

Samoilenko, Anton; Law, David

**Working Paper**

## KiwiSaver: Comparing Survey and Administrative Data

New Zealand Treasury Working Paper, No. 14/06

**Provided in Cooperation with:**

The Treasury, New Zealand Government

*Suggested Citation:* Samoilenko, Anton; Law, David (2014) : KiwiSaver: Comparing Survey and Administrative Data, New Zealand Treasury Working Paper, No. 14/06, ISBN 978-0-478-42120-0, New Zealand Government, The Treasury, Wellington

This Version is available at:

<https://hdl.handle.net/10419/205661>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# KiwiSaver: Comparing Survey and Administrative Data

Anton Samoilenko and David Law

---

New Zealand Treasury Working Paper 14/06



**NZ TREASURY WORKING  
PAPER 14/06**

KiwiSaver: Comparing Survey and Administrative Data

**MONTH/YEAR**

February 2014

**AUTHORS**

Anton Samoilenko  
New Zealand Treasury  
Email           anton.samoilenko@treasury.govt.nz  
Telephone       +64 4 471-5084

David Law  
New Zealand Treasury  
Email           david.law@treasury.govt.nz  
Telephone       +64 4 471-5929  
Fax              +64 4 471-5191

**ISBN (ONLINE)**

978-0-478-42120-0

**URL**

Treasury website at February 2014:  
<http://www.treasury.govt.nz/publications/research-policy/wp/2014/14-06>  
Persistent URL: <http://purl.oclc.org/nzt/p-1635>

**ACKNOWLEDGEMENTS**

Authors would like to acknowledge the helpful comments from Katherine Meerman, Malcolm Menzies, Kristie Carter, Barb Lash and Ian McGregor.

**NZ TREASURY**

New Zealand Treasury  
PO Box 3724  
Wellington 6008  
NEW ZEALAND  
Email           information@treasury.govt.nz  
Telephone       64-4-472 2733  
Website         www.treasury.govt.nz

**DISCLAIMER**

The views, opinions, findings, and conclusions or recommendations expressed in this Working Paper are strictly those of the author(s). They do not necessarily reflect the views of the New Zealand Treasury or the New Zealand Government. The New Zealand Treasury and the New Zealand Government take no responsibility for any errors or omissions in, or for the correctness of, the information contained in these working papers. The paper is presented not as policy, but with a view to inform and stimulate wider debate.

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975.

The results presented in this study are the work of the author, not Statistics New Zealand.

# Abstract

---

This paper explores the KiwiSaver information contained in two sources: the administrative data from the Inland Revenue Department (IRD) and the Survey of Family, Income and Employment (SoFIE). In particular, the paper explores the membership and contribution information, explaining significant patterns and highlighting any differences that exist between the two data sources. At the aggregate level, the paper shows noticeable difference in membership levels, tenure, annual employer and employee contributions and total cumulative contributions. At the individual level, regression results identify the main determinants of the observed differences and quantify their impact.

**JEL CLASSIFICATION**

E21: Consumption; Saving; Wealth  
D91: Intertemporal Consumer Choice; Life Cycle Models and Saving

**KEYWORDS**

KiwiSaver; Saving; New Zealand; Administrative data; Survey Data

## Executive Summary

---

The purpose of this paper is to compare KiwiSaver information that was collected by the Survey of Family, Income and Employment (SoFIE) to the individual's IRD administrative records for the period that covers October 2007 to September 2010 (SoFIE waves 6 to 8).

The paper begins by describing the challenges that exist in trying to align survey and administrative data. In particular, the paper notes the difficulty of creating measures of contributions that cover a comparable period of time. SoFIE collects information that covers individual's interview period, which is often not aligned with a particular tax year. At the same time, administrative data are compiled on quarterly or yearly basis, creating issues in adjusting administrative data to match the survey interview period.

In addition, major differences exist in the information provided by respondents and the information contained in the administrative data. More specifically, SoFIE appears to underreport the number of KiwiSaver members relative to administrative data over the period of 2009 to 2010. This is particularly apparent in the difference in the number of automatically enrolled members. The paper finds that SoFIE underreports the number of automatically enrolled members by around 50% of the administrative measure of membership during that period.

Moreover, individuals tend to report KiwiSaver enrolment dates that do not match their administratively recorded enrolment dates. In most cases, individuals appear to report an enrolment date that precedes their administrative enrolment date by up to 3 months.

At the aggregate level, these differences in membership and enrolment dates result in SoFIE reporting longer average tenure in KiwiSaver and higher values for annual member, employer and total cumulative contributions relative to the administrative data.

At the individual level, regression results suggest that SoFIE underestimates the values of annual member and total cumulative contributions relative to administrative data at the lower end of the contributions distribution, while overestimate the value of these measures at the higher end of the contributions distribution. This pattern is observed over all waves for annual member contributions and total cumulative contributions.

Similar pattern is observed for annual employer contributions in wave 8. However, in wave 6 and 7, SoFIE underestimates the value of annual employer contributions relative to administrative data across virtually all levels of contributions. This pattern can be potentially explained by the fact that a number of employers contributed above the mandatory minimum rate of 1% during that period.

The paper concludes by noting that even though there are some challenges presented by the use of the administratively linked survey data, the addition of administrative data to survey data provides extra precision and choice in the KiwiSaver related variables. These advantages could potentially lead to higher quality evaluation of the effects of KiwiSaver on savings than what would have been possible if the evaluation relied exclusively on survey data.

# Table of Contents

---

<b>Abstract</b> .....	<b>i</b>
<b>Executive Summary</b> .....	<b>ii</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>2 Data</b> .....	<b>2</b>
2.1 Survey of Family, Income and Employment (SoFIE) .....	2
2.2 IRD administrative data .....	3
2.3 Sample 4 .....	
<b>3 IRD administrative dataset</b> .....	<b>4</b>
3.1 KiwiSaver membership patterns .....	4
3.2 KiwiSaver membership tenure .....	6
3.3 Annual member contributions and rates .....	7
3.4 Annual employer contributions and rates .....	10
3.5 Member Tax Credit .....	12
3.6 Extra contributions via a provider .....	13
3.7 Total cumulative contributions to KiwiSaver .....	14
<b>4 SoFIE dataset</b> .....	<b>15</b>
4.1 SoFIE KiwiSaver membership patterns .....	16
4.2 SoFIE KiwiSaver membership tenure .....	18
4.3 Annual member contributions and rates .....	19
4.4 Annual employer contributions .....	21
4.5 Total cumulative contributions to KiwiSaver .....	22
<b>5 Key points of difference</b> .....	<b>23</b>
5.1 KiwiSaver membership .....	23
5.2 KiwiSaver membership tenure .....	23
5.3 KiwiSaver member contribution rates .....	24
<b>6 Analysis of differences</b> .....	<b>25</b>
6.1 KiwiSaver membership .....	25
6.2 Enrolment dates .....	26
6.3 Regression results .....	27
6.4 Main determinants .....	34
<b>7 Discussion and Conclusion</b> .....	<b>36</b>
<b>References</b> .....	<b>39</b>

## List of Tables

---

Table 1 – Membership levels across waves .....	5
Table 2 – Members by type of enrolment .....	5
Table 3 – Contributing members .....	6
Table 4 – Summary of membership tenure by type of enrolment (days) .....	7
Table 5 – Summary of annual member contributions .....	8
Table 6 – Summary of annual employer contributions .....	10
Table 7 – Proportion of members receiving full value of the MTC .....	12
Table 8 – Cumulative value of contributions: summary statistics .....	14
Table 9 – Membership levels across waves (SoFIE) .....	16
Table 10 – Members by type of enrolment (SoFIE) .....	16
Table 11 – Contributing members (SoFIE) .....	17
Table 12 – Summary of membership tenure (in days) by type of enrolment (SoFIE) .....	18

