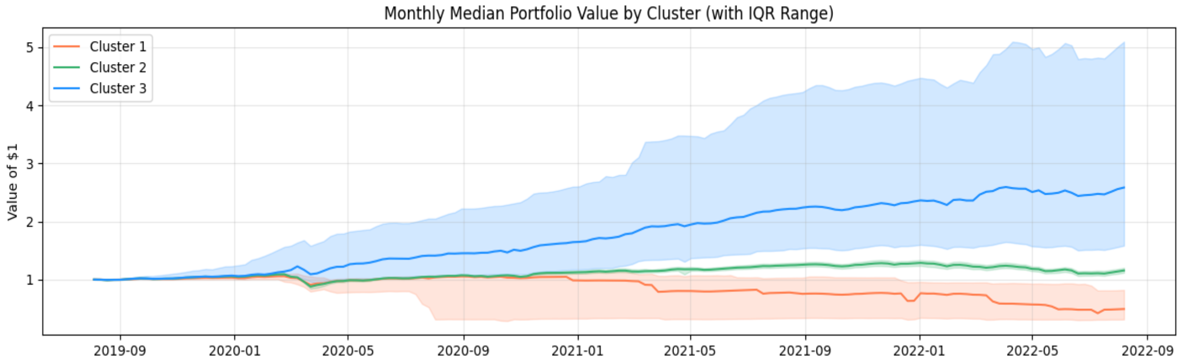
We found no evidence of investment returns driving significant wealth accumulation. Instead returns followed a normal distribution curve with random deviations from the mean (see Appendix 6 for a description of our investment return methodology).

Figure 6. Dataset 1—Change in Wealth Over Time (Normalized to $1 on Day 1)



Note. Canada Emergency Response Benefit (CERB) payments began in March 2020 and ended in May 2022.

Figure 7. Dataset 2—Change in Wealth Over Time (Normalized to $1 on Day 1)



In Figures 6 and 7 above, it should be noted that there is no overlap between Cluster 2 and 3 or Cluster 4, 5 and 6 – indicating statistically unique patterns.

We noted that savings frequency had a marginal impact on investment returns as measured by IRR but a significant impact on savings rates (CR) (Appendix 7). Investors who saved more frequently (biweekly vs quarterly for example) had a higher CR. And investors who saved systematically and regularly, had significantly higher savings rates than investors who saved periodically. Given the administrative burden and the consistency of our observed trading behaviour, we have assumed that weekly, biweekly, and monthly trades were automated. Our conclusions led to a discussion of the parsimony principle – to keep it simple. We concluded that when searching for wealth strategies with a powerful impact on financial resilience, keeping it simple – saving and saving often - is not only easy to prescribe but effective.

Finally, we observed that the advised investors (Dataset 1) had higher savings rates and lower withdrawal rates than the DIY investors (Dataset 2) although the size of Dataset 2 is significantly smaller and the investors significantly younger than Dataset 1.

Discussion

Limitations

Our conclusions are constrained by the datasets provided and the timeframe they cover. It is possible that additional data could influence the feature engineering deployed during our clustering. For example, we were not able to examine savings or trading behaviour in the context of fees or taxes.

We observed that the data was not 'perfect'. It included cases where the data was erroneous. It is not unusual with 'real world' data to encounter incorrect values or administrative challenges. These values would eventually be corrected over time, but our dataset was a point in time snapshot. We made efforts to curate the data and account for these outliers. Subsequent testing and modelling determined that our curation did not materially impact our final calculated values.

Likewise, our conclusions are constrained by the unique time-period they cover and its relatively short (3-year) duration. The time-period (2019 to 2022) represents a particularly unique period given the Pandemic. It is probable that investment returns would play a stronger role over a longer time-period. Over the last 25 years, Canadian fixed income yields have averaged closer to 3.6% annually compared to the -4.2% observed in our dataset. We would note however that historical investment returns would still pale in comparison to our strongest observed savings patterns.

Our conclusions are also limited to a specific view of an investors saving patterns – Registered Retirement Saving Plans. We did not have access to an investor's savings at other institutions such as an employer sponsored pension plan or a savings account at their bank. It is possible that some investors would seek to optimize their savings across multiple accounts and our observations will not reflect those tendencies.

Nor have we attempted to answer the question 'how much is enough'. It could be argued that some of our strongest observed savings' patterns would be difficult to maintain over the long run. But we have left that question for the 'further research' section.

Implications

Our analysis concluded that:

1. An active savings strategy was more effective at building wealth than relying on investment returns or complex trading strategies alone.
2. Saving is a simple, reliable, and powerful technique to build wealth.
3. Frequent and disciplined saving is more effective than irregular or just-in-time saving.
4. Saving is a universal strategy - the observed results were the same regardless of age groups, genders, risk tolerances and income levels.

Our conclusions offer participants a simple tool to cut through the complexity of the background noise and focus on actions that have a tangible impact on wealth accumulation and, by extension, financial resilience. The key would appear to be a focus on participation in savings plans, automated if possible, and incented to run as long as possible. There is no need to make this complicated. The simple act of opening an account and automating regular contributions is the most powerful technique that investors, policy makers, asset managers and advisors can deploy in the pursuit of wealth accumulation.

For policy makers, we would encourage their continued sponsorship of savings plans such as retirement savings plans, tax free savings accounts, education saving plans and homeownership saving plans. We would also encourage careful consideration for raising the annual contribution limits in Canada for RSPs and TFSAs, in particular.

Globally, policy makers have become strong advocates for financial literacy. We would encourage them to make saving a cornerstone strategy within their financial literacy plans. In Canada, the Financial Consumer Agency of Canada (FCAC) has embarked on a project to measure financial resilience. We would strongly advocate to integrate savings into such a measure.

In our research we determined that the impact of strong savings behaviour crossed demographic lines such as age, gender, income and location. We would therefore advocate for the sponsorship of saving across the widest breadth of society and not narrowly focused on any one segment.

For regulators, we would encourage a balanced perspective that combines transparency and a fiduciary perspective with incentives to save and processes that create simplicity for investors. Transparency is a noble objective but when taken too far, it can create complexity and noise for decision makers (investors). Complexity has been shown to be a barrier for the saving behaviours for which we advocate.

