CANADA'S FINANCIAL WELLNESS LAB

The Impact of Saving on Financial Resilience

Keeping It Simple

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

***** Correspondence: fwl@uwo.ca

September 2023

[1] Ivey Business School, London, ON N6G 0N1, Canada; cgrace@ivey.ca
[2] Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada
[3] Department of Mathematics, Western University, London, ON N6G ON1, Canada
[4] Financial Wellness Lab, Western University, London, ON N6G 0N1, Canada
[5] Financial Wellness Lab, Western University, London, ON N6G ON1, Canada

The Financial Wellness Lab's mandate is to develop quantitative finance and data analytics solutions that will enable Canadian households to enhance their financial resilience. The Lab resides in Western University's Faculty of Science, leveraging strengths in actuarial science, financial modelling, and data science. Academic partners from the Ivey Business School, Wilfrid Laurier University, UBC Okanagan and the University of Winnipeg are also part of the team. The academic team is joined by six industry partners who are providing the lab with anonymized data on the financial circumstances of thousands of Canadians showing how much they spend, save and earn.

An earlier white paper from the Lab (December 2021) concluded that savings, spending, and debt play uniquely powerful roles in financial resilience. The lab's analysis to date has concluded that each is a determining factor in the lab's clustering algorithms.

In this paper, we take a deeper dive into the issue of saving. To do so, we examined unique dataset(s) of investor transactions to determine the relationship between investor behaviours, household savings, and investment outcomes. We examined real-world observed behaviours through advanced data analytics in the form of machine learning to explore previously unknown patterns (referred to as clusters) and seek a determination of any causal relationships.

We examined trading over a 3-year period ending August 2022, providing us with the opportunity to observe behaviour during rising markets, declining markets and the turbulent phases during transitions. Our datasets included investors who work with financial advisors and those who prefer "do it yourself".

Trading behaviours over this period, demonstrated an active savings strategy to be the most effective strategy for building wealth. On average, an active savings strategy was 5X more effective at building wealth and resilience than relying on investment returns or complex trading strategies. We found little evidence to suggest active trading resulted in superior investment returns or unusual growth in wealth.

We concluded that;

1. Saving is a 'force of nature'. The math isn't new, but it works and we observed it working in the 'real world".
2. Saving is simpler, more reliable, and more powerful than investment returns for building wealth and financial resilience.
3. Frequent and disciplined saving is more effective than periodic 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.
5. Keeping it simple is a legitimate strategy for building wealth.

In summary, 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.

There are several implications for policy makers, regulators, asset managers, financial advisors, employers and consumers. The dominant theme amongst of the implications is an encouragement for the intermediaries to "get out of the way", simplify our approaches to savings and help Canadians achieve financial resiliency through the power of saving.

TABLE OF CONTENTS

TABLES AND FIGURES

SECTION 1: INTRODUCTION

In this paper we examined unique dataset(s) of investor transactions to determine the relationship between investor behaviours, household savings and investment outcomes. The topic is important because savings have been shown to be a cornerstone for financial resilience.

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 (Kempson et al., 2017, Metzler et al., 2021) and financial stress can lead to more serious health problems (Chen, 2013). Financial stress is alarmingly widespread among the Canadian population. Metzler et al., find that between 2009 and 2019, one third of employed Canadians were financially stressed. If a large fraction of Canadian households are financially fragile, Canadian Society is less able to withstand unexpected shocks (Bank of Canada, 2017).

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 but a better understanding of the prevalence and type of financial fragility in Canadian households may allow more targeted policy interventions in the future.

We contribute to our understanding of this topic by examining real-world observed behaviours using advance analytics. We examined these behaviours by deploying advanced data analytics in the form of machine learning to examine previously unknown patterns (referred to as clusters) and a determination of a causal relationship.

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.

The datasets include anonymized data on investors under two discrete 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 (Do It Yourself).

Both datasets were limited to retirement savings accounts, allowing us to control for such things as time horizon, risk tolerance and structural constraints. The datasets included data points down to the daily transaction level.

The machine learning algorithms determined that investors could be clustered into one of 3 groups (for each dataset) that were defined by their trading behavior and portfolio outcomes. A brief description of the resulting clusters is noted in Table 1.

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 EI 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 2). As a result, our study provided a unique opportunity to observe the impact of unusual cash inflows on household resiliency – as measured by wealth.

In our analysis, we hypothesized that investors who contribute to their savings plans in a regular and systematic manner, regardless of market conditions, will maximize their wealth over time. We also assumed that the accumulation of wealth and regular saving habits contribute to a positive state of financial wellness.

We observed that in both Datasets, the 'winning' cluster followed an aggressive savings pattern. They saved systematically and frequently, regardless of market conditions, and grew their wealth, on average, 5 times[6] faster than the investors who relied on the markets or complex trading strategies alone.

Our analysis concluded that a consistent pattern of saving – even in turbulent markets - is a 'winning' strategy for wealth accumulation and financial resilience. We noted 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 periodically. The observed results were consistent with stochastic simulations implying that 'the math works.

Investment performance played a role in wealth accumulation, but it was muted when compared to savings behaviour. Investment performance appeared to be driven by risk tolerance and asset mix.

TABLE 1 CLUSTER DESCRIPTIONS

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.

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

We observed 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. It would therefore be premature to draw definitive conclusions.

We also concluded that 'keeping 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.

The bottom line - saving patterns were by far the most powerful determinant of wealth accumulation and financial resiliency. Or stated a little differently, it turns out Grampa was right. The secret to accumulating wealth – and financial wellness - is to **"pay yourself first."**

[6] See Appendix 5 for detailed calculations

The paper reads as follows:

Section 2 provides important context regarding the Canadian investment industry and its regulatory regime. It includes a literature review on pertinent research covering retail financial advice, product choice, risk tolerance, trading strategies and savings. It also includes background information and regarding the application of machine learning algorithms on retail finances.

Section 3 articulates our working hypothesis' that investors who contribute to their savings in a systematic fashion will optimize their wealth accumulation.

Section 4 provides a detailed description of our clustering methodologies including Dynamic Time Warping and K-means.

Section 5 provides a description of our datasets and the features used in our clustering.

Section 6 provides a summary of our analysis and conclusions.

Section 7 includes industry implications, the limitations inherent in our datasets and the case for further research.

Section 8 provides pertinent disclosures and an abbreviations glossary.

Section 9 provides our appendices including detailed methodologies, our simulations and our references.

SECTION 2: BACKGROUND AND CONTEXT

SECTION 2.1: INDUSTRY BACKGROUND

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[7], 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

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 IRR calculations and our savings calculations.

THE ROLE OF ADVICE

Dataset 1 encompassed investors who are working 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 (IIROC)[8]. Under the IIROC regime, advisors provide a broad range of services and can recommend investment solutions from thousands of investment choices[9]. The IIROC regime

---

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

[8] 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

[9] 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.

encompasses 174 dealers, 32,105 'approved persons' or advisors and approximately $8T in assets under administration[10].

RISK TOLERANCE

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 (Thompson et al 2021) we 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.

TRADING STRATEGIES

Canadian investors generally follow one of two trading strategies with their RSPs – active or passive.

Active trading refers to the periodic trading in specific securities, typically to deliver alpha (unusual returns) or minimize risk (volatility). DIY investors can trade as often as they wish while advised investors are presumably influenced by their advisor's recommendations and availability.

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.

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.

SECTION 2.2: LITERATURE REVIEW

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 – to name a few. 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 (Kempson et al 2017, Metzler et al 2021, FCAC 2021). And it is even more recent that scientists have been able to turn to advanced data analytics such as machine learning to explore the topic.

Our literature review is broken down into seven subsections including economic drivers and policy interventions, behavioural interventions, demographic drivers, an industry view, and applications of machine learning in finance.

2.2.1 Economic drivers and 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, W. 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[11].