Table 13 – Summary of annual member contributions (SoFIE).....	20
Table 14 – Summary of annual employer contributions (SoFIE) .....	21
Table 15 – Cumulative value of KiwiSaver contributions: summary statistics (SoFIE).....	22
Table 16 – Summary table: KiwiSaver membership and automatically enrolled members.....	23
Table 17 – Summary table: KiwiSaver membership tenure for automatically enrolled members .....	24
Table 18 – Differences in membership .....	25
Table 19 – KiwiSaver members who don't appear in SoFIE by wave 8, by enrolment type .....	26
Table 20 – Regression results (total cumulative contributions) .....	28
Table 21 – Regressions results (annual member contributions) .....	30
Table 22 – Regression results (annual employer contributions).....	32
Table 23 – Main Determinants: summary of regression variables.....	34
Table 24 – Regressions on the difference in total cumulative, member and employer contributions.....	35

## List of Figures

---

Figure 1 – Membership tenure (in days) by type of enrolment .....	6
Figure 2 – Distribution of annual member contributions .....	8
Figure 3 – Annual member contribution rates.....	9
Figure 4 – Distribution of annual employer contributions.....	10
Figure 5 – Annual employer contribution rates .....	11
Figure 6 – Member Tax Credit distribution.....	12
Figure 7 – Distribution of extra contributions to KiwiSaver via a provider.....	13
Figure 8 – Distribution of combined member and extra contributions .....	13
Figure 9 – Distribution cumulative value of KiwiSaver contributions .....	14
Figure 10 – Membership tenure (in days) by type of enrolment (SoFIE).....	18
Figure 11 – Distribution of annual member contributions (SoFIE).....	19
Figure 12 – Annual member contribution rates (SoFIE) .....	20
Figure 13 – Distribution of annual employer contributions (SoFIE) .....	21
Figure 14 – Distribution of cumulative value of KiwiSaver contributions (SoFIE).....	22
Figure 15 – Difference between the IRD and SoFIE enrolment dates in days .....	26
Figure 16 – Total cumulative contributions: regression graphs .....	29
Figure 17 – Annual member contributions: regression graphs .....	31
Figure 18 – Annual employer contributions: regression graphs .....	32

# KiwiSaver: Comparing Survey and Administrative Data

## 1 Introduction

---

The purpose of the working paper is to compare KiwiSaver data that were collected by the Survey of Family, Income and Employment (SoFIE) to the individual's IRD administrative KiwiSaver records for the period that covers October 2007 to September 2010 (SoFIE waves 6 to 8)<sup>1</sup>.

Comparing these two datasets serves a number of objectives. Initially, the paper describes the variables contained in both datasets and summarises any notable patterns within these variables. It highlights the differences that exist between the two datasets, which could shed some light into how members view their KiwiSaver. By exploring the differences between survey and administrative data, this paper also highlights the possible implications the exclusive reliance on SoFIE data would have had on the results from the evaluation of the impact of KiwiSaver on savings.

Differences in survey and administrative data can occur for a number of reasons. Errors in survey data can arise due to respondent error from erroneous recollection of specific details or through failure to understand the nature of the question. Errors can also be a product of issues related to timing, either of survey itself or through the process of calculating annual amounts from collected data.

In contrast, while administrative data do not suffer from such issues, the data could be affected by reporting lags, inexact matching of survey and administrative data, limited observation window or limited coverage of variables.

A quick scan of the international literature shows that most studies that examine the differences between survey and administrative data concentrate on labour market related variables, namely income. Most recent literature notes that survey data tend to overestimate income at the lower end of the income distribution, while underestimating income at the higher end of the income distribution when compared to administrative records (for an example of such studies see (Kim and Tamborini, C. (2009))). These papers also find that the difference increases with the level of income.

---

<sup>1</sup> The introduction of KiwiSaver actually coincided with the tail-end of SoFIE wave 5. However, by the time KiwiSaver was introduced, many of the respondents would have already been interviewed. This means that there are only a small number of cases where individuals would have been eligible for KiwiSaver prior to their interview in wave 5 and for the vast majority of members the introduction of KiwiSaver would have occurred in wave 6. Due to the small sample in wave 5, results for this wave were omitted.



These results suggest that if all other factors were held equal, SoFIE would overestimate the contributions to KiwiSaver at the lower end of the income distribution, while also underestimating the contributions at the higher end. Depending on the evaluation techniques used to study the effects of KiwiSaver on saving, this overestimation of income at the lower end and underestimation at the higher end can have an adverse effect on the robustness of the results. In view of these, the paper aims to showcase the advantages that exist in individually linking survey data to administrative records.

Section 2 begins by providing a short overview of the SoFIE and the IRD datasets. Section 3 describes the IRD dataset in detail, noting any specific patterns in the data. Section 4 describes the variables that were derived using the information contained in the SoFIE dataset and explores any differences that can be observed between the SoFIE and the IRD variables at the aggregate level. Section 5 compares the SoFIE and the IRD variables at the individual level using a combination of descriptive and simple econometric techniques. Section 6 concludes with a broad discussion of the findings.

## 2 Data

---

### 2.1 Survey of Family, Income and Employment (SoFIE)

The Survey of Family, Income and Employment (SoFIE) is a longitudinal survey that was run by Statistics New Zealand. Respondents were interviewed once a year over 8 years (waves), covering the period from October 2002 to end of September 2010. At each interview, individuals were asked to answer questions on income, family and household as well as questions on their involvement in the labour market. Detailed socio-demographic data were also collected in each wave. In waves 2, 4, 6 and 8 the questionnaire included questions on individual's assets and liabilities. Questionnaire for waves 3, 5 and 7 included questions on individual's mental and physical health.

Most importantly, in wave 8 of SoFIE the questionnaire included specific questions about KiwiSaver. Two broad categories of questions were included in the survey: questions on individuals' membership status and questions on their contributions.

Questions related to KiwiSaver membership included:

- Whether the individual is a KiwiSaver member.
- If not a members, the individual is asked to provide a reason why not.
- If a member, the individual is asked how they first joined KiwiSaver (automatically enrolled or through an employer or a provider).
- The individual is then asked to provide the year and month in which they joined the Scheme.

Questions related to contributions to KiwiSaver included:

- If a member, the individual is asked to indicate whether they are currently contributing to KiwiSaver.
- If not contributing, they are asked to provide a reason why they are not contributing (contribution holiday or other reason).
- If they are on a contribution holiday, they are then asked to indicate how long their contribution holiday is.
- Finally, the individual is asked to indicate their contribution rate and whether their employer is currently paying into their KiwiSaver account.

There are two major elements of KiwiSaver that were not included in the SoFIE questionnaire. First, the questionnaire did not ask whether the individual made any extra contributions to KiwiSaver since they joined the Scheme. In addition, the questionnaire did not include any information on employer contribution rates.

Regardless of these limitations, SoFIE presents a good collection of information that not only describes how individuals contribute to KiwiSaver, but provides some insight into how individuals think about KiwiSaver.

## 2.2 IRD administrative data

In 2012, Statistics New Zealand undertook a project to link SoFIE to the IRD administrative KiwiSaver data. The IRD dataset covers the period of 16 quarters from 1 July 2007 to 30 June 2011. For each quarter, the dataset contains information on individuals' taxable income, main source of income and number of jobs held. In relation to KiwiSaver, the dataset contains detailed information on the individuals' current KiwiSaver membership including:

- Enrolment date
- Type of enrolment and current membership status
- Member contributions (in dollars) for each quarter
- Employer contributions (in dollars) for each quarter
- Member's contribution rate for each quarter
- Employer's contribution rate for each quarter
- Government's contributions for the year

Altogether, the IRD administrative dataset provides a comprehensive record of KiwiSaver membership and contributions across time.