For asset managers, we would encourage more balanced market facing activities that spend more time encouraging saving in general and less time overwhelming investors with investment and economic jargon and communications that create confusion rather than a tangible impact on client outcomes. We encourage a specific focus on the use of systematic saving routines (Preauthorized Chequing or PACs) as a tangible, simple mechanism for capital accumulation.

For advisors we encourage a specific focus on systematic saving routines (Preauthorized Chequing or PACs). In addition to a strong impact on customer outcomes, automated savings routines can help streamline an advisor's operation and represent a low cost means to increase assets under administration. Advisors could also consider a goal-based approach that helps clients keep their investing activities in perspective – i.e., encouraging activities, such as saving, that will have the strongest impact on their end goals.

Employers are in a unique position to encourage systematic saving through payroll deduction. We would encourage employers to strongly support saving plans for such things as retirement but to also include plans for children's education and emergency accounts. Robust sponsorship and participation in these plans have been shown to improve employee wellness with downstream benefits to the employer in terms of reduced absence, higher productivity and stronger employee loyalty.

For consumers, the overwhelming conclusion from our research is to embrace saving and simplicity. A simple, automated savings plan into a diversified portfolio is the strongest way to achieve financial resilience. Factors such as fees, taxes, rebalancing, asset mix etc. can also be important for some investors, but it starts with saving and the accumulation of investable capital. And the first step need not be intimidating.

Further Research

Our research points towards 'what to do' but not necessarily 'how to do it'. We would support future research into how policymakers, regulators, industry, and advisors can ensure strong savings behaviours over the long run. In our datasets, we observed a 55% to 60% participation rate in systematic trades. We would ask 'what would it take' to move participation rates to 80% or 90%? Or to drive the behaviours observed in Clusters 3 and 6 from 44% of investors to 60% or 70%?

In Canada, industry sponsored research has explored the concept of Advisor Alpha – a measure of the value derived from advice. Our research hinted at higher savings rates in our advised dataset, but our DIY dataset was too small to be definitive. We would support future research, in collaboration with industry partners, into the role advisors play in encouraging strong savings behaviour.

Our research was specific to retirement accounts. Our conclusions would be strengthened by an examination of other forms of savings such as saving accounts or tax-free accounts. We plan to explore those areas next.

Our research deployed several methodological approaches to clustering financial data. While we are confident in the robustness of our analysis, further research would be beneficial in helping to identify best practices when researching with financial data.

Finally, our research has not addressed the question of 'how much is enough'. Is there a recommended minimum saving amount? Does that amount change depending on the goal or the time horizon? Is it better to save or eliminate debt? We have left all these questions for future papers and collaborations.

Disclosures

Acknowledgements: the authors wish to thank and acknowledge Matt Davison and Nicole Carleton for their expertise and advice during the final edits on the paper. The authors also thank Mark Reesor (Wilfrid Laurier University), Kristina Sendova (Western University) and the many members of our data donor team for their invaluable insights that improved the content and conclusions of this document.

Author Contributions: Conceptualization, C.G. and A.M; data curation, L.F., Y.M and A.F.; methodology, C.G, A.M., Y.M., AF; software, L.F.; validation, A.M., Y.M.; formal analysis, A.M., Y.M. L.F., A.F.; resources, C.G. and A.M.; writing - original draft preparation, C.G.; writing—review and editing, C.G, A.M., L.F. Y.M. and A.F.; visualization, A.M. and L.F.; supervision, C.G., and A.M.; project administration, C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported, in part, by funding from the Financial Wellness Lab at Western University, Natural Sciences and Engineering Research Council of Canada, and our anonymous industry partners.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Government of Canada's Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2) and approved by the Research Ethics Board of Western University (REB #118582, approved April 2022).

Informed Consent Statement: The data source is secondary and provided to us from the private data donor. Additionally, the data has been anonymized so individuals cannot be identified by their accounts. This was approved by the ethics board above.

Data Availability Statement: The data used in this paper contains personal information for a number of Canadians and cannot be shared due to a non-disclosure agreement with our private data donor.

Conflicts of Interest: The dataset discussed herein was designed and collected by a private industry financial dealership that, in part, funded this research. The funders had no role in the analyses or interpretation of the data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
AUA	Assets under administration
CAD	Canadian dollars
CR	Contribution Rate (savings or deposits)
DIY	Do It Yourself
DRIP	Dividend Reinvestment Program
ETF	Exchange Traded Fund
FINRA	Financial Industry Regulatory Authority
IIROC	Investment Industry Regulatory Organization of Canada
IRR	Internal Rate of Return
KYC	Know your client
KYP	Know your product
PAC	Preauthorized Contribution
RSP	Registered Savings Plan
RRSP	Registered Retirement Savings Plan
TFSA	Tax Free Savings Account

Appendices

Appendix 1: Industry Investment Return Benchmarks (August 2, 2019 to August 5, 2022)

			Asset Mix (Equity/FI)		
Asset Class	Investment Proxy	Returns	80/20	60/40	50/50
Fixed Income	iShares Core Canadian Universe Bond Index ETF (XBB.TO)	-4.2%	20%	40%	50%
Cdn Equity	iShares Core S&P/TSX Capped Composite Index ETF (XIC.TO)	5.7%	50%	35%	30%
US Equity	iShares Core S&P 500 Index ETF (XUS.TO)	10.8%	30%	25%	20%
Portfolio Returns			5.2%	3.0%	1.8%

Source: https://ca.finance.yahoo.com/quote/ Sourced July 11, 2023, returns annualized (CAGR)

Appendix 2: Quantitative Modelling

This paper considers the importance of contributions (as opposed to investment returns) during the accumulation phase of the retirement savings cycle. Specifically, we present theoretical and empirical evidence that maintaining a high contribution rate early in the lifetime of the retirement account is crucial to building wealth. This observation, in and of itself, is hardly a new insight. The primary contribution of the paper is not the observation, rather it is the use of unique transaction-level data from thousands of retirement savings accounts in Canada, to support both the conventional wisdom and theoretical model. Our hope is that this paper provides a reference point for both financial advisors and policy makers, when providing advice and designing incentive programs.