[10] Investment Industry Regulator Organization of Canada, https://www.iiroc.ca
[11] Statistics Canada, www150.statcan.gc.ca/, Table 11-10-0016-01, released 2020-12-22, sourced June 2023

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, C. 2021)

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).

FIGURE 1 HOUSEHOLD SAVINGS RATES AS % OF DISPOSABLE INCOME[12]

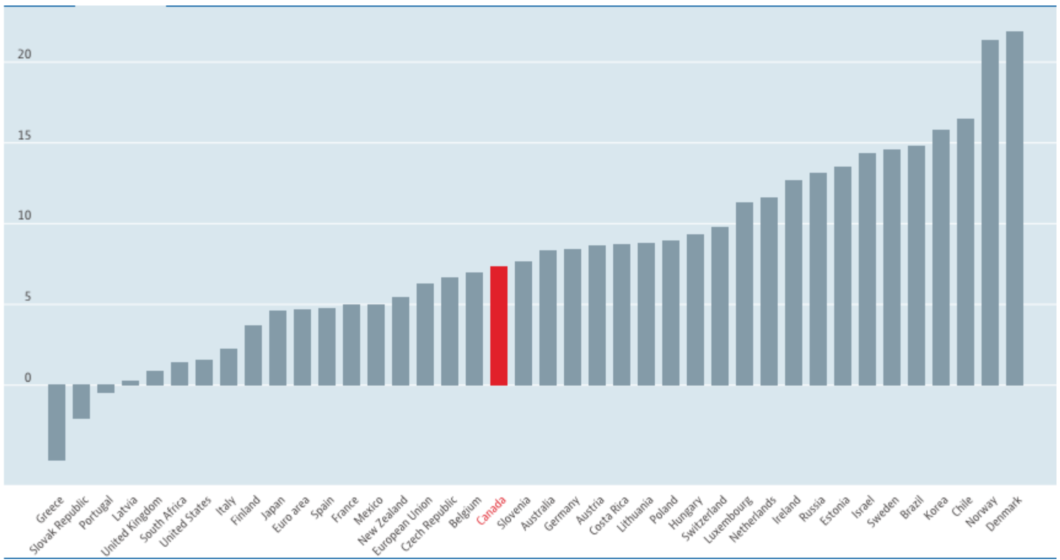
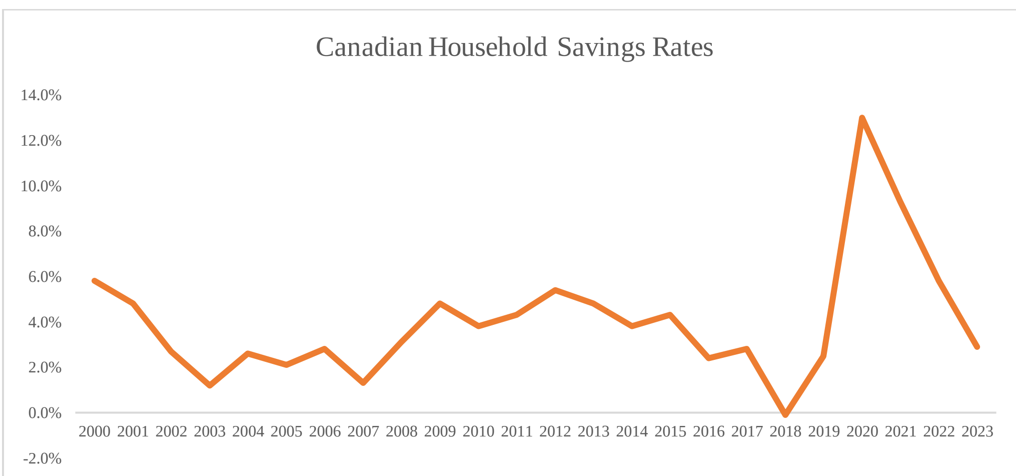


FIGURE 2 CANADIAN HOUSEHOLD SAVINGS RATES AS % OF DISPOSABLE INCOME[13]



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

2.2.2 Behavioural Interventions

Kahneman (Kahneman 2012) and Thaler (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), feelings of power (Garbinsky et al. 2014) and risk aversion (Cagetti 2003) have all been linked to savings behaviour. Dholakia, in turn, (Dholakia 2006) 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 (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).

2.2.3 Demographic Drivers

Researchers have linked savings rates and resiliency to several demographic factors including age (Tufino 2008, Baldwin 2022), income ((Dynan et al 2004, Turner et al 2009, MacGee 2022, Cruz 2016), household composition (Cobb-Clark et al. 2016), and gender (Fisher 2010). These factors are often combined under the title 'life cycle model' (Feiveson et al 2019).

However, Metzler (Metzler et 2021) has noted that while these demographic traits can be linked to financial resilience, the data does not support a causal relationship.

2.2.4 Finance and Actuarial Sciences

In the 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.

Further to these approaches is the examination of a widespread practice of dollar-cost averaging (DCA) or what we refer to as systematic saving in this paper. DCA refers to the practice of investing a constant amount in a security at regular intervals regardless of price or market conditions. DCA has been widely panned by finance scholars (Knight et al 1992, Statman 1995, Milevsky et al 2003, Brennan et al 2005) since Constantinides demonstrated its inefficiency in 1979 (Constantinides, 1979). In this paper, we haven't attempted to duck the debate on the effectiveness of DCA as an investment technique but rather to examine the holistic impact of systematic saving on wealth accumulation and financial resiliency.

2.2.5 Industry Perspectives

The wealth management industry in Canada is estimated to encompass USD $4.5T in assets under management[14]. The typical Canadian has savings with their financial institution (chequing and savings accounts), through a pension/RSP or a TFSA (see Table 2)

[14] www.mordorintelligence.com/industry-reports/, source50.4d June 29 2023

Within these assets, Canadians have access to 400 pension providers[15] and approximately 6,900+ listed securities across 7 different stock exchanges[16], 4,000+ mutual funds[17] and 1,100+ exchange traded funds[18]. The saturated level of competition has led to confusion amongst retail investors and a push by regulators for more transparent and simplified disclosures[19].

TABLE 2 CANADIAN HOUSEHOLD FINANCIAL ASSETS[20]

	Percent of Canadians holding	Average Value
Deposits in financial institutions	95.1	$5,000
Private pension assets	71.8	$158,700
Registered Retirement Savings Plans (RRSPs)/Registered Retirement Income Funds (RRIFs)	58.5	$50,000
Employer sponsored Registered Pension Plans (EPPs)	50.4	$163,900
Tax Free Saving Accounts (TFSAs)	45.8	$15,000
Other financial assets	21.5	$14,000
Mutual funds, investment funds and income trusts	10.7	$50,000
Stocks	8	$23,000
Bonds	2.3	$5,000

2.2.6 The Role of Advice

The role of a financial advisor with respect to household savings patterns has not been widely researched. Industry studies (Russell[21], Vanguard[22]) estimate advisors add 150 to 200 basis points (bps) to portfolio growth through coaching and investor discipline. Foerster et al (Foerster 2017, 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 do not work with advisors held fewer funds and did not take advantage of automated savings plans. In contrast, investors who did work with advisors benefited from greater savings.

2.2.7 Financial Wellness

Recently, savings have been linked to the concept of financial wellness (Vlaev et al 2014, Kempson et al 2017, Suh 2021, Metzler 2021). Financial stress – the antithesis of wellbeing - can affect anyone – young or old, married or single, rich or poor. And when it does, its impact can overwhelm us at home and at work. Lost

[15] Federally regulated only (www.osfi-bsif.gc.ca), sourced June 29 2023
[16] TSX, TSX Venture Exchange, Canadian Securities Exchange, Montreal Exchange, NASDAQ Canada, Vancouver Stock Exchange and Alberta Stock Exchange
[17] Investment Funds Institute of Canada (www.ific.ca/en/), sourced June 29 2023
[18] Canadian ETF Association (www.cetfa.ca), sourced June 29 2023
[19] Canadian Securities Administrator (www.securities-administrators.ca), sourced June 29 2023
[20] Statistics Canada, www150.statcan.gc.ca/, Table 11-10-0016-01, released 2020-12-22, sourced June 2023
[21] Russell Investments (www.russellinvestments.com), sourced June 29 2023
[22] Vanguard (www.vanguard.com/pdf/ISGQVAA.pdf), sourced June 29 2023

productivity due to employees worrying about personal finances is estimated to cost Canadian employers over $20 billion[23] each year or 7 to 14 working days per year[24].