## 2.3 Sample

The paper uses a sample selected from SoFIE respondents who had the following characteristics:

- 1) They were eligible and responding individuals in wave 1 of the survey. This means that they either personally answered a Personal Questionnaire in wave 1, or were a child of an eligible and responding adult.
- 2) They have a complete longitudinal history. This means they provided a complete response in all 8 waves of the survey. SoFIE respondents were initially selected using a stratified and clustered sampling methodology. Subsequently, sample weights were created to account for the uneven probability of selection and attrition across waves. Using respondents with a complete longitudinal history allows the sample to be weighted using the appropriate weights for each wave.
- 3) They were between the ages of 15 and 62 in wave 2 (wave 2 ran from October 2003 to September 2004). This ensures that they were eligible to be part of KiwiSaver when the scheme was first introduced in 2007, which coincides with the end of SoFIE wave 5.

Individuals who satisfied these conditions were selected from SoFIE and their administrative records and SoFIE KiwiSaver related responses form the basis of the subsequent analysis. This ensures that the analysis of administrative and SoFIE data is conducted using the same set of individuals, eliminating any potential differences that might exist between the populations covered by the two datasets.

Survey weights were applied to the sample to produce the population totals.

## 3 IRD administrative dataset

---

This section presents the KiwiSaver data from the IRD administrative dataset. While all of the data presented in this section were taken from the administrative dataset, the data were not always originally organised in the way it is presented in this section. Some variables had to be recreated using administrative data in order to better match the interview period covered by SoFIE to enable effective comparison. In other instances, a number of potentially important variables were not part of the original administrative dataset and had to be created from the available information.

In the majority of cases, the use of administrative data ensured that the created variables align closely to the period they cover. In a small minority of cases, reporting lags within the administrative data resulted in some mismatch between the period covered by the variable and the period covered by one of the elements used during the creation process.

### 3.1 KiwiSaver membership patterns

Tables 1-3 summarise membership and contributions patterns across waves. Originally, administrative data only contained records of individuals' current membership enrolment date and membership type. This information, combined with the individual's interview date, was used to create a membership indicator for each wave and the individual's record of contributions was used to determine the proportion of members who contributed in a given wave.

Table 1 shows consistent growth in the membership level within the sample across time, with membership reaching 37.6% of the sample by wave 8. This level of membership is consistent with the level of membership observed in other papers (Law, Meehan and Scobie (2011)) and suggests that SoFIE-IRD linked dataset does not fundamentally underestimate or overestimates KiwiSaver membership.

**Table 1 – Membership levels across waves**

<i>KiwiSaver membership</i>		<i>Wave 6</i>	<i>Wave 7</i>	<i>Wave 8</i>
Not a Member	<i>Count</i>	1,718,600	1,507,500	1,387,400
	<i>Percent</i>	82.3%	70.3%	62.4%
Member	<i>Count</i>	368,800	636,600	836,600
	<i>Percent</i>	17.7%	29.7%	37.6%
Total		2,087,400	2,144,000	2,224,000

Table 2 shows a large proportion of members who joined KiwiSaver through their employer or directly through a provider in wave 6, with roughly equal proportions choosing to actively enrol via either of the two types of enrolment. Table 2 also shows the changing composition of actively enrolled members by type of enrolment over time. In particular, the table shows that the proportion of individuals who enrolled via their employer decreases substantially in waves 7 and 8. In contrast, the proportion of members who actively chose their provider grows, which can indicate that new members prefer to make their own decision about their KiwiSaver provider instead of enrolling with their employer elected provider.

**Table 2 – Members by type of enrolment**

<i>Enrolment Type</i>		<i>Wave 6</i>	<i>Wave 7</i>	<i>Wave 8</i>
Opt-in via Employer	<i>Count</i>	131,900	176,500	199,100
	<i>Percent</i>	35.8%	27.7%	23.8%
Actively chose provider	<i>Count</i>	130,400	237,700	331,400
	<i>Percent</i>	35.4%	37.3%	39.6%
Automatically enrolled	<i>Count</i>	106,600	222,400	306,100
	<i>Percent</i>	28.9%	34.9%	36.6%
Total (Members)		368,800	636,600	836,600

Table 2 also shows a consistent growth in the proportion of KiwiSaver members who automatically enrolled. This highlights the growing importance of automatic enrolment as a way for new members to join the Scheme.

In Table 3, the numbers of contributing and non-contributing members can be partially explained by the fact that some individuals would not have had a full quarter of contributions between their enrolment date and the date of their interview. Hence, when determining whether they contributed during wave 5 or in any of the following waves, difference in the timing of the interview date and the timing of the end of the first contribution quarter means that the person is recorded as a member for a particular wave, but is not considered as having contributed.

**Table 3 – Contributing members**

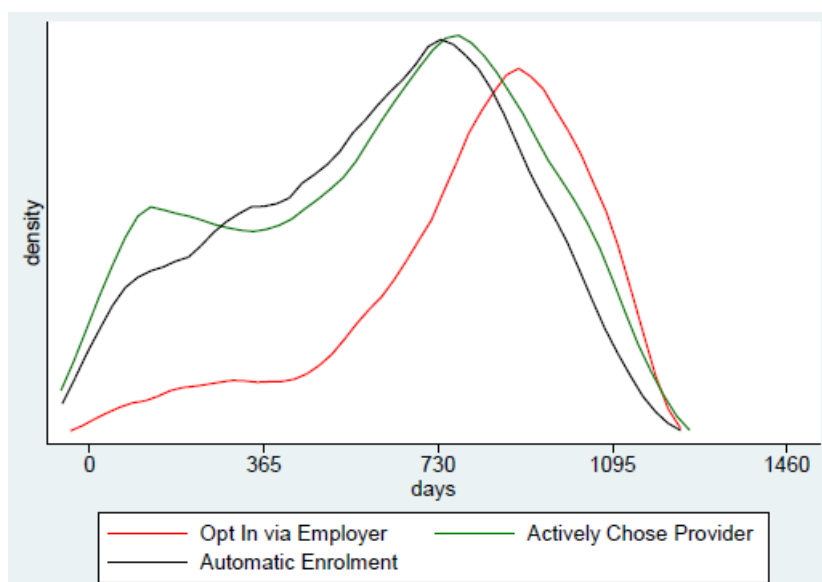
Contribution Status of KiwiSaver Members across waves (Admin Data)		Wave 6	Wave 7	Wave 8
Contributing	Count	332,600	578,200	724,500
	Percent	90.2%	90.8%	86.6%
Contribution Holiday	Count	4900	17500	22800
	Percent	1.3%	2.7%	2.7%
Not Contributing	Count	31,300	40,900	89,300
	Percent	8.5%	6.4%	10.7%
KiwiSaver Members (Total)		368,800	636,600	836,600

Moreover, the increase in the number of non-contributing members could be due to an increase in the unemployment rate following the Global Financial Crisis (GFC). It is also possible that some of the self-employed members would have stopped contributing due to a reduction in their income following the GFC.

## 3.2 KiwiSaver membership tenure

Administratively recorded enrolment date was used to create measures of KiwiSaver membership tenure in each wave. This variable indicates the number of days between the individual's enrolment date and the date of their interview. Figure 1 shows the distribution of tenure across types of enrolment at the end of wave 8, while Table 4 summarises the distribution of tenure for each wave.