Theoretical Framework

Perhaps the simplest model for portfolio growth is to assume a deterministic per period investment growth rate of μ, a constant per period salary of X of which a constant fraction f is saved, and that all investment income is reinvested. The original wealth is V0. This model yields the difference equations:

Discrete
$V_{k+1} = (1+\mu)V_k + fX$
V_0 = given

Where V_k is the value of the investment portfolio at the kth time step.

This has a solution $V_k = V_0(1+\mu)^k + (fX/\mu)[(1+\mu)^k - 1]$

Expand for small mu. Using the binomial theorem $(1+\mu)^k = 1 + k\mu + k(k-1)/2\ \mu^2$ + higher order terms
And $[(1+\mu)k - 1]/\mu = k + k\mu(k-1)/2$ so…
To linear order $V_k = V_0 + \mu k V_0 + fkX$, or $V_k - V_0 = fkX + \mu k V_0$
To quadratic order $V_k - V_0 = fkX + \mu k V_0 + ½\ \mu^2 k(k-1)V_0 + ½\ f\mu k(k-1)X$

This will also be true for beginning at time n and moving forward k steps:

$V_{n+k} - V_n = fkX + \mu k V_n + ½\ \mu^2 k(k-1)V_n + ½\ f\mu k(k-1)X$
If k is about 3, mu 5% and f 10%, the 2nd order terms are going to be fairly neglible and so linear order is fine.

Note that when $fX = \mu V_n$ the first term, which is the savings term, and the second term, which is the growth term, are about equal in dollar value.

$fX = \mu V_n = \mu\{V_0(1+\mu)^k + (fX/\mu)[(1+\mu)^k - 1]\}$ which, if $V_0 = 0$ as is quite reasonable, occurs when

$fX = fX[(1+\mu)^k - 1]$ or $1 = (1+\mu)^k - 1$ or $(1+\mu)^k = 2$. This is when $k \ln(1+\mu) = \ln(2)$ or when
$k^* = \ln(2)/\ln(1+\mu)$. Using the linearization, good for small μ, that $\ln(1+\mu) = \mu$, this is approximately $k^* = \ln(2)/\mu$. For $\mu = 5\%$ this occurs when k is about 14 years.

Appendix 3: Retirement Savings Plans in Canada (RSPs)

Both of our datasets encompass registered retirements savings plans. A Registered Retirement Savings Plan (RRSP or RSP) is a savings plan, registered with the Canadian federal government. Investors who contribute funds to an RSP, gain a "tax-advantage" in that the contribution is exempt from income taxes in the year they make the contribution. Any investment income earned from investments held within the RSP also grows tax-deferred until it's withdrawn.

According to Statistics Canada [17] , in 2020, over 6.2 million Canadians made contributions to a registered retirement savings plan (totalling $50.1 billion). Twenty two percent of Canadian tax filers made RRSP contributions in 2020 with a median contribution of $3,600.

Canadians can open an RSP at their financial institution either by

- Working through a licensed investment representative (advisor)
- Opening a DIY account or
- Through their employer.

Saving in the context of RSPs generally takes the form of either periodic lump sum payments or automated deposits referred to by the industry as PACs (pre-authorized contributions). In Canada, lump sum payments are frequently made in late February each year, just before the RSP contribution deadline for the previous year. PACs are generally set on a monthly or quarterly frequency and are electronically withdrawn from the investors bank account. Deposits or savings into an RSP are traditionally referred to as ‘contributions.

Participants can also transfer funds from other RSPs they may own to consolidate their investments. For the purposes of this paper, we did not classify transfers as a saving activity since the savings behaviour was exhibited in a separate account prior to our research.

It could be argued that re-invested dividends (DRIPs) or the roll-over of interest payments are a form of saving but also represent a return on the original capital. For this reason, we include DRIPs and re-invested interest payments in both our Internal Rate of Return (IRR) calculations and our Contribution Rates (CR) calculations.

Appendix 4: Detailed Clustering Methodology

Machine learning algorithms have been widely used in financial applications, such as risk modelling, return forecasting, and portfolio construction (Emerson et al., 2019), quantitative finance (Rundo et al., 2019), financial distress prediction (Huang et al., 2019), banking risk management (Leo et al., 2019), credit-scoring models and financial crisis prediction (Lin et al., 2011), automation through artificial intelligence (Donepudi, 2019), market prediction (Henrique et al., 2019), and credit risk modeling, detection of credit card fraud and money laundering, and surveillance of conduct breaches at financial institutions (Van Liebergen, 2017). Popular algorithms used in these applications are support vector machines (Kim, 2003), neural networks (West et al., 2005), and random forests (Patel et al., 2015).

In this paper we are particularly interested in clustering methods for financial trades and transactions. Recent work in this area includes agglomerative hierarchical clustering for asset allocation (Raffinot, 2017) and aggregating stocks using dynamic time-series warping as a distance measure (Lim et al., 2020), self-organizing maps and k-means clustering methods in combination with classifier techniques to predict financial distress (Tsai, 2014), fuzzy C-medoids clustering method for classifying financial time series (D’Urso et al., 2013), and clustering algorithms for financial risk analysis using multiple criteria decision-making methods (Kou et al., 2014). Absent from this body of work is the use of this broad class of techniques to analyze trading behaviours, the focus of this paper.

In our study, we deployed machine learning to help uncover previously unknown patterns in the data. Machine learning – and in particular,

[17] Statistic Canada, https://www150.statcan.gc.ca/n1/daily-quotidien/220401/dq220401a-eng.htm

clustering – has proven to be invaluable when examining large, complex datasets. Our datasets encompassed over 200 discrete variables, some of which changed daily over the 36 months of observation. We deployed two machine learning techniques: Dynamic Time Warping and K-Means clustering using Python, PyCharm, tslearn and dtaidistance software.

We used Dynamic Time Warping (DTW) to quantify the degree of similarity among various portfolios' weekly average market values. We trained our DTW models using the RRSP portfolio's weekly average market values (the sole variable utilized in our time series analysis). We then conducted a deeper within-cluster analysis on KYC variables such as income and retirement indicators, however these variables weren't included in our model training. By capturing the temporal dynamics of these portfolios, we were able to identify patterns in clients' trading behaviors.