A financially resilient individual is able to 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[25]. 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.

In their research paper entitled "The financial resilience and financial well-being of Canadians during the Covid-19 pandemic", Statistics Canada[26] notes: "As Canada begins the path to recovery from the health and economic impacts of the COVID-19 pandemic, there is a heightened awareness of the need for households to maintain or build their financial resilience. Global uncertainty, increasingly longer life spans, the changing nature of work, changing work-life patterns, housing affordability, high debt loads, and the impact of unplanned life events, amongst other factors, mean that many Canadians are needing to manage and, where possible, improve their financial resilience. The unprecedented situation brought on by the global pandemic has had significant economic impacts on most people in Canada, challenged the finances of many households and exacerbated inequalities that existed before the crisis. While some households have profited financially from the pandemic, others have experienced a considerable downturn."

In their ground-breaking research, Kempson et al. 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) 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.

2.2.8 Machine Learning Algorithms in Finance

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 ma- chines (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

[23] Canadian Payroll Association, *Nothing is Normal*, 2020
[24] Lifeworks, The Financial Wellbeing Index™Report, 2021
[25] Manulife, 2016 Financial Wellness Index
[26] Statistics Canada, *The financial resilience and financial well-being of Canadians during the Covid-19 pandemic,*

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 the suitability and client trading behaviours, the focus of this paper.

2.2.8. Introducing the Parsimony Principle to Financial Resilience

The parsimony principle (lex parsimoniae in Latin) is typically associated with Occam's Razor which states that entities should not be multiplied beyond necessity. Whenever we have different explanations of the observed data, the simplest one is preferable. "If you hear hoofbeats, think horse -- not zebra."

Parsimony has played a role in scientific research for years. In statistics and machine learning for example, parsimony means that if two models can explain the data equally well then, the one that contains the fewest parameters should be preferred. In mathematical models, the relationship between Occam's razor and rational probabilistic inference was first pointed out by Jeffreys (1939/1961). Jeffreys argued that the simplest interpretation was indeed the most likely one and advocated for penalizing complexity. Edwards (1972) has argued that probability theory inherently favors simpler inductions.

More recently, parsimony has been applied to prescriptions and not just the analytics – and, specific to this paper, the topics of habit forming and behaviour. In psychology and cognitive science, the simplicity principle suggests that the mind draws interpretations of the world—mental models or mental representations—that are as simple as possible, or, at least, that are biased towards simplicity (Chater, 1997; Chater & Vitányi, 2003). The idea takes different forms in different areas of cognition, depending on the nature of the many perceptual and cognitive problems the mind encounters. In many areas of cognition, researchers have found that human thought has a bias towards simplicity because, presumably, it helps us navigate a complex world.

The 'western' or commercialized world has made personal financial management complex in their pursuit of competitive advantage, scale and their own P&L motivations. New policies and government interventions are 'blunt instruments' that can foster unintended consequences including competing messages and complexity. Seeking solutions through a behavioural lens is informative but behaviour is often difficult to predict in the moment and solutions are a challenge to institutionalize at scale. Policy makers and industry have attempted to counter this complexity through financial literacy programs but the curriculum in those programs is, by definition, always one step behind.

SECTION 3: OPERATING HYPOTHESIS

In our analysis, we examined our datasets through the lens of an operating hypothesis:

We hypothesize that the investors who contribute to their savings plans in a regular and systemic manner, regardless of market conditions, will optimize their wealth accumulation over time.

We tested our hypothesis by examining investor behaviour at the transaction level using machine learning algorithms widely used in financial applications. The use of machine learning techniques such as clustering allowed the data to speak without bias.

We were also cognizant of testing the hypothesis from a practical perspective and searching for patterns that led to easy to implement solutions and practices controllable by the investor.

We have assumed, based on a preponderance of evidence, that wealth accumulation is a key foundation in financial resilience and wellness.

SECTION 4: CLUSTERING METHODOLOGIES

In our study, we deployed machine learning to uncover previously unknown patterns in the data. Machine learning – and in particular a technique known as clustering – has proven to be invaluable when examining large, complex datasets such as financial transactions. Our datasets encompassed over 200 discrete variables that sometimes change daily. Traditional analytics such as regression models are able to identify correlations amongst many of these variables or features, but machine learning allows us to get closer to causal relationships.

Specifically, we deployed two machine learning techniques: Dynamic Time Warping and K-Means clustering.

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' investment behaviors.

Afterward, we applied K-Means clustering algorithms to categorize portfolios exhibiting similar movements. 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 validates our data-driven grouping and helps to highlight the unique trends within each cluster.

A complete description of our methodology can be found in Appendix 6.

FIGURE 3 DATASET 1 CLUSTER VISUAL

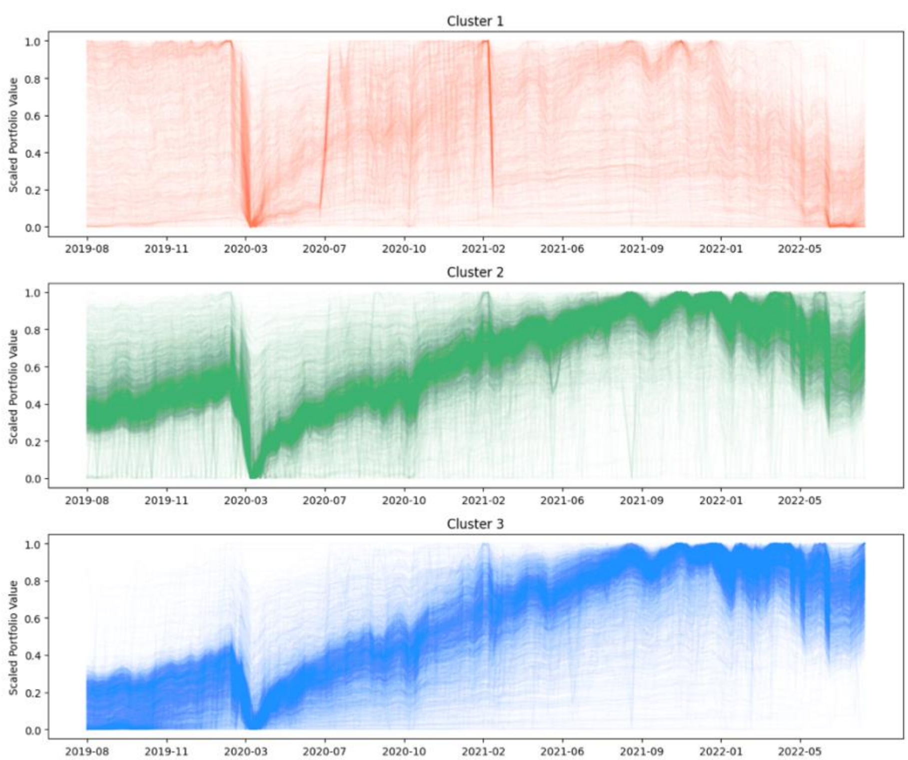
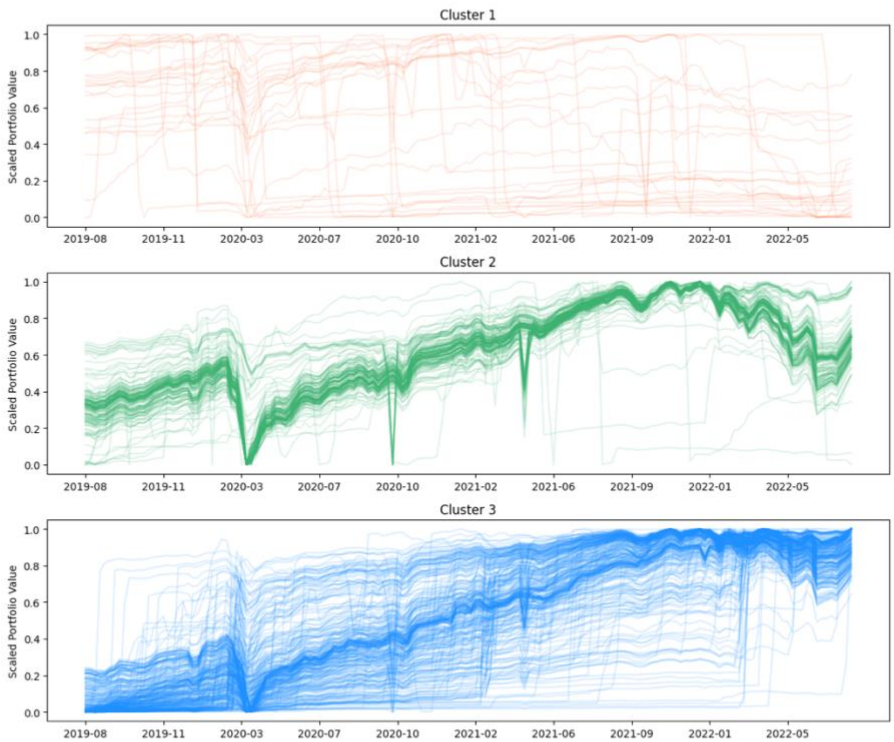