**Figure 1<sup>2</sup> – Membership tenure (in days) by type of enrolment**



<sup>2</sup> Figure 1 shows the kernel density of membership tenure by type of KiwiSaver enrolment. The advantage of using kernel density is that it provides a high-resolution representation of the distribution of the variable without the need to aggregate the data into discrete intervals. This should better capture the peaks and troughs within the distribution, particularly if there are high concentrations of individuals around specific values. It also allows the figure to show long tails of the distributions for the cases where long tails occur. The main drawback of using the kernel density is that the scale of the density estimate is not easily interpreted, which is why it was decided to remove any ticks from the density scale to avoid unnecessary confusions. Kernel density is used throughout the paper.

**Table 4 – Summary of membership tenure by type of enrolment (days)**

Wave	Type of Enrolment	Lower Quartile	Median	Upper Quartile	Mean
6	Opted in via employer	109	201	293	203
	Actively chose provider	72	156	262	175
	Automatically enrolled	67	138	230	153
	Average (all types)	83	168	264	179
7	Opted in via employer	379	503	623	486
	Actively chose provider	244	396	538	392
	Automatically enrolled	222	366	496	362
	Average (all types)	265	417	554	408
8	Opted in via employer	671	848	971	793
	Actively chose provider	348	658	847	603
	Automatically enrolled	370	635	807	594
	Average (all types)	421	699	879	645

Tenure and enrolment patterns across waves show a strong initial uptake of KiwiSaver via employers. This can be seen both in the large proportion of members in wave 6 who opted in via their employer and in the longer average tenure during that wave for this group of members. Furthermore, tenure patterns across waves show that members who opted in via an employer have consistently longer tenure, indicating a low number of new members who chose to join KiwiSaver via this method of enrolment in later waves.

Perhaps not surprisingly, individuals who automatically enrolled have the shortest average tenure in KiwiSaver across all waves. This is largely due to the slower than average uptake of the Scheme among members who were automatically enrolled and due to the steady inflow of new KiwiSaver members via this type of enrolment.

The observed decrease in the number of automatically enrolled KiwiSaver members over the course of waves 7 and 8 can be at least partially attributed to the difficult labour market environment at the time. This particular period is characterised by high unemployment (particularly among young adults) as well as a lower number of individuals who change their jobs. Altogether, these factors are likely to have contributed to the observed patterns in uptake and tenure by members who automatically enrolled in the Scheme.

Individuals who actively chose a provider exhibit an interesting tenure pattern across waves. In the first instance, the enrolment patterns show a large number of individuals who joined KiwiSaver by choosing a provider shortly after the Scheme was introduced. Moreover, the enrolment and tenure patterns of this group suggest that this method of enrolment is consistently favoured by individuals joining the Scheme.

Further, Figure 1 shows a large concentration of members with low tenure who actively chose a provider. This pattern could be indicative of the improving economic conditions in wave 8, which would enable new members to join KiwiSaver via a provider, especially if they are self-employed.

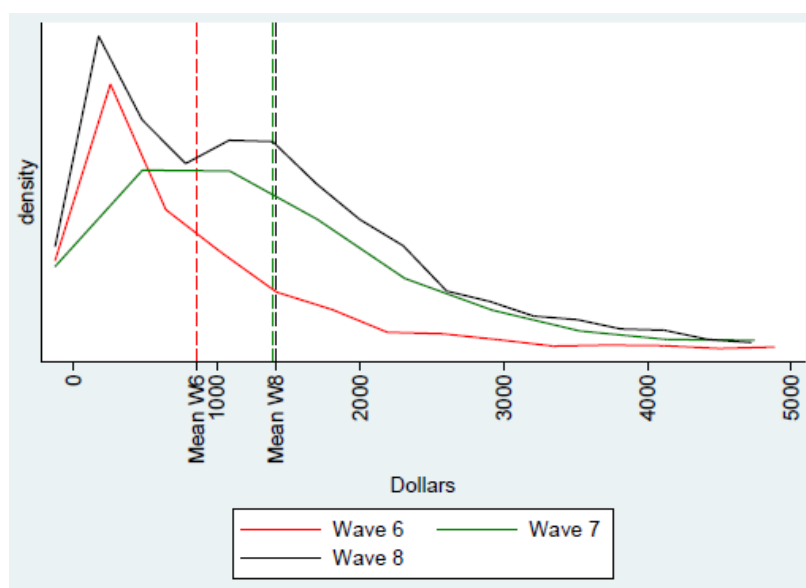
### 3.3 Annual member contributions and rates

Figure 2 shows the distribution and Table 5 provides summary statistics for annual member contributions. Before describing the patterns in the annual member contributions,

it is important to note the difficulty of aligning KiwiSaver administrative data with the individuals' interview period. In the derivation of annual member contributions using administrative data, quarterly contributions to KiwiSaver were used to create a measure that, in most cases, closely aligned to the individuals' interview period.

However, since individuals were not interviewed at the end of a particular quarter, the value for the quarter in which the individual was interviewed was not used to derive the amount of annual member contributions for the interview period. Instead, the value for that quarter was used in the calculation of annual member contributions in the subsequent interview period. Altogether, it is hoped that this treatment of the data results in only a small underestimation of annual member and employer contributions as well as total cumulative contributions for any given period.

**Figure 2 – Distribution of annual member contributions**



**Table 5 – Summary of annual member contributions**

Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$170	\$525	\$1,133	\$859
7	\$418	\$1,097	\$1,847	\$1,392
8	\$415	\$1,146	\$1,932	\$1,409

Figure 2 shows a number of features in the distribution of annual member contributions across waves. In waves 6, the shape of the distribution is likely to be affected by the fact that many KiwiSaver members would have only just joined the Scheme and would have had low membership tenure. These members would not have had the time to contribute for an entire year, which could explain the large number of observations below \$1,000.

By wave 7, the distribution of annual member contributions changes considerably. With a larger group of individuals who have been part of KiwiSaver for at least one year, there is a noticeable increase in the number of individuals who contribute \$1,042 or more. This change is to be expected, since anyone earning more than the minimum wage at the time would have contributed over \$1,000 to KiwiSaver over the course of a full year.

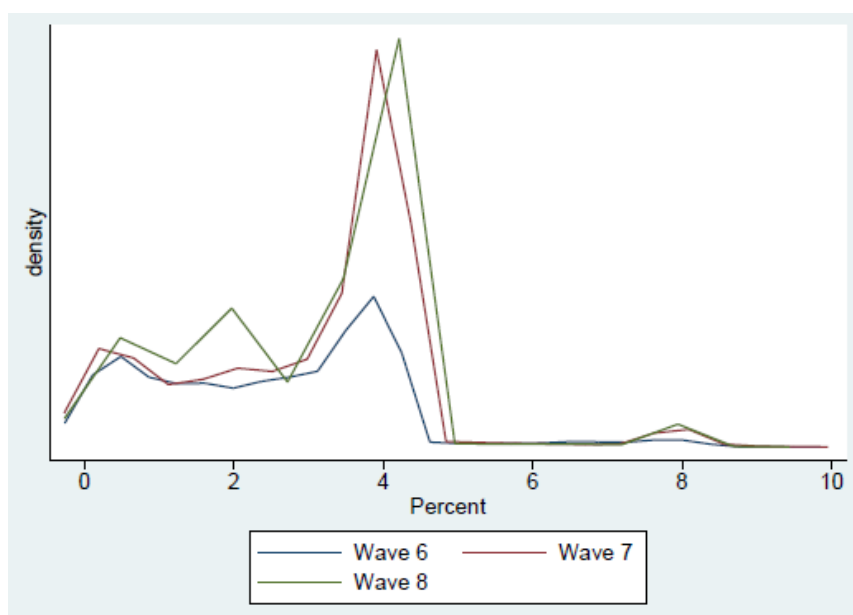
The change observed in wave 8 is possibly due to the change in the minimum member contribution rate from 4% to 2% that was introduced on the 1<sup>st</sup> of April 2009. It is possible that individuals whose 2% yearly contributions are around \$1,042 per year would have



switched from contributing 4% to 2% of their pay. In addition, some of the new members would have also had the opportunity to choose to start contributing at a lower rate. This means that any new members who did not earn more than \$52,000 per year and contributed 2% of their salary or wages would have contributed less than \$1,042 to KiwiSaver. Altogether, this change in the minimum contribution rate potentially increased the number of observations at the lower end of contributions.