Afterward, we applied K-Means clustering algorithms to categorize portfolios exhibiting similar trajectories. This approach allowed us to systematically organize portfolios into clusters that demonstrate distinct investment behaviors. The figures below (Figures 3 and 4) represent the trajectories of three distinct client groups within each of Dataset 1 and Dataset 2. This visual representation helps validate our data-driven grouping and helps to highlight the unique trends within each cluster.

Classification of Investor Accounts by Contribution Frequency

Each of the datasets under study included transaction level detail at a daily level. To examine contribution patterns in terms of frequency, we first curated the data to eliminate transactions that did not affect portfolio market values. For example, administrative adjustments, corrective transactions, and some fees collected directly from the client. We then aggregated the remaining transactions at a daily level for each client to see whether the net sum of these activities was positive for a given day. If the total number of days with positive net sum was greater than 90% of the total number of business days in the interval, the account was classified as a daily contributor, otherwise, the aggregation was repeated, respectively, on weekly, biweekly, monthly, and quarterly levels to classify every account according to its contribution pattern. This bottom-up approach used the same threshold for pattern similarity (i.e., 90%) at every level of aggregation. If an account failed to be classified as one of the predefined periodic contributors, it was categorized as an irregular contributor. Results are included in Appendix 7.

Time series data clustering has become an important part of financial data analysis due to its ability to reveal hidden patterns and correlations in time-series data. By grouping similar time series together, it allows for a more efficient and targeted analysis, enabling analysts to draw conclusions about collective behaviour or attributes. Studies have corroborated the efficiency of using time series clustering for financial data analysis, highlighting its validity as an approach (Dose et al., 2005).

Min-Max Scaling

Before proceeding with the clustering process, it's essential to scale the time series data to ensure that the variance in scale of different features does not distort the distances between data points, which in turn would impact the performance of the clustering algorithm. Min-Max scaling is an effective method in this regard, as it brings all values within a predetermined range, typically between 0 and 1. This prevents features with larger scales from dominating the calculation of distances.

When applying Min-Max scaling to portfolio weekly market values, it's important to consider the structure of the input. In our case, each time series is associated with a unique account ID and scaling must be performed on an account-by-account basis. Suppose we have a time series associated with a particular account ID, $X = [x_1, x_2, \ldots, x_n]$. The Min-Max Scaler operation for each account ID can be expressed as follows:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where X_{min} and X_{max} are the minimum and maximum values of the time series X associated with that account ID. This scales the time series X_{scaled} such that all values lie between 0 and 1. This transformation ensures that we're comparing the shape of the time series, rather than being influenced by their magnitude when performing the subsequent clustering with DTW and K-means.

Dynamic Time Warping (DTW) Algorithm

The Dynamic Time Warping algorithm is a technique used to measure similarity between two sequences which may vary in time or speed. The algorithm considers all possible alignments between the sequences and identifies the optimal alignment that minimizes the total distance between them.

For two time series $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_m)$, which are represented as arrays of respective shapes (n, 1) and (m, 1), the steps involved in the DTW algorithm are as follows:

1. Initialization: Create an n-by-m matrix where the (i , j)-th element of the matrix contains the distance $d(x_i, y_j)$ between the points x_i and y_j. The distance can be computed using a selected distance metric, commonly the Euclidean distance. The calculation formula for Euclidean distance is: $d(x_i, y_j) = \sqrt{\sum (x_i - y_j)^2}$.

   Create a second n-by-m matrix D for storing the accumulated distances, where $D(i, j)$ represents the sum of $d(x_i, y_j)$ and the minimum among $D(i-1, j)$, $D(i, j-1)$, $D(i-1, j-1)$.

2. Matrix Filling: Iterate over the matrix D, starting from $D(1,1)$, and compute the accumulated distance for each cell using:

$$D(i, j) = d(x_i, y_j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)]$$

   According to this equation, the accumulated distance is the sum of the distance at that point and the minimum accumulated distance among its neighboring points.

3. Path Identification: Starting from $D(n, m)$, move backwards to $D(1,1)$ by choosing at each step the cell $(i-1, j)$, $(i, j-1)$, or $(i-1, j-1)$ that has the smallest accumulated distance. The path that is formed, known as the warping path, represents the optimal alignment between the two-time series.

The DTW distance between the two time-series is then given by the value at D(n,m), which represents the minimum sum of distances for aligning the two sequences. The whole process considers the temporal dynamics and can provide a more accurate measure of similarity between time series data compared to traditional Euclidean Distance, especially when dealing with sequences of different lengths or speeds. The flexibility of the DTW algorithm makes it particularly suited for financial time series analysis, where data can exhibit significant temporal variations.

K-Means Clustering

For our research, we used K-Means clustering, an iterative technique widely used in machine learning and data mining. The fundamental idea behind K-Means clustering is to classify dataset into K different clusters in such a manner that the within-cluster variations are minimized. The iterative process of the K-Means algorithm involves partitioning the portfolios into K clusters, computing the centroid of each cluster, and reassigning the portfolio to the cluster whose centroid is closest. The process continues until the positions of the centroids stabilize, indicating the optimal clustering of the data.

Since the nature of time-series data and the flexibility of DTW in aligning sequences, the centroid calculation can't be as straightforward as simply taking the arithmetic mean of the points in each cluster (Petitjean et al., 2011). We use a variant of K-Means known as Time Series K-Means that utilizes the DTW distance as the dissimilarity measure. In this context, the 'centroid' of a cluster is defined using the DTW Barycenter Averaging (DBA) method, which provides an averaged sequence that minimizes the distances to the sequences of the cluster. In each iteration, DBA performs three main steps:

1. Computing DTW alignments: In this step, we calculate the DTW between the temporary average sequence (also known as the centroid) and every individual sequence within our set of sequences, denoted as $S = \{S_1, \ldots, S_n\}$. This DTW computation allows us to establish links between the coordinates of the average sequence and the coordinates of the individual sequences.
2. Updating Centroid Coordinates: Each coordinate of the average sequence is updated as the barycenter (or geometric center) of coordinates linked to it in the previous step. The average sequence at iteration i is represented as $C = C_1, \ldots, C_T$, and we aim to update its coordinates for the next iteration (i+1), represented as $C' = C'_1, \ldots, C'_T$.