FIGURE 4 DATASET 2 CLUSTER VISUAL



SECTION 5: DATASETS

DATASET 1

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. In total, clients with advisors who work with the dealer have over 50,000 accounts. 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.

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.

Table 3 shows the details of the available KYC information.

DATASET 2

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 using anonymization practices, there is still 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.

SECTION 6: ANALYSIS

6.1 CLUSTERING RESULTS

Our clustering identified 6 unique groups – 3 per dataset – based on their trading behaviour (Table 3). Within each dataset, the 3 clusters were very similar in terms of KYC data (age, income, gender, risk tolerance etc. but the clusters were differentiated by their trading behaviour.

An analysis of the trading behaviours for this set of investors, over this sample period, demonstrated that an active savings strategy is the most effective strategy for building wealth. We found little evidence to suggest active trading resulted in superior returns (for neither the advised or the DIY investors) or unusual growth in wealth. An active savings strategy was, on average, 5X more powerful at building wealth than relying on investment returns (see Appendix 5).

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 would suggest that these investors were under unusual financial stress over this period of time.

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

We also noted that with respect to "the math" an investor is probably indifferent between IRR and CR[27] since both have very similar effects on the final outcomes (Appendix 4). However, IRR is largely determined by the markets, asset mix and risk tolerance[28] – and the investor has little control over the markets. CR on the other hand is entirely at the discretion of the investor and the outcomes are directly controllable through CR.

The math isn't new, but it works and it was observed to work in the 'real world".

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.

[27] See Appendix 1
[28] Foerster et al

TABLE 3 KYC DATA AND CLUSTER TRADING PATTERNS

	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%

6.2 CLIENT OUTCOMES

In terms of impact, the 6 clusters delivered dramatically different outcomes. 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[29]. In Figures 5 and 6, it should be noted that there is no overlap between Cluster 2 and 3 or Cluster 4, 5 and 6 – indicating unique trading patterns.

All 6 clusters had investment returns that were consistent with their risk tolerance and the general market conditions at the time. 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%[30]. A balanced portfolio (60% fixed income, 40% equities) would have had a return of approximately 3.0% before fees.

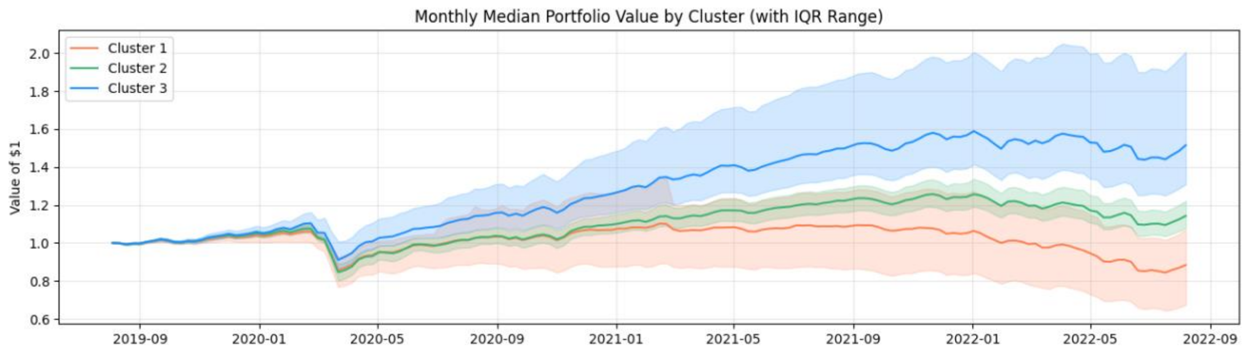
Where the clusters differed was in terms of savings rate (net contributions). Median contribution rates ranged from a low of -59% to a high of 235% compared to their opening balance.

We found no evidence of investment returns driving significant wealth accumulation. Instead returns followed a normal distribution curve with deviations from the mean seemingly random (Figures 7 and8, Table 3).

[29] See Appendix 3

[30] See Appendix 1 and 3

FIGURE 5 DATASET 1: CHANGE IN WEALTH OVER TIME (NORMALIZED TO $1 ON DAY 1)