Member contribution rates presented in Figure 3 differ somewhat from the contribution rates reported in the administrative data supplied by the IRD. The contribution rate recorded in the administrative data shows the rate of contributions selected by the individual, but does not necessarily reflect actual rate of contributions for any given quarter.

**Figure 3 – Annual member contribution rates**



For example, consider the case where the administrative record indicates that the individual chose an 8% contribution rate for a particular quarter. However, the amount contributed shows that the individual contributed less than what would be expected at the indicated rate. This can occur in cases where individuals did not contribute for the entire quarter, or increase their contribution rate midway through the quarter.

In both scenarios, a more accurate way of reporting the contribution rate would be to calculate a contribution rate based on the income earned and contributions made to KiwiSaver by the individual. Quarterly contribution rates are then averaged over the interview period to produce the measure of the member contribution rate for that year. The resulting contribution rate is often slightly higher or lower than the indicated rate for a particular quarter.

Figure 3 presents the annual contribution rates to KiwiSaver across waves and shows that the 4% contribution rate was the preferred rate for all waves, with only a small number of individuals contributing at 8% before wave 8 and a small fraction contributing at 2% in wave 8 when the minimum contribution rate changed. This result could indicate a certain degree of investor inertia by KiwiSaver members. At the same time, it can be a product of active choice by individuals who might want to contribute more than the mandatory minimum to receive the full value of the Member Tax Credit at the end of the

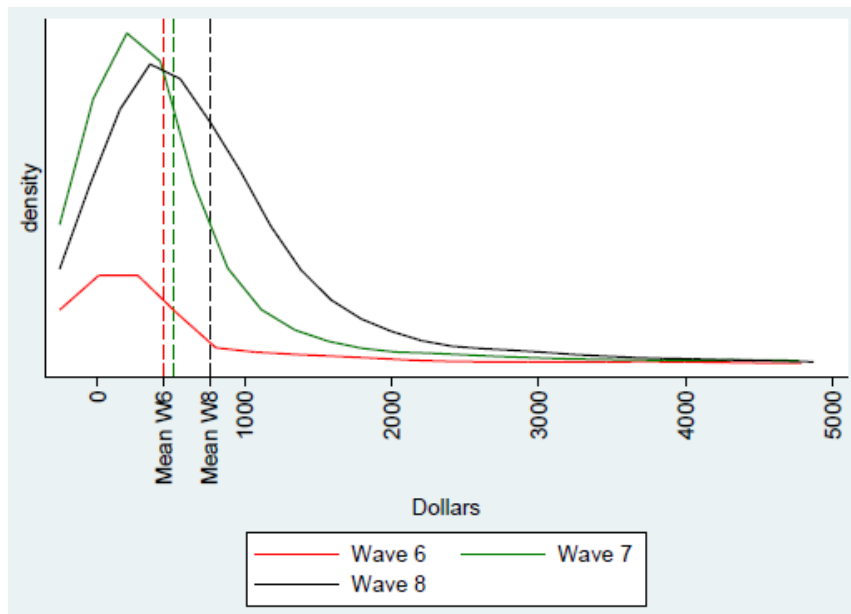


year. Further analysis might be required in order to evaluate the determinants of the observed contributions patterns.

### 3.4 Annual employer contributions and rates

Figure 4 shows the distribution of annual employer contributions across waves and Table 6 provides summary statistics. Similarly to the distribution of annual member contributions, the distribution of annual employer contributions is affected by two factors: the number of new members joining the scheme during a particular wave and the rate at which the employers contribute.

**Figure 4 – Distribution of annual employer contributions**



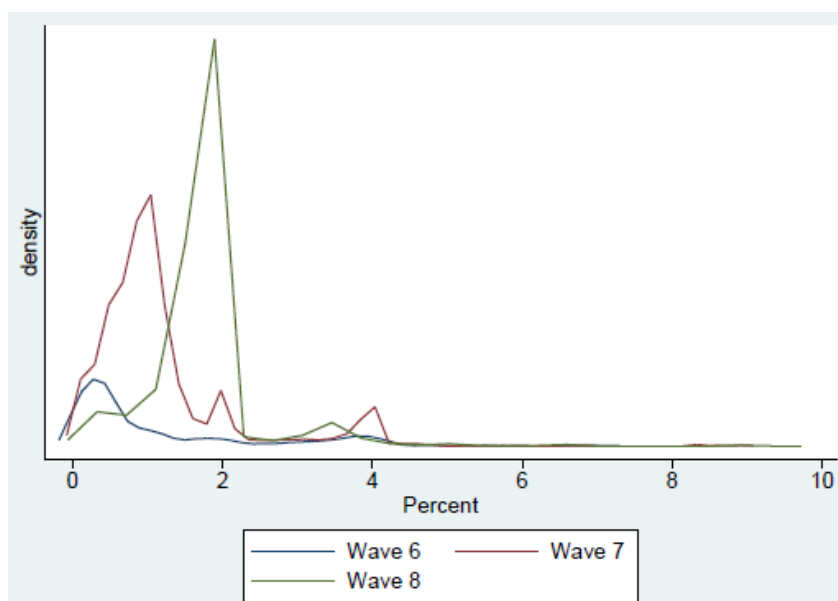
**Table 6 – Summary of annual employer contributions**

Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$58	\$112	\$333	\$451
7	\$89	\$257	\$536	\$518
8	\$218	\$585	\$997	\$760

New members would have contributed less to KiwiSaver by the time they were interviewed based solely on the fact that new members didn't contribute for a full year. Consequently, this means that the corresponding employer contributions for new members are lower than they would be if these members contributed for the entire year. This potentially explains the large number of observations at the lower end of the distribution, particularly in wave 6.

The rate at which employers contribute can also have a strong affect on the distribution of annual employer contributions, particularly when comparing waves 7 and 8. In wave 7, the majority of employers contributed around 1% of the employee's salary or wages, with a small group of employers contributing at 2% and 4%. By wave 8, this pattern changes with virtually all employers contributing at 2%. This shift in the contribution rate is likely to have made the largest impact on the distribution, resulting in an almost universal shift in the distribution towards higher contribution levels.

**Figure 5 – Annual employer contribution rates**



Looking specifically at the pattern of annual employer contribution rates, Wave 7 offers the most interesting result and shows a distinct group of employers who were contributing at 2% or 4%. The wide dispersion around the 1% rate indicates that employers contributed at a higher than mandatory minimum rate of 1% at least for a part of wave 7. These results are partially consistent with employers responding to the tax incentives that were part of the Scheme at that time. The Employer Superannuation Contributions Tax (ESCT) exemption applied to employer contributions up to 4% and the Employer Tax Credit reimbursed employers up to a maximum of \$20 per week for the contributions they made to their employee's KiwiSaver.