Now, we use a function 'assoc' that associates each coordinate of the average sequence with one or more coordinates of the sequences in S. This function is computed during the DTW calculation between C and each sequence in S.

3. We can then define the t-th coordinate of the average sequence C_t as:

   $$C_t = \text{barycenter}(\text{assoc}(C_t))$$

4. the barycenter is the arithmetic mean of a set of points {X1,...,Xα} in the vector space:

$$\text{barycenter}\{X_1, \ldots, X_n\} = \frac{(X_1 + \ldots + X_n)}{n}$$

After computing the new centroid, we then repeat the DTW computation between this updated average sequence and all sequences in S. The associations created by the DTW may change as a result, which is why we iteratively perform this process until the average sequence converges to a stable configuration.

Together, the combination of min-max scaling, DTW and K-means clustering forms an effective methodology for time series data clustering in our research.

Appendix 5: Investment Returns versus Contribution Rates

Figure 8. Dataset 1—Investment Returns (IRR) and Contribution Rates (CR) by Cluster

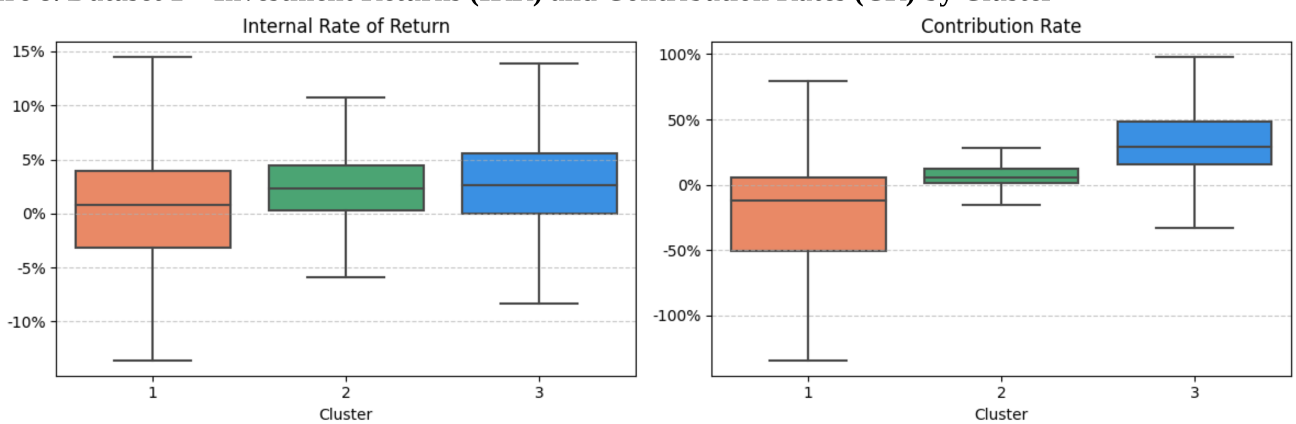
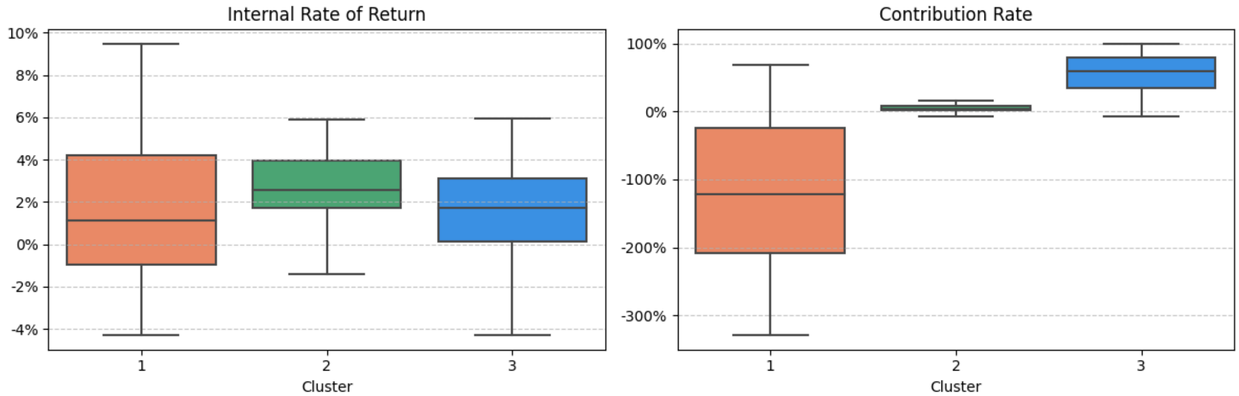


Figure 9. Dataset 2—Investment Returns (IRR) and Contribution Rates (CR) by Cluster



Appendix 6: Investment Return Calculations (IRR)

To compare investment performance across different portfolios, we needed a comparison criterion that considers the cash flows into and out of the portfolio, as well as the timing of such cash flows. One simple and commonly used criterion is the internal rate of return (IRR). The IRR is defined as the discount rate at which the present value of all cash flows in a given period of time equals to 0. Consider a portfolio that is invested from time 0 to time T. Assume that the market value of the portfolio at time 0 is S_0, and that the market value at the conclusion of the investment is S_T. Assume that the investor makes N transactions before time T, where the ith transaction happens at time t_i and has amount C_i. We further assume that $C_i > 0$ if the ith transaction is an additional contribution to the portfolio, and that $C_i < 0$ otherwise. The IRR of this investment is the root to the equation

$$S_0 + \sum_{i=1}^{N} C_i \, e^{-Rt_i} - S_T e^{-RT} = 0,$$

where $\sum_{i=1}^{N} C_i \, e^{-t_i R} = 0$ if $N = 0$. Notice more than one root may exist if one or more C_i is negative, i.e., the investor withdraws at least once from the investment. Some approaches have been proposed in the existing literature to select the most useful IRR in this case, see, for example, Hartman and Schafrick (2004). For the RRSP accounts that are analyzed in this project, there is a strong incentive for investors to refrain from withdrawing prematurely. Consequently, large withdrawals from the portfolio are less commonly observed compared to other account types.