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

FIGURE 6 DATASET 2: CHANGE IN WEALTH OVER TIME (NORMALIZED TO $1 ON DAY 1) DATASET 2

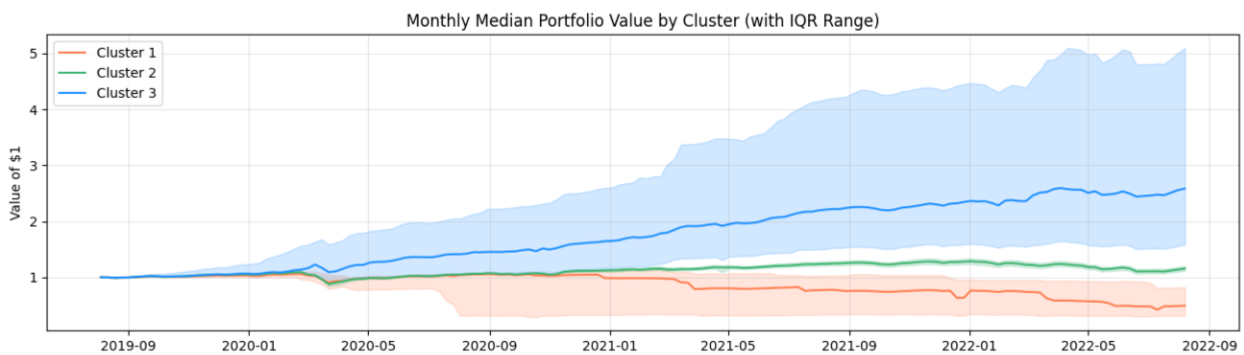


FIGURE 7 DATASET 1: INVESTMENT RETURNS AND SAVINGS RATES

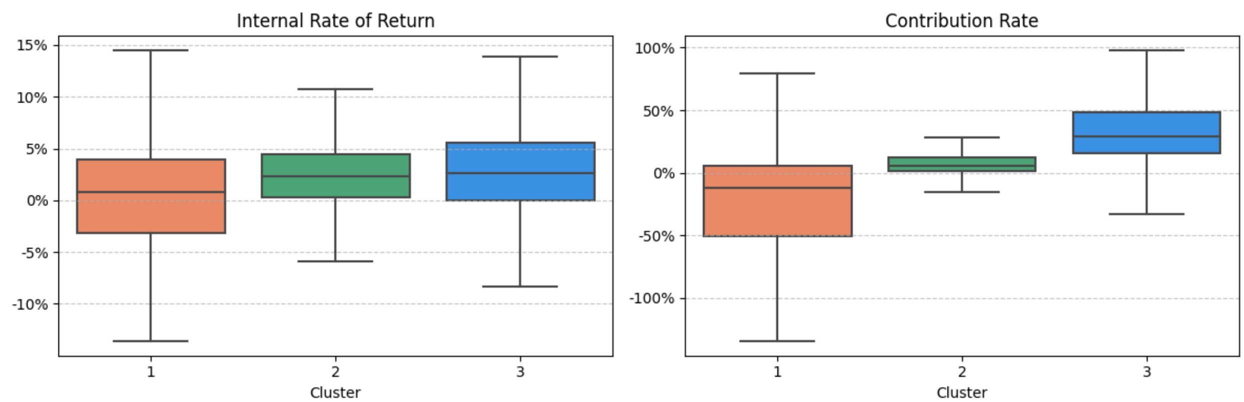
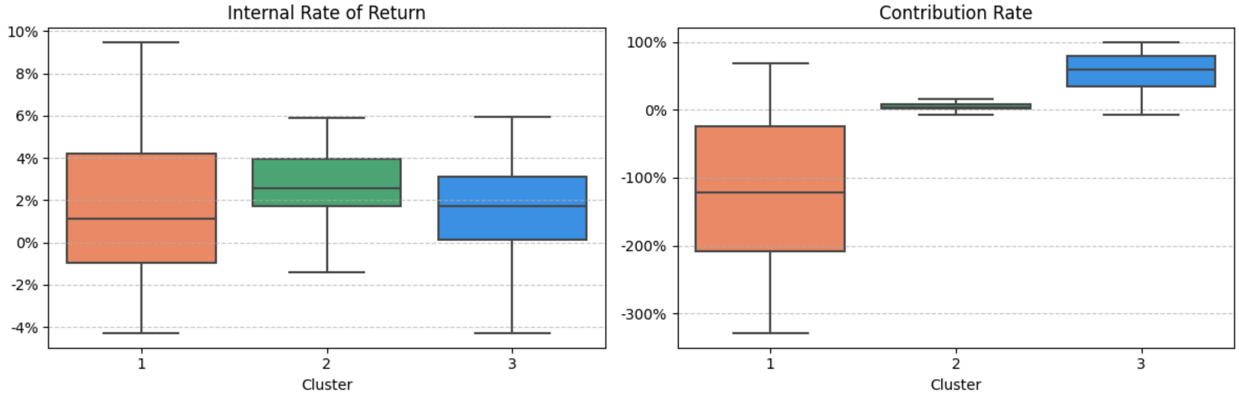


FIGURE 8 DATASET 2: INVESTMENT RETURNS AND SAVINGS RATES



We noted that savings frequency had a marginal impact on investment returns as measured by IRR but a significant impact on savings rates (CR) (Figure 10 and 11, Appendix 4). 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 observation of the impact of payment frequency on investment returns is consistent with simple compounding but inconsistent with the academic literature on dollar-cost averaging. However, as previously noted, the academic literature appears to focus on the impact of dollar cost averaging on the risk/return trade off and ignores the impact on savings rates.

We also 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. It would therefore be premature to draw definitive conclusions.