The picture changes completely in wave 8, with the vast majority of employers contributing at the mandatory minimum of 2%. This is most likely due to the following changes to KiwiSaver:

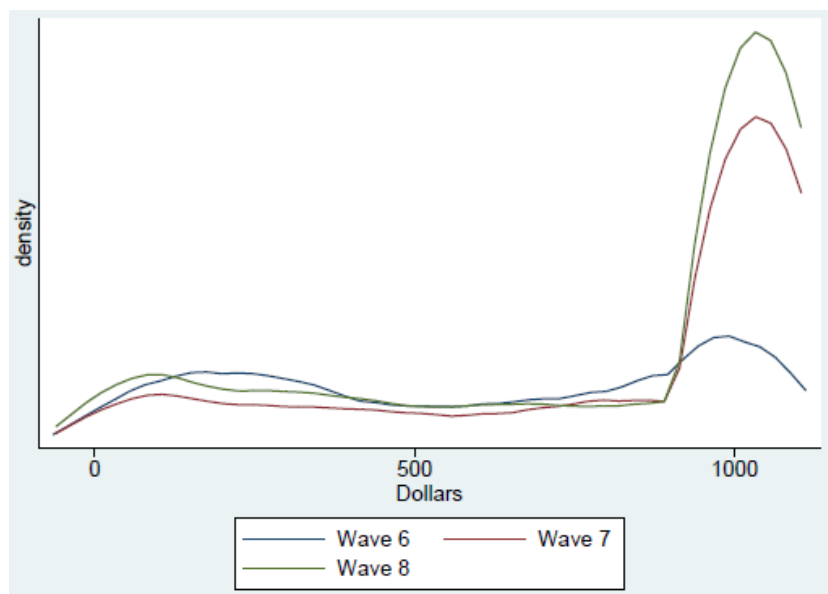
- Minimum contribution rate was increased from 1% to 2%
- Employer Tax Credit was discontinued
- Employer Superannuation Contributions Tax exemption was limited from 4% to 2% of the employees' salary or wages

Altogether, these changes to the Scheme would have removed any incentives for employers to contribute above the minimum contribution rate.

### 3.5 Member Tax Credit

Figure 6 shows that the number of KiwiSaver members receiving full or close to full value of the Member Tax Credit (MTC) increases from wave to wave. Table 7 shows that in wave 6, the proportion of members receiving the full value of the Member Tax Credit is substantially smaller due to a large number of new members who would not have contributed for an entire year. In contrast, in waves 7 and 8 a larger proportion of members would have contributed for an entire year or made enough extra contributions to qualify for the full value of the Member Tax Credit.

**Figure 6 – Member Tax Credit distribution**



**Table 7 – Proportion of members receiving full value of the MTC**

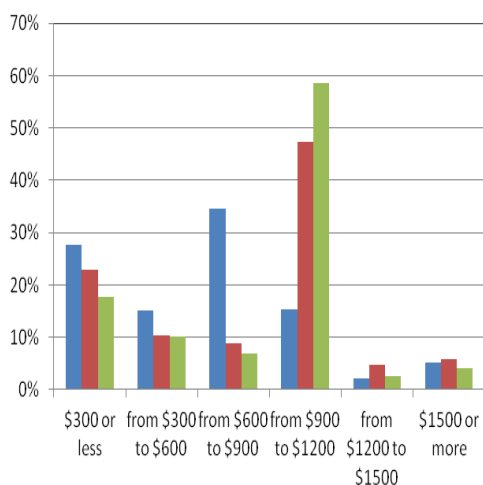
<i>Member Tax Credit received by the KiwiSaver member</i>	<i>wave 6</i>	<i>wave 7</i>	<i>wave 8</i>
Received less than \$1042 from MTC	83.4%	51.4%	53.6%
Received at least \$1042 from MTC	16.5%	48.6%	46.4%
Total	100%	100%	100%

The small decrease in the proportion of members receiving at least \$1042 from the Member Tax Credit between waves 7 and 8 is likely due to a variety of factors. The most plausible explanation could be in the large number of new members joining the Scheme in wave 8. Some of the members would have started contributing at the minimum rate of 2%, while others would only have been part of the Scheme for a very short time, leading to some of them not making enough contributions to qualify for the full amount of the Member Tax Credit.

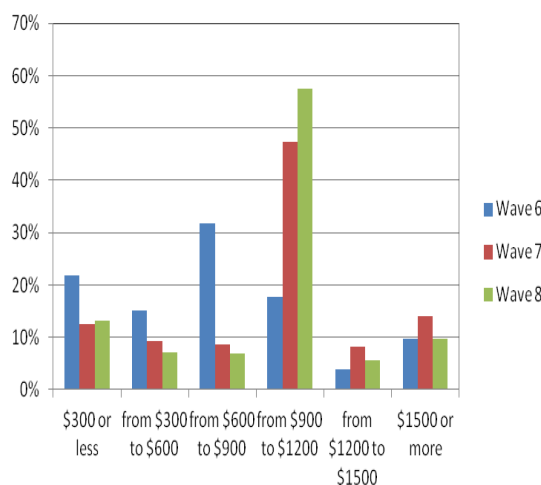
### 3.6 Extra contributions via a provider

Figure 7 shows the distribution of extra contributions that were made directly to the providers and Figure 8 shows the combined value of annual member contributions and extra contributions. These tables are produced only for those individuals who made extra contributions, since not all members made contributions above their mandatory minimum within a given wave.

**Figure 7 – Distribution of extra contributions to KiwiSaver via a provider**



**Figure 8 – Distribution of combined member and extra contributions**



In the administrative data, the information on the extra contributions individuals make directly to KiwiSaver providers is reported only once a year and is based on extra contributions for the year ending on 30 June. This reporting feature of the data can make it difficult to examine the instances where individuals make extra contributions within a particular interview period, because the interview period and the year ending on 30 June almost never match with precision.

This mismatch in the coverage period can also create problems when looking at the patterns of combined annual member contributions and extra contributions. If the two period matched completely, it would be possible to identify cases where individuals made combined contributions of exactly \$1042.86 per year to maximise the value of the Member Tax Credit they receive from the Government. The mismatch in the timing means that the combined value of annual member and extra contributions is likely to be either underestimated, or overestimated depending on when the individual was interviewed.

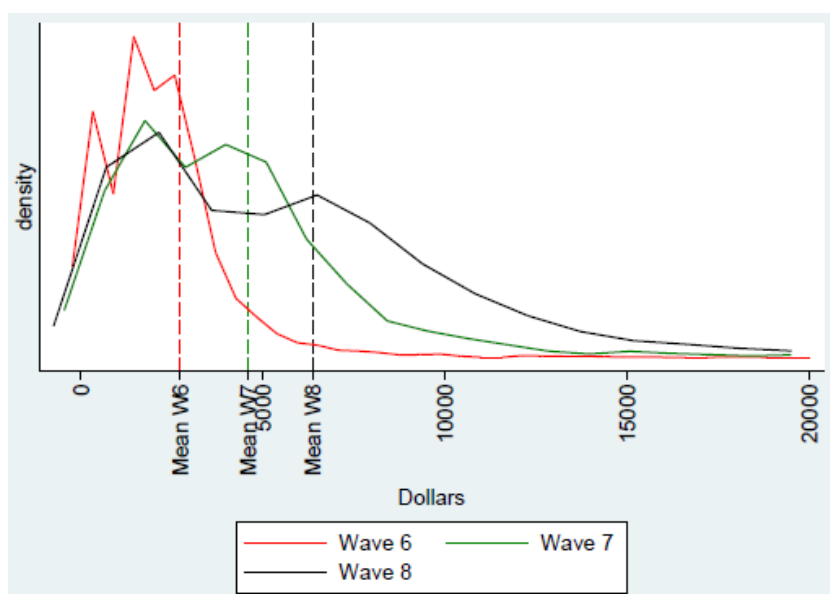
In view of these limitations, the patterns across waves show an increase in the proportion of individuals who make extra contributions of between \$900 and \$1,200 per year. This can indicate that as time passes, individuals adjust their behaviour to take full advantage of Government's incentives. Further, the two figures show that many of the individuals who make extra contribution either have very small or no member contributions and use extra contributions as their main way of contributing to KiwiSaver. This observation is reinforced by the fact that a large proportion of individuals who make extra contributions also have investment or business income as their main source of income.