To calculate the IRR of different portfolios, we first need to obtain the amount and the timing of all the cash flows. However, as the trading records of the raw data sets contain errors that are challenging, if not impossible, to distinguish from correct records, we need to resort to approximations. To this end, we use the following procedure:

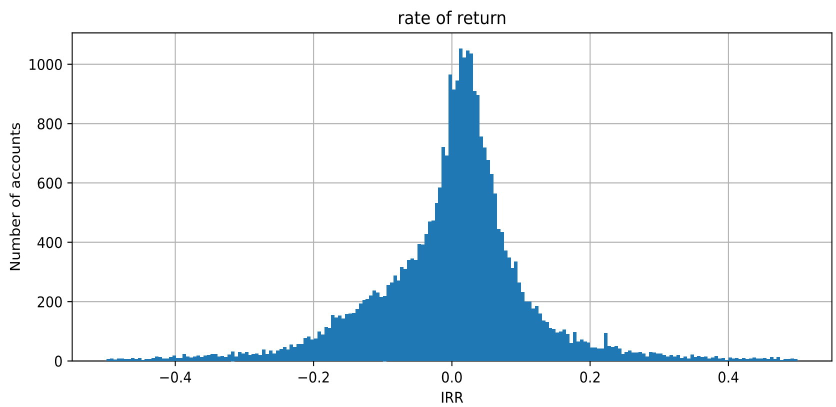
1. Calculate the daily change of number of shares for all the securities in a portfolio. Determine the reason for such changes and keep only those caused by a trading decision from the investor. For example, reinvested dividends would cause the number of shares to change, but they are not counted as cash flows.
2. Calculate the average trading price for each security on each day. Calculate the amount of the changes in step 1).
3. Add other cash flows that does not cause changes in the number of shares. For example, dividends paid out as cash do not cause the number of shares to change, but they are counted as cash flows since they are returns from the investment.

There are two main sources of error in the approximation procedure: the rounding error of the number of shares that are exchanged, and the difference between average trading price and actual trading price. The errors are not material and do not affect the results significantly.

A histogram of the observed IRRs is given in Figure 10. The table below summarizes the quantiles of the IRR.

Figure 10. Histogram of the Realized IRR

Quantile	0.01	0.25	0.50	0.75	0.99
IRR	-0.433257	-0.049519	0.011676	0.054847	.653425