FIGURE 9 DATASET 1: SAVINGS FREQUENCY AND WEALTH ACCUMULATION

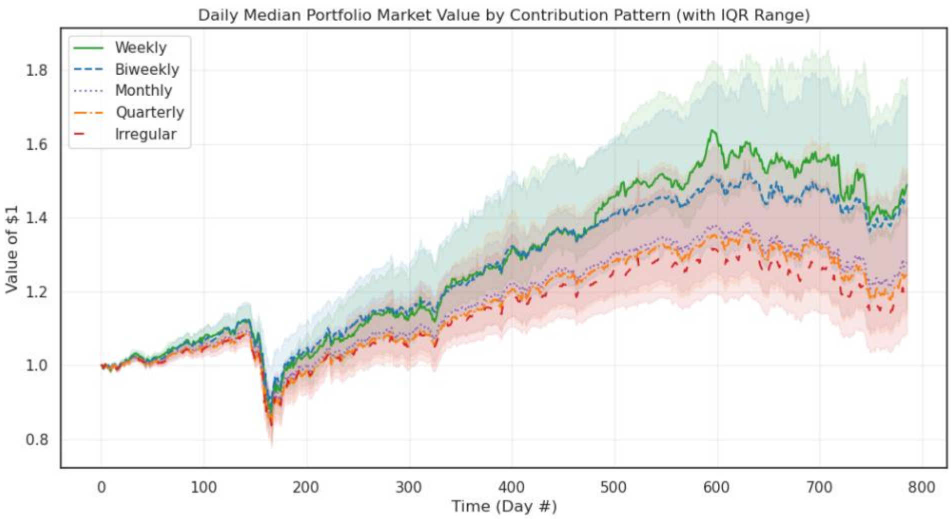


FIGURE 10 DATASET 1: SAVINGS FREQUENCY, IRR AND CR

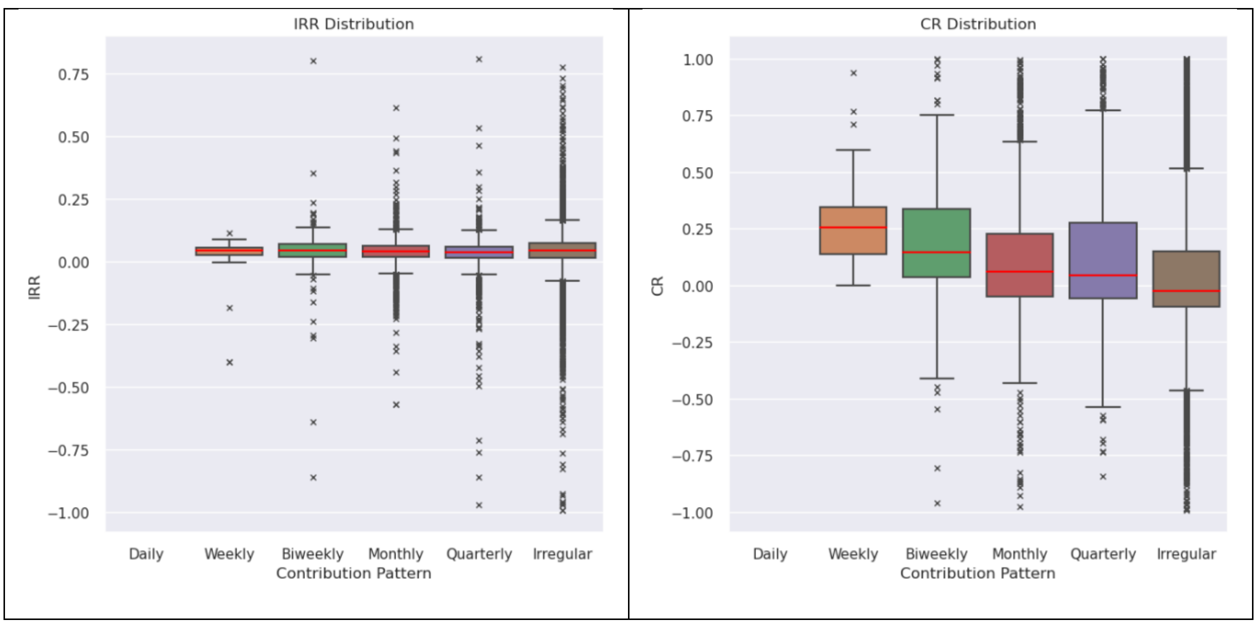


FIGURE 11 DATASET 2: SAVINGS FREQUENCY AND WEALTH ACCUMULATION

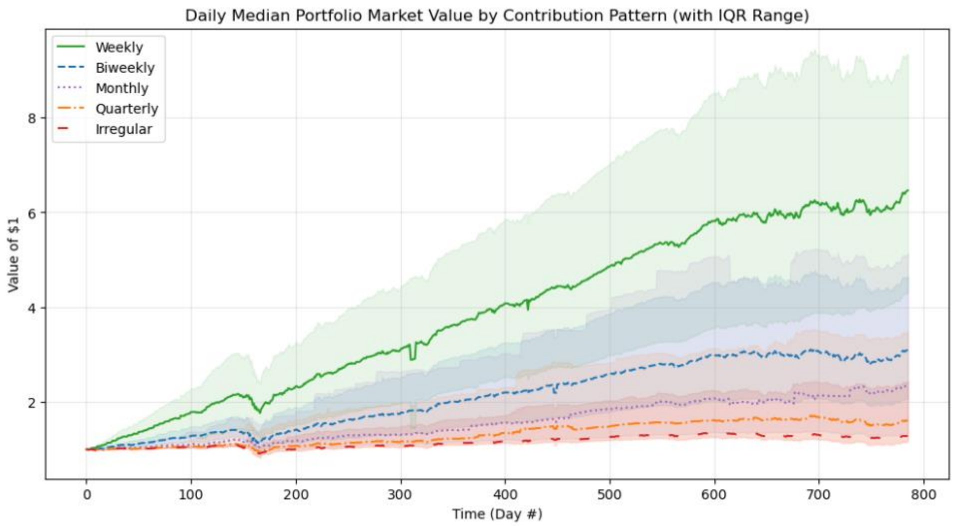
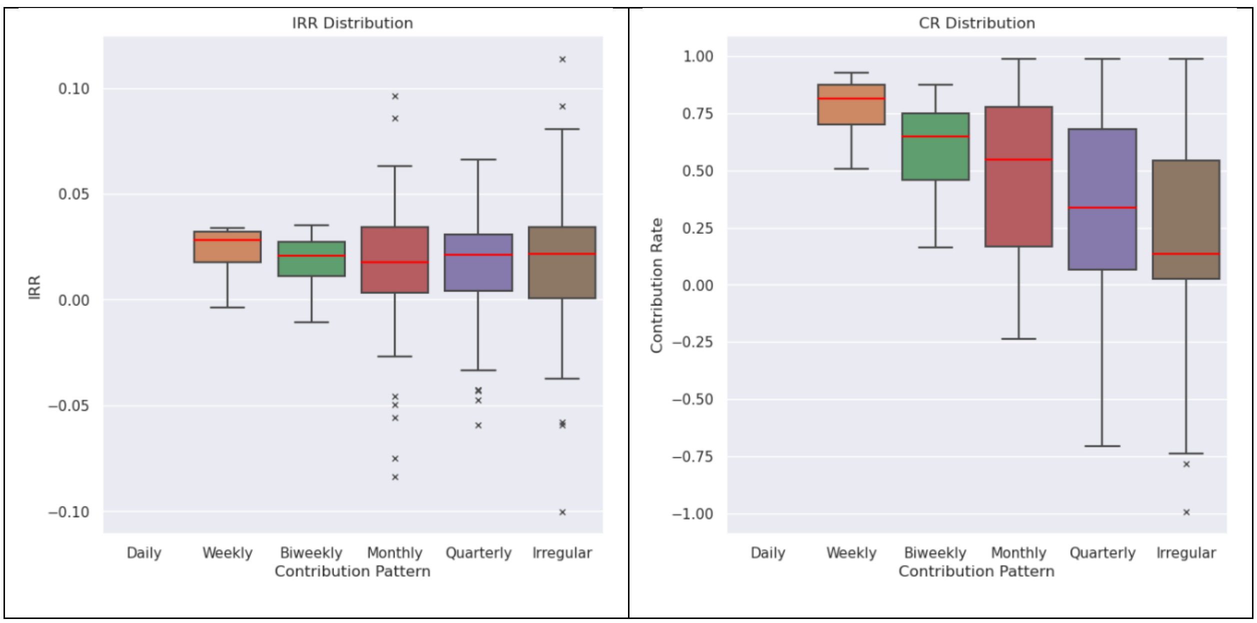


FIGURE 12 DATASET 2: SAVINGS FREQUENCY AND IRR AND CR



SECTION 7: DISCUSSION – LIMITATIONS, IMPLICATIONS AND FURTHER RESEARCH

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 also 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

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%? We would look to behavioural scientists to take the lead on this topic.

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 in the near future.

Our research experimented with a number of 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.

SECTION 8: DISCLOSURES, ABBREVIATIONS

DISCLOSURES

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.

Acknowledgments: The authors would like to thank Matt Davison (Western University), Mark Reesor (Wilfrid Laurier University), Kristina Sendova (Western University) and the many members of our data donor team for their valuable input and insights that improved the content and writing of this document.

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

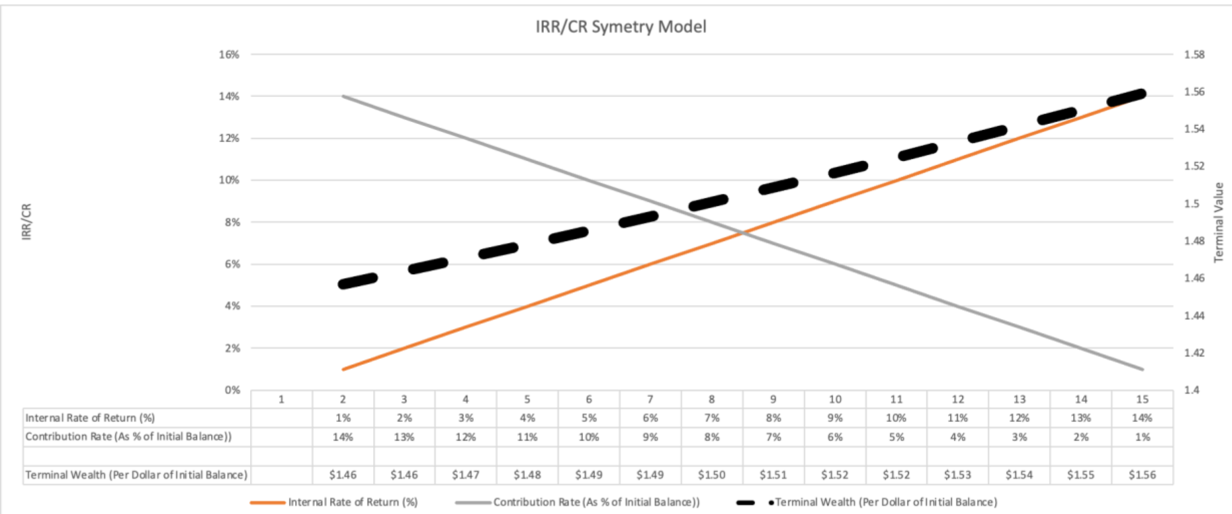
The following abbreviations are used in this manuscript:

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

SECTION 9: APPENDICES

APPENDIX 1: IRR/CR SYMMETRY AND OUTCOMES

In the model below we have projected the terminal value for portfolios assuming a combined mix of IRR and CR. We assumed a 3-year duration and an opening balance of $1.00. Focusing on IRR, to the exclusion of CR, can generate a higher terminal value. However, it is important to note that IRR is subject to market volatility. As well, the model begins to break down as IRRs exceed 10%. Since 1985 the S&P/TSX has had a 5.2%[31] CAGR. In only 17 of those years has the annualize return exceeded 10%.



APPENDIX 2: PORTFOLIO GROWTH ATTRIBUTION

	IRR (annualized)[32]	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%

[31] Yahoo Finance, https://ca.finance.yahoo.com/quote/%5EGSPTSE/history?period1=1660694400&period2=1692230400&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

[32] 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)

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.