### 3.7 Total cumulative contributions to KiwiSaver

Figure 9 and Table 8 show the distribution of total cumulative contributions towards KiwiSaver. This includes the combined value of annual member contributions, annual employer contributions, the Member Tax Credit and the KickStart payment as well as any extra contributions individuals paid directly through a provider from the time they enrolled to their current interview date in a particular wave.

The data do not contain any information on returns on past contributions. As the result, this measure only shows the net cumulative inflows into individuals' KiwiSaver accounts over time. Moreover, as a cumulative measure of all contributions this measure is affected by the various trends and behaviours already described in the previous sections.

**Figure 9 – Distribution cumulative value of KiwiSaver contributions**



**Table 8 – Cumulative value of contributions: summary statistics**

Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$1,106	\$1,957	\$2,876	\$2,172
7	\$1,810	\$3,721	\$5,461	\$4,222
8	\$1,935	\$5,080	\$8,182	\$5,968

In view of this, Figure 9 shows two distinct patterns. The figure suggests a presence of a group of KiwiSaver members who joined at the start of the Scheme and consistently contributed to it across time. A person who continuously contributed around \$1,000 for 2 to 2.5 years (roughly the period between wave 6 and wave 8) would have approximately \$7,000 to \$8,500 in total cumulative contributions. This aligns well with the value of the Upper Quartile (\$8,182), which suggests that around 25% of KiwiSaver members consistently contributed for 2 years or more.

Further, the Figure 9 also highlights the constant inflow of new members into the Scheme, characterised by the large concentrations of individuals around the lower end of the contributions distribution. This observation is reinforced both by the fact that 25% of the contributions are below \$1,935 and that 25% of members would have contributed for just over a year by the time they were interviewed in wave 8.

## 4 SoFIE dataset

---

SoFIE collected information on KiwiSaver membership and contributions only once, during wave 8. While most of the information only covers one period, it can be used to create a record of KiwiSaver membership and contributions across waves 6, 7 and 8 for each individual who answered the KiwiSaver questions.

SoFIE respondents were asked to provide a month and year in which they first joined the Scheme, which were then used to produce membership levels and measurement of tenure for each wave.

SoFIE collected information on individual's member contribution rate which was used to impute contribution rates for earlier waves. In cases where individuals reported a contribution rate of 4% or more in wave 8, that rate was used for waves 6 and 7. In cases where the individual reported a contribution rate below 4% in wave 8, 4% rate was filled in for waves 6 and 7, since 4% was the minimum mandatory contribution rate at the time. These contribution rates and the members' yearly income from salary or wages were then used to calculate the value of annual member contributions for each wave.

SoFIE did not collect any information on employer contributions, which made it necessary to impute the contribution rates and amounts based on a set of assumptions. For waves 6 and 7, it was assumed that all employers contributed at the mandatory minimum rate of 1%. In wave 8, 2% rate was imputed to reflect the increase in the minimum rate. These rates were then used to calculate the value of annual employer contributions based on members' income from salary or wages.

SoFIE also did not collect any information on Government contribution such as the KickStart payment or the Member Tax Credit. These values were easily imputed using the rules that govern these payments.

Member, employer and Government contributions were then combined to create a measure of total cumulative contributions across waves. This measure of total cumulative contributions effectively mimics the measure that was created using the administrative data. The only notable difference is the exclusion of extra contributions, since SoFIE does not include any information on whether or when the individuals made extra contributions directly to their providers.

As with any data that are created using a set of assumptions instead of being directly collected, there is a risk that the assumptions do not effectively reflect reality. However, given the available information in SoFIE and the conservative approach to creating the variables, we believe that these derived variables represent a considered approach towards creating a reasonable set of information that could have been used in the evaluation of KiwiSaver in the absence of administrative data.

The following sections cover each variable that was created using the SoFIE data, describing the patterns across waves and noting when and how these patterns differ from those observed in the administrative data.

## 4.1 SoFIE KiwiSaver membership patterns

Tables 9 and 10 show KiwiSaver membership patterns and KiwiSaver membership by type of enrolment derived from the information collected in SoFIE. These tables contain information that can be closely compared to the information contained in Tables 1 and 2 in Section 3.

Table 9 shows that SoFIE KiwiSaver aggregate membership level closely matches the membership level observed in the IRD administrative data in wave 6. However, SoFIE underreports the number of members relative to the IRD measures of membership in subsequent waves, with the difference increasing from wave 7 to wave 8.

**Table 9 – Membership levels across waves (SoFIE)**

<i>KiwiSaver Membership (SoFIE)</i>		<i>Wave 6</i>	<i>Wave 7</i>	<i>Wave 8</i>
Not a Member	<i>Count</i>	1,730,600	1,636,200	1,568,500
	<i>Percent</i>	82.9%	76.3%	70.5%
Member	<i>Count</i>	356,800	507,800	655,500
	<i>Percent</i>	17.1%	23.7%	29.5%
Total		2,087,400	2,144,000	2,224,000

Table 10 offers some insight into the KiwiSaver member group that could be responsible for the observed difference in membership between SoFIE and the administrative data. In particular, Table 10 shows that SoFIE reports a lower level of automatically enrolled members for all waves, with the size of the difference between the two data sources increasing in waves 7 and 8. In wave 6, SoFIE measure of automatically enrolled members is lower by about 35% of the administrative measure. In waves 7 and 8, the difference between SoFIE and the administrative measures increases to over 50%.

**Table 10 – Members by type of enrolment (SoFIE)**

<i>Enrolment Type</i>		<i>Wave 6</i>	<i>Wave 7</i>	<i>Wave 8</i>
Opt-in via Employer or Provider	<i>Count</i>	283,800	394,500	502,000
	<i>Percent</i>	79.5%	77.7%	76.6%
Automatically enrolled	<i>Count</i>	69,800	109,400	147,700
	<i>Percent</i>	19.6%	21.5%	22.5%
Enrolled by parent	<i>Count</i>	3,100	3,900	5,700
	<i>Percent</i>	0.9%	0.8%	0.9%
Total (Members)		356,800	507,800	655,500

In contrast, the table shows that the number of individuals who reported joining KiwiSaver by opting-in via an employer or a provider is similar to the corresponding number of members in the administrative data. This suggests that individuals who made an active decision to join KiwiSaver could be more aware that KiwiSaver is a private superannuation scheme and can correctly identify the period in which they joined the Scheme.

Difference between the administrative and survey data could be due to a combination of factors, two of which are: SoFIE questionnaire routing or individuals' awareness/financial literacy. In the questionnaire, individuals were asked whether they have life insurance or if



they contribute to a superannuation scheme. If the respondent indicated that they have neither life insurance nor a superannuation scheme, they were not probed any further about whether they are a KiwiSaver member.

Routing the questionnaire in this manner means that more than 50% of those who are eligible to be KiwiSaver members were not asked whether they were part of KiwiSaver or probed about their reasons for not participating in the Scheme.

While it appears that the routing plays a major role in the observed differences in KiwiSaver membership, it is not altogether clear why some KiwiSaver members did not indicate that they are contributing to a superannuation scheme. One possible explanation could be that automatically enrolled KiwiSaver members have a lower level of financial literacy than members who actively joined KiwiSaver and are not aware that KiwiSaver is a private superannuation scheme.

It is also possible that some of the automatically enrolled individuals are not aware that they are part of KiwiSaver. Automatic enrolment is inherently a more passive way of joining KiwiSaver than opting in, which means that some individuals might not be aware that they were enrolled into KiwiSaver when they started their new job. This could be a possible explanation for at least some of the members, particularly if they have not received any information about their membership from their KiwiSaver provider before the interview.