Appendix 7: Savings Frequency and Wealth Accumulation Figures

Figure 11. Dataset 1—Savings frequency and Wealth Accumulation

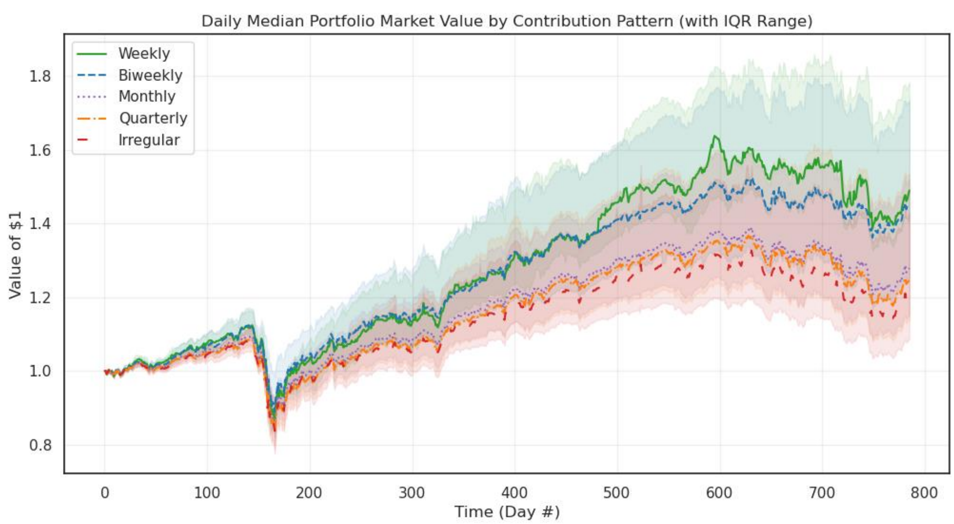


Figure 12. Dataset 1—Savings frequency, IRR versus CR

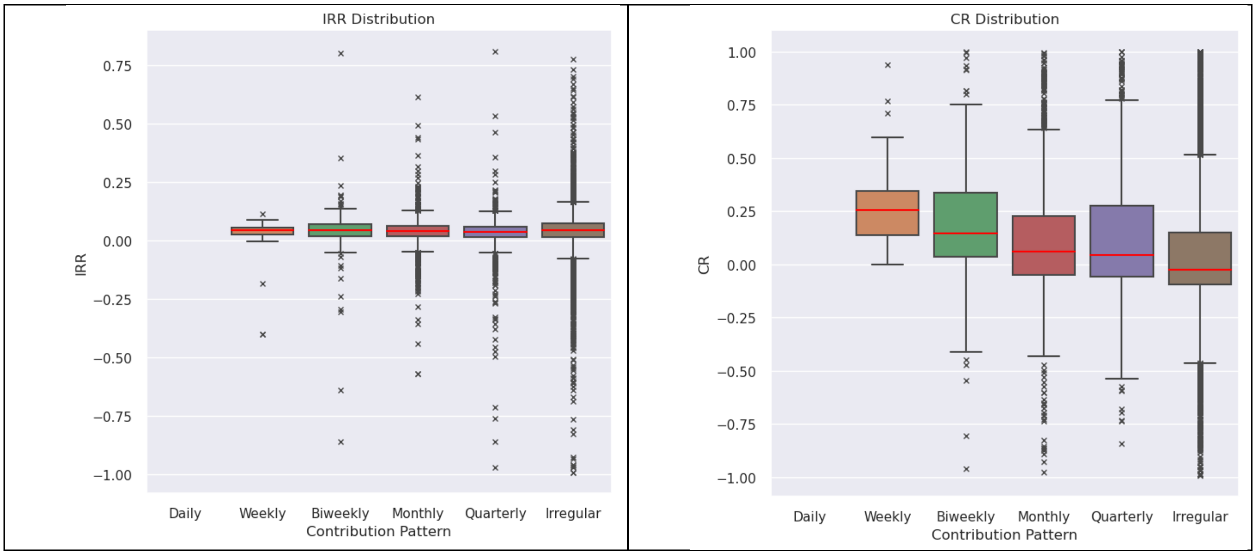


Figure 13. Dataset 2—Savings Frequency and Wealth Accumulation

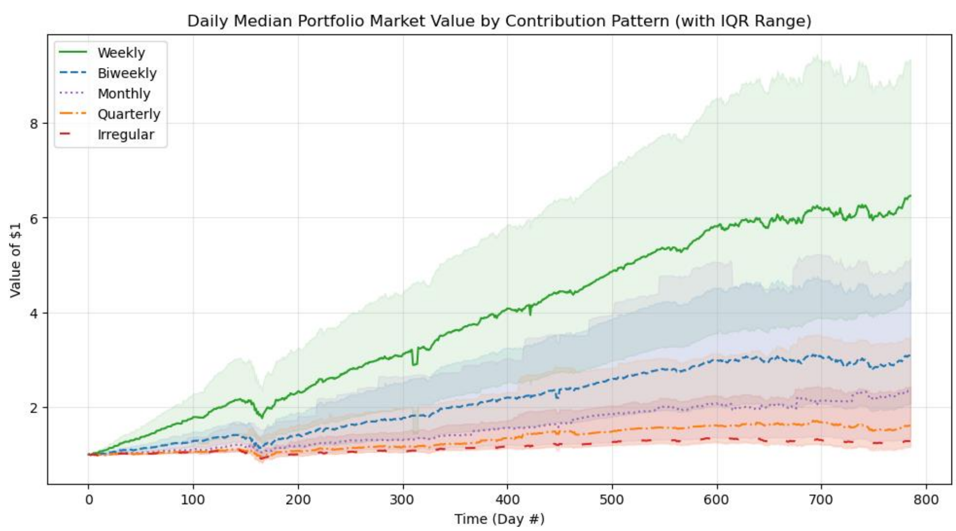
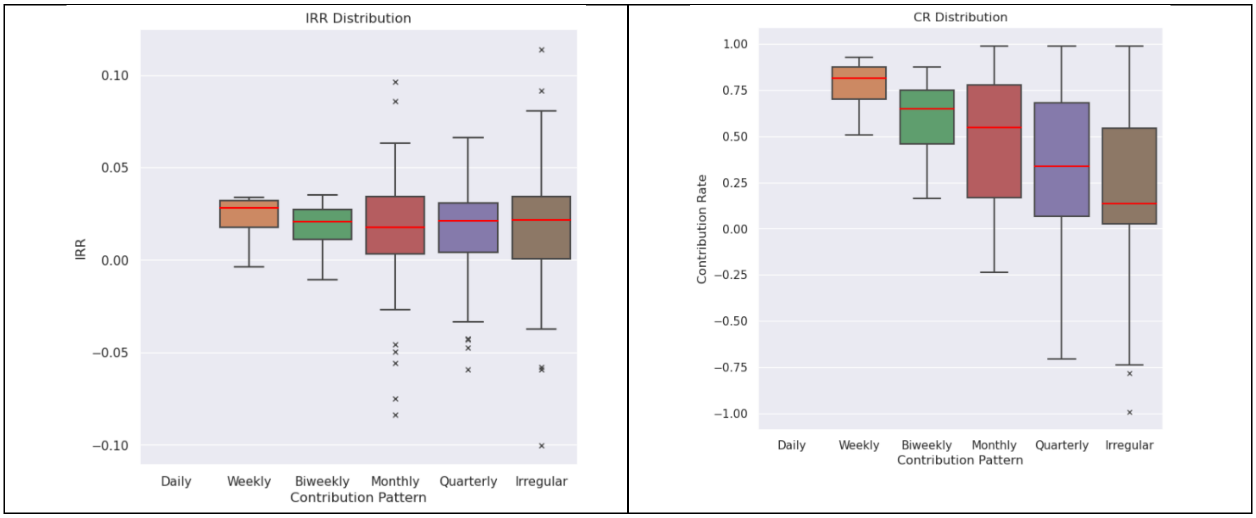


Figure 14. Dataset 2—Savings Frequency and IRR versus CR



References

Avanzi, B., Taylor, G., Wong, B., & Xian, A. (2021). Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *European Journal of Operational Research*, *290*(1), 177-195.

Baldwin, R. D. (2022). The Evolving Wealth of Canadians: Who Is Better Fixed for Retirement? Who is Not?. *Who is Not*.

Bayraktar, E., & Young, V. R. (2010). Optimal investment strategy to minimize occupation time. *Annals of Operations Research*, *176*(1), 389-408.

Blake, D., Wright, D., & Zhang, Y. (2013). Target-driven investing: Optimal investment strategies in defined contribution pension plans under loss aversion. *Journal of Economic Dynamics and Control*, *37*(1), 195-209.

Cagetti, M. (2003). Wealth accumulation over the life cycle and precautionary savings. *Journal of Business & Economic Statistics*, *21*(3), 339-353.

Ülkümen, G., & Cheema, A. (2011). Framing goals to influence personal savings: The role of specificity and construal level. *Journal of marketing research*, *48*(6), 958-969.

Cobb-Clark, D. A., Kassenboehmer, S. C., & Sinning, M. G. (2016). Locus of control and savings. *Journal of Banking & Finance*, *73*(C), 113-130.

Cruz, M., & Ahmed, S. (2016). On the impact of demographic change on growth, savings, and poverty. *World Bank Policy Research Working Paper*, (7805).

D'Urso, P., Cappelli, C., Di Lallo, D., & Massari, R. (2013). Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications*, *392*(9), 2114-2129.

Dholakia, U., Tam, L., Yoon, S., & Wong, N. (2016). The ant and the grasshopper: Understanding personal saving orientation of consumers. *Journal of Consumer Research*, *43*(1), 134-155.

Donepudi, P. K. (2019). Automation and machine learning in transforming the financial industry. *Asian Business Review*, *9*(3), 129-138.

Dose, C., & Cincotti, S. (2005). Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A: Statistical*

Mechanics and its Applications, *355*(1), 145-151.