APPENDIX 3: INDUSTRY IRR BENCHMARKS

IRR Benchmarks			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 4: SAVINGS FREQUENCY AND ITS EFFECT ON IRR AND CR

Saving Frequency	**n**	**Median IRR (Internal Rate of Return)**	**Median CR (Contribution Rate)**
Dataset 1			
Weekly	27	4.43%	25.68%
Biweekly	213	4.29%	14.64%
Monthly	815	3.96%	6.03%
Quarterly	1,288	3.78%	4.31%
Irregular	4,758	4.33%	-2.39%
Dataset 2			
Weekly	4	2.81%	81.41%
Biweekly	8	2.07%	64.74%
Monthly	81	1.78%	54.9%
Quarterly	155	2.11%	33.58%
Irregular	256	2.18%	13.47%

APPENDIX 5: IRR/CR OUTCOME LEVERAGE

In the model below we have projected the terminal value for portfolios assuming differing durations (based on the investor's age and assuming a retirement age of 65. We included scenarios for 'average' savers (Cluster 2 and 5) and 'aggressive' savers (Clusters 3 and 6). Opening balances, IRR and CR values are based on the corresponding averages from our datasets. On average, Dataset 2 investors were younger than Dataset 1 so we relied on their median values for 30 and 40 year olds, and Dataset 1 for 50 and 60 year olds.

On average, relying on CR alone was 5X more powerful than a relying on IRR alone (buy and hold).

Average Savers								
Flat % of opening balance	Cluster Benchmark	Opening Balance	Duration (years)	IRR	Ending Bal. (IRR only)	CR	Ending Bal. (CR Only)	Leverage (CR/IRR)
30 year old	5	$ 11,845	35	3.1%	$34,482	2.5%	$ 28,111	0.82
40 year old	5	$ 19,008	25	2.4%	$34,390	4.0%	$ 50,672	1.47
50 year old	2	$ 87,083	15	2.8%	$131,774	29.8%	$ 4,355,658	33.05
60 year old	2	$ 94,485	5	2.2%	$105,346	29.2%	$ 340,154	3.23
								9.6
Flat $$ amount based on Yr 1 % of opening balance								
30 year old	5	$ 11,845	35	3.1%	$34,482	2.5%	$ 52,737	1.5
40 year old	5	$ 19,008	25	2.4%	$34,390	4.0%	$ 60,027	1.7
50 year old	2	$ 87,083	15	2.8%	$131,774	29.8%	$ 607,415	4.6
60 year old	2	$ 94,485	5	2.2%	$105,346	29.2%	$ 249,499	2.4
								2.6
Aggressive Savers								
Flat % of opening balance	Cluster Benchmark	Opening Balance	Duration (years)	IRR (from cluster)	Ending Bal (IRR only)	CR (from cluster)	Ending Bal (CR Only)	Leverage (CR over IRR)
30 year old	6	$ 6,108	35	1.6%	$10,646	65.1%		
40 year old	6	$ 9,267	25	2.5%	$17,180	48.1%		
50 year old	3	$ 97,931	15	2.5%	$141,833	6.5%	$ 251,863	1.78
60 year old	3	$ 130,350	5	2.4%	$146,761	5.3%	$ 168,754	1.15
								1.5
Flat $$ amount based on Yr 1 % of opening balance								
30 year old	6	$ 6,108	35	1.6%	$10,646	65.1%	$ 195,278	18.34
40 year old	6	$ 9,267	25	2.5%	$17,180	48.1%	$ 169,436	9.86
50 year old	3	$ 97,931	15	2.5%	$141,833	6.5%	$ 255,979	1.80
60 year old	3	$ 130,350	5	2.4%	$146,761	5.3%	$ 183,002	1.25
								7.81

APPENDIX 6: CLUSTERING METHODOLOGY

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:

$$C_t = barycenter(assoc(C_t))$$

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

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

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.

IRR calculations

To compare the 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.

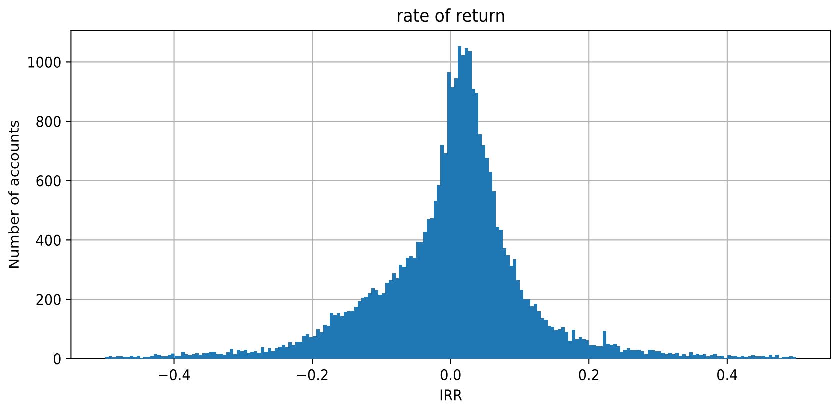
There are two main sources of error in the approximation procedure: the rounding error of the number of shares that are exchanged, and the difference between average trading price and actual trading price. The errors are not material and do not affect the results significantly.

A histogram of the IRRs is given in Figure 13. The table below summarizes the quantiles of the IRR.

Summary of the IRR

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

FIGURE 13 HISTOGRAM OF THE REALIZED IRR



4.2.4 Classification of Accounts by Contribution Patterns

Each of the datasets under study included transaction level detail. In our examination of contribution patterns, we first curated the data to eliminate transactions that did not affect portfolio market values. For example, a client's withdrawal of cash from the cash account or non-reinvested dividends.

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.

A summary of the results is included in Appendix 3.

APPENDIX 7: REFERENCES

- Avanzi, Benjamin, Greg Taylor, Bernard Wong, and Alan Xian. 2021. "Modelling and Understanding Count Processes Through a Markov-Modulated Non-Homogeneous Poisson Process Framework." *European Journal of Operational Research* 290 (1): 177–95.
- Baldwin, 2022, *The evolving wealth of Canadians: Who is better fixed for Retirement? Who is not?*, C.D. Howe Institute, May 2022
- Bayrakter, E., & Young, V., 2010, *Optimal Investment Strategy to Minimize Occupation Time*. Annals of Operations Research, 176 (1), 389-408
- Blake, D., Wright, D., & Zhang, Y. l, *Target-driven investing: Optimal investment strategies in defined contribution pension plans under loss aversion*. Journal of Economic Dynamics & Control, 37(1), 195-209, 2013
- Brennan Li & Torous, 2005, *Dollar Cost Averaging*, European Finance Review
- Cagetti, 2003, *Wealth Accumulation Over the Life Cycle and Precautionary Savings*, Journal of Business & Economic Statistics
- Chater N. Simplicity and the mind. *Psychologist.* 1997;10(11):495–498.
- Chater N, Vitányi P. Simplicity: a unifying principle in cognitive science. *Trends in Cognitive Sciences.* 2003;7(1):19–22.
- Cheema & Ulkumen, 2011, *Framing Goals to Influence Personal Savings: The Role of Specificity and Construal Level*, Journal of Marketing Research
- Cobb-Clark, Kassenbhohmer, & Sinning, 2016, *Locus of control and savings*, Journal of Banking & Finance
- Constantinides, 1979, A Note on the Suboptimality of Dollar-Cost Averaging as an Investment Policy, Journal of Financial and Quantitative Analysis
- Cruz & Ahmed 2016, *On the Impact of Demographic Change on Growth, Savings, & Poverty*, World Bank Policy Research Working Paper No. 7805
- Dose, C., & Cincotti, S., 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, 2005
- D'Urso, Cappelli, Di Lallo, & Massari. 2013. *Clustering of financial time series*. Physica A: Statistical Mechanics and Its Applications 392: 2114–29.
- Dholakia et al 2016, *The Ant & the Grasshopper: Understanding Personal Saving Orientation of Consumers*, SSRN: https://ssrn.com/abstract=2706943 or http://dx.doi.org/10.2139/ssrn.2706943
- Donepudi 2019, *Automation and Machine Learning in Transforming the Financial Industry*, Asian Business Review
- Dynan, Skinner, & Zeldes, 2004, *Do the rich save more*? Journal of Political Economy
- Edwards AWF. *Likelihood.* Cambridge: Cambridge University Press; 1972.
- Emerson, Kennedy, O'Shea, & O'Brien. 2019. *Trends and applications of machine learning in quantitative finance*. Paper presented at 8th International Conference on Economics and Finance Research (ICEFR 2019), Lyon, France, June 18–21.
- FCAC, 2021, *Make Change that Counts: National Financial Literacy Strategy* 2021-2026
- Feiveson, & Sabelhaus, 2019, *Lifecycle Patterns of Saving & Wealth Accumulation*, FEDS Working Paper No. 2019-10
- Fisher 2010, *Gender Differences in Personal Saving Behaviors*, Journal of Financial Counseling & Planning, Vol. 21, No. 1
- Foerster, Linnainmaa, Melzer, Previtero, 2017, *Retail Financial Advice: Does One Size Fit All?* Journal of Finance August.
- Gale, Iwry & Orszag. et al, 2005, *The Automatic 401(K): A Simple Way to Strengthen Retirement S*aving, Tax Notes, https://www.urban.org/sites/default/files/publication/51271/1000751-The-Automatic--k-A-Simple-Way-To-Strengthen-Retirement-Saving.PDF
- Garbinsky, Klesse & Aaker, 2014, Money in the Bank: Feeling Powerful Increases Saving, Journal of Consumer Research
- Goutte, Stéphane, Amine Ismail, and Huyên Pham. 2017. "Regime-Switching Stochastic Volatility Model: Estimation and Calibration to VIX Options." *Applied Mathematical Finance* 24 (1): 38–75.

- Hall, 2021, *Promoting Savings for Financial Resilience: Expanding the Psychological Perspective*, Current Directions in Psychological Science
- Henrique Henrique, Sobreiro, & Kimura. 2019. *Literature review: Machine learning techniques applied to financial market prediction*. Expert Systems with Applications 124: 226–51.
- Hershfield et al. 2011, *Increasing saving behaviour through age-progressed renderings of future self*, doi: 10.1509/jmkr.48.SPL.S23
- Huang & Yen, 2019, *A new perspective of performance comparison among machine learning algorithms for financial distress prediction*. Applied Soft Computing 83: 105663.
- Jeffreys H. *Theory of probability*. third. Oxford: Clarendon Press; 1939/1961.
- Justera, James, Smith & Stafford, 1999, *The measurement & structure of household wealth*, Labour Economics Vol 6, Issue 2
- Kahneman 2012 *The Human Side of Decision Making: Thinking Things Through with Daniel Kahneman*, Journal of Investment Consulting 13, no. 1, (2012): 5–14
- Kempson, Finney & Poppe, 2017, *Financial Well-Being: A Conceptual Model & Preliminary Analysis*, DOI: 10.13140/RG.2.2.18737.68961
- Kim 2003, *Financial time series forecasting using support vector machines*. Neurocomputing 55: 307–19.
- Kim, Chang-Jin. 1994. "Dynamic Linear Models with Markov-Switching." *Journal of Econometric*s 60 (1-2): 1–22.
- Knight & Mandell, 1992, *Nobody gains from dollar cost averaging: analytical, numerical and empirical results*, Financial Services Review
- Kou, Peng, & Wang. 2014. *Evaluation of clustering algorithms for financial risk analysis using MCDM methods*. Information Sciences 275: 1–12
- Leo, Sharma, & Maddulety. 2019. *Machine learning in banking risk management: A literature review*. Risks 7: 29.
- Lim& Ong. 2020. *Portfolio diversification using shape-based clustering*. The Journal of Financial Data Science. doi:10.3905/jfds.2020.1.054.
- Lin, Hu, & Tsai. 2011. *Machine learning in financial crisis prediction: A survey*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42: 421–36.
- Linnainmaa, Melzer & Previtero, 2020, *The Misguided Beliefs of Financial Advisors*, The Journal of Finance (Volume 76, Issue 2).
- MacGee, Pugh & See, 2022, *The heterogeneous effects of COVID-19 on Canadian household consumption, debt and savings*, Canadian Journal of Economics
- MacInnis et al 2009, *How Do Emotions Influence Saving Behavior?*, Research Brief, Center for Retirement Research, Boston College
- Markowitz 1991, *Foundations of Portfolio Theory*, The Journal of Finance
- Markowitz 2010, *Portfolio Theory: As I Still See It*, Annual Review of Financial Economics, Vol. 2:1-23 (Volume publication date December 2010)
- Metzler, Grace & Zhou, 2021, *Learning About Financial Health in Canada*, Quantitative Finance & Economics, 5(3): 542-570. 2021.
- Merton 1973, *An intertemporal asset pricing model*., Econometrica 41, 867–887.
- Milevsky & Posner, 2003, *A Continuous-Time Reexamination of Dollar-cost Averaging*, International Journal of Theoretical and Applied Finance
- Montmarquette & Viennot-Briot, 2016, *The Gamma Factor & the Value of Financial Advice*, Cirano Allier Saviour et Decision
- Newmeyer 2020, *Is Savings Automation Helpful to Liquid Savings? It Depends on Whether You Have a Savings Habit*, Volume 40, Issue 2
- Papanicolaou, Andrew, and Ronnie Sircar. 2014. "A Regime-Switching Heston Model for VIX and s&p 500 Implied Volatilities." *Quantitative Finance* 14 (10): 1811–27.
- Patel, Shah, Thakkar, & Kotecha. 2015. *Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques*. Expert Systems with Applications 42: 259–68.
- Raffinot 2017. *Hierarchical clustering-based asset allocation*. The Journal of Portfolio Management 44: 89–99.

- Rocher, Hendrickx, & De Montjoye. 2019. *Estimating the success of re-identifications in incomplete datasets using generative models.* Nature Communications 10: 1–9
- Rundo, Trenta, Stallo, & Battiato. 2019. *Machine learning for quantitative finance applications: A survey.* Applied Sciences 9: 5574.
- Shefrin & Thaler, 2004, *Mental accounting matters*, Chapter 14, Advance in Behavioural Economics
- Soldatos & Kyriazis, 2022, *Big Data and Artificial Intelligence in Digital Finance*, Springer Open Access,
- Soman & Cheema 2011, *Earmarking & Partitioning: Increasing saving by low-income households*, Journal of Marketing Research, Vol, XLVIII
- Statman, 1995, *A Behavioural Framework for Dollar-Cost Averaging*, The Journal of Portfolio Management
- Suh 2021, *Can't save or won't save: financial resilience & discretionary retirement saving among British adults in their thirties & forties*, Published online by Cambridge University Press
- Thaler & Benartzi, 2004, *Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving*, University of California, Los Angeles
- Thompson, Feng, Reesor & Grace 2021, *Know your clients' Behaviours: A Cluster Analysis of Financial Transactions*, Journal of risk and Financial Management
- Tsai,. 2014. *Combining cluster analysis with classifier ensembles to predict financial distress*. Information Fusion 16: 46–58.
- Tufino, Maynard & De Neve, 2008, Consumer Demand for Prize-linked Savings: A preliminary analysis, Harvard Business School
- Turner et al 2009, *Homeownership, wealth accumulation & income status*, Journal of Housing Economics
- West, Dellana, & Qian. 2005. *Neural network ensemble strategies for financial decision applications*. Computers & Operations Research 32: 2543–59.
- Van Liebergen, 2017. *Machine learning: A revolution in risk management and compliance?* Journal of Financial Transformation 45: 60–67.
- Vlaev Ivo & Elliot, 2014, *Financial well-being components*, Social Indicators Research, 118(3):1103-1123.