Dynan, K. E., Skinner, J., & Zeldes, S. P. (2004). Do the rich save more?. *Journal of political economy*, *112*(2), 397-444.

Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J. (2019, May). Trends and applications of machine learning in quantitative finance. In *8th international conference on economics and finance research (ICEFR 2019)*.

FCAC. (2021). Make change that counts: National financial literacy strategy 2021-2026. https://www.canada.ca/en/financial-consumer-agency/programs/financial-literacy/financial-literacy-strategy-2021-2026.html

Feiveson, L., & Sabelhaus, J. (2019). Lifecycle patterns of saving and wealth accumulation.

Fisher, P. (2010). Gender differences in personal saving behaviors. *Journal of Financial Counseling and Planning*, *21*(1).

Foerster, S., Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2017). Retail financial advice: does one size fit all?. *The Journal of Finance*, *72*(4), 1441-1482.

Gale, W. G., Iwry, J. M., & Orszag, P. R. (2005). *The automatic 401 (k): A simple way to strengthen retirement savings* (pp. 2006-1). Retirement Security Project.

Hartman, J. C., & Schafrick, I. C. (2004). The relevant internal rate of return. *The Engineering Economist*, *49*(2), 139-158.

Hall, C. C. (2021). Promoting savings for financial resilience: expanding the psychological perspective. *Current Directions in Psychological Science*, *30*(1), 49-54.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226-251.

Hershfield, H. E., Goldstein, D. G., Sharpe, W. F., Fox, J., Yeykelis, L., Carstensen, L. L., & Bailenson, J. N. (2011). Increasing saving behavior through age-progressed renderings of the future self. *Journal of marketing research*, *48*(SPL), S23-S37.

Huang, Y. P., & Yen, M. F. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. *Applied Soft Computing*, *83*, 105663.

Juster, F. T., Smith, J. P., & Stafford, F. (1999). The measurement and structure of household wealth. *Labour Economics*, *6*(2), 253-275.

Kahneman, D. (2012). The human side of decision making. *Journal of Investment Consulting*, *13*, 9-14.

Kahneman, D., & Thaler, R. H. (2006). Anomalies: Utility maximization and experienced utility. *Journal of economic perspectives*, *20*(1), 221-234.

Kempson, E., Finney, A., & Poppe, C. (2017). *Financial well-being: A conceptual model and preliminary analysis*. Working Paper. https://www.researchgate.net/publication/318852257_Financial_Well-Being_A_Conceptual_Model_and_Preliminary_Analysis?channel=doi&linkId=59819dd14585150575c00b54&showFulltext=true

Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1-2), 307-319.

Kou, G., Peng, Y., & Wang, G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information sciences*, *275*, 1-12.

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29.

Lim, T., & Ong, C. S. (2020). Portfolio diversification using shape-based clustering. *The Journal of Financial Data Science*.

Lin, W. Y., Hu, Y. H., & Tsai, C. F. (2011). Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(4), 421-436.

Linnainmaa, J. T., Melzer, B. T., & Previtero, A. (2021). The misguided beliefs of financial advisors. *The Journal of Finance*, *76*(2), 587-621.

MacGee, J., Pugh, T. M., & See, K. (2022). The heterogeneous effects of COVID-19 on Canadian household consumption, debt and savings. *Canadian Journal of Economics/Revue canadienne d'économique*, *55*, 54-87.

Nenkov, G. Y., MacInnis, D. J., & Morrin, M. (2009). How do emotions influence saving behavior. *Center for Retirement Research*, *9*(8), 1-11.

Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance*, *46*(2), 469-477.

Markowitz, H. M. (2010). Portfolio theory: as I still see it. *Annu. Rev. Financ. Econ.*, *2*(1), 1-23.

Maynard, N., De Neve, J. E., & Tufano, P. (2008). Consumer demand for prize-linked savings: A preliminary analysis. *Harvard Business School Finance Working Paper*, (08-061).

Metzler, Z., & Grace, C. (2021). Learning about financial health in Canada. *Quantitative Finance & Economics*, *5*(3): 542-570.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867-887.

Modigliani, F., & Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. *Franco Modigliani*, *1*(1), 388-436.

Montmarquette, C., & Viennot-Briot, N. (2015). The value of financial advice. *Annals of Economics and Finance*, *16*(1), 69-94.

Newmeyer, C., Warmath, D., O'Connor, G. E., & Wong, N. (2021). Is savings automation helpful to liquid savings? It depends on whether you have a savings habit. *Journal of public policy & marketing*, *40*(2), 285-297.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, *42*(1), 259-268.

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition*, *44*(3), 678-693.

Raffinot, T. (2017). Hierarchical clustering-based asset allocation. *The Journal of Portfolio Management*, *44*(2), 89-99.

Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, *10*(1), 1-9.

Rundo, F., Trenta, F., di Stallo, A. L., & Battiato, S. (2019). Machine learning for quantitative finance applications: A survey. *Applied Sciences*, *9*(24), 5574.

Shefrin, H. M., & Thaler, R. H. (2004), Mental accounting matters (chapter 14). *Advances in Behavioural Economics*. Princeton University Press

Soman, D., & Cheema, A. (2011). Earmarking and partitioning: Increasing saving by low-income households. *Journal of Marketing Research*, *48*(SPL), S14-S22.

Suh, E. (2022). Can't save or won't save: financial resilience and discretionary retirement saving among British adults in their thirties and forties.

Ageing & Society, *42*(12), 2940-2967.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, *112*(S1), S164-S187.

Thompson, J. R., Feng, L., Reesor, R. M., & Grace, C. (2021). Know Your Clients' behaviours: a cluster analysis of financial transactions. *Journal of Risk and Financial Management*, *14*(2), 50.

Tsai, C. F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, *16*, 46-58.

Turner, T. M., & Luea, H. (2009). Homeownership, wealth accumulation and income status. *Journal of Housing Economics*, *18*(2), 104-114.

West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & operations research*, *32*(10), 2543-2559.

Van Liebergen, B. (2017). Machine learning: a revolution in risk management and compliance?. *Journal of Financial Transformation*, *45*, 60-67.

Vlaev, I., & Elliott, A. (2014). Financial well-being components. *Social indicators research*, *118*, 1103-1123.