Table 11 shows that a smaller proportion of KiwiSaver members in SoFIE are not currently contributing to the Scheme. The table shows that the proportion of those on a contribution holiday is roughly the same as what is observed in the administrative data (see Table 3). However, the proportion of those who are not contributing to KiwiSaver for some other reason is considerably smaller.

**Table 11 – Contributing members (SoFIE)**

<i>Contribution Status of KiwiSaver Members in Wave 8 (SoFIE)</i>	<i>Count</i>	<i>Percent</i>
Contributing	596,400	91%
Contribution Holiday	13,800	2.1%
Not Contributing (Other reason)	45,300	6.9%
Wave 8 Members	655,500	100%

The difference could be explained by the fact that contributions are reported on the quarterly basis in the administrative data. This results in some members not being recorded as contributing if they don't have a full quarter of contributions before their interview for any given wave. In contrast, SoFIE data were used to impute contributions straight from the date of enrolment. This means there is no delay between the date of enrolment and the date at which the individuals count as having contributed to the Scheme.

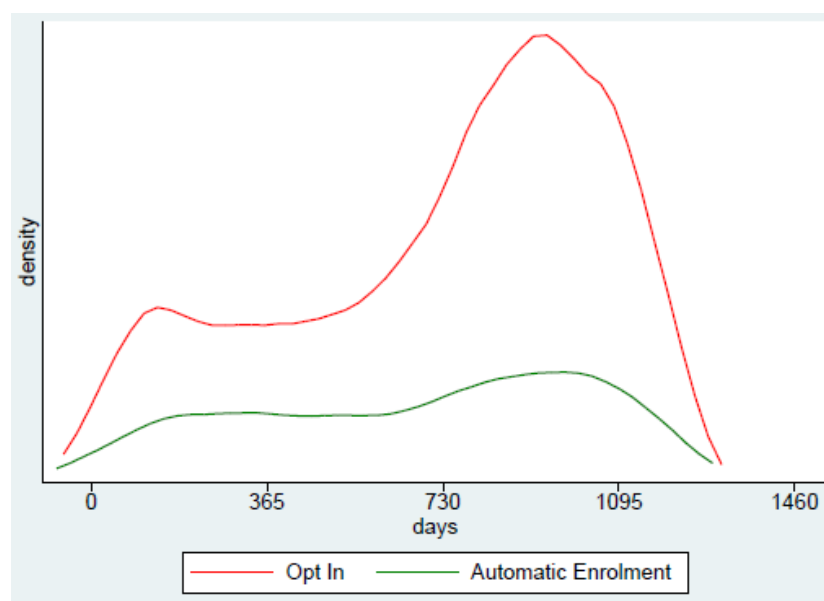


## 4.2 SoFIE KiwiSaver membership tenure

Figure 10 and Table 12 summarise the pattern and distribution of KiwiSaver tenure derived from SoFIE. The figure and the table can be compared to Figure 1 and Table 4 in Section 3.

Before discussing the tenure patterns, it should be noted that SoFIE respondents were asked only to provide the month and the year during which they enrolled into KiwiSaver. This information was combined to produce the enrolment date, which was set to the start of the month in which the individual enrolled. This treatment of the enrolment information will likely lead to slightly longer derived tenure in KiwiSaver for SoFIE respondents, since most members would not have joined KiwiSaver at the start of the month.

**Figure 10 – Membership tenure (in days) by type of enrolment (SoFIE)**



**Table 12 – Summary of membership tenure (in days) by type of enrolment (SoFIE)**

Wave	Type of Enrolment	Lower Quartile	Median	Upper Quartile	Mean
6	Opted in	128	231	339	233
	Automatically enrolled	132	233	346	235
7	Opted in	356	520	662	494
	Automatically enrolled	286	494	654	459
8	Opted in	499	824	994	734
	Automatically enrolled	401	757	970	682

Figure 10 suggests that SoFIE captures the general trends in membership tenure. Similarly to administrative data, SoFIE picks up on a strong rush in enrolments by members who opted in to KiwiSaver at the start of the Scheme. SoFIE also picks up on the increase in the number of members who opted in to the Scheme during wave 8. Finally, the tenure pattern of automatically enrolled individuals appears to closely mirror the pattern observed in the administrative data.

Table 12 provides a more in-depth look at the tenure patterns across waves. The most striking feature of the table can be observed in wave 6 and shows that automatically enrolled members have virtually the same distribution of tenure as those who actively joined the Scheme.

The reason for this pattern is unclear. The observed pattern could in part be due to recall error resulting from asking respondents to recall a date that can be up to two and a half years removed from their current interview date. It is also possible that that the recall error has a stronger effect on the individuals who automatically enrolled into KiwiSaver and could lead to some members reporting a much earlier date of enrolment, while others report a date that is substantially later than actual date of enrolment.

Patterns in waves 7 and 8 are closer to the patterns observed in the administrative data. However, consistently longer tenure in SoFIE suggests that the recall error could be affecting the SoFIE measurement of tenure even in later waves. Furthermore, the fact that SoFIE KiwiSaver membership levels in waves 7 and 8 is over 20% lower than the observed level in the IRD data suggests that SoFIE might not be as effective in capturing members who join during these waves. Smaller inflow of new members in SoFIE would result in longer tenure, particularly at the lower end of the tenure distribution.

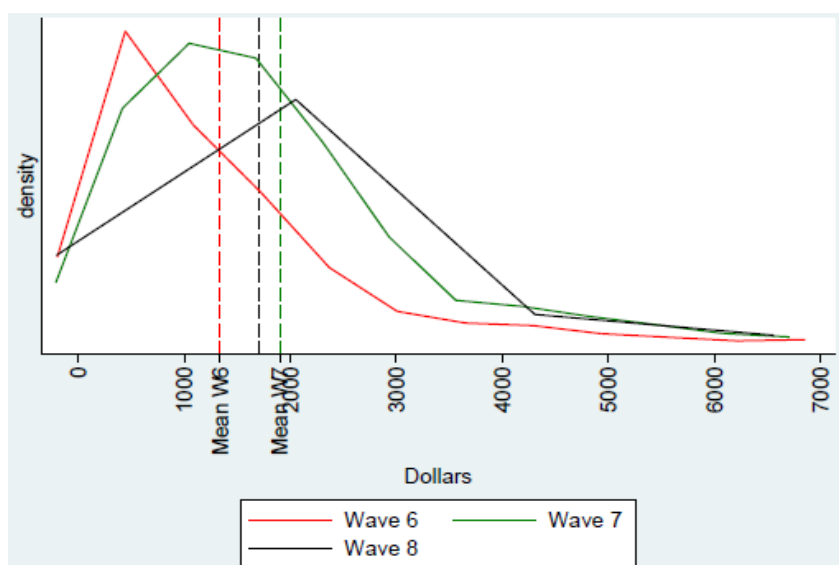
Overall, the difference in average tenure and membership levels between SoFIE and administrative data is likely to have flow-on effects on the remaining variables. The observed effects on each are covered separately in subsequent sections.

### 4.3 Annual member contributions and rates

Figure 11 shows that while the shape of the distributions of annual member contributions in SoFIE for wave 6 and 7 resemble the distributions of annual member contributions in the administrative data, SoFIE values of annual member contributions are noticeably higher than values from administrative data over that period (see Figure 2).

The observed difference between SoFIE and the IRD measures of annual member contributions could be partially explained by the differences in average tenure between the two data sources. If that's indeed the case, the difference will most severely affect new members, since even a minor difference in tenure for recent members could result in a noticeable difference in annual contributions.

**Figure 11 – Distribution of annual member contributions (SoFIE)**



**Table 13 – Summary of annual member contributions (SoFIE)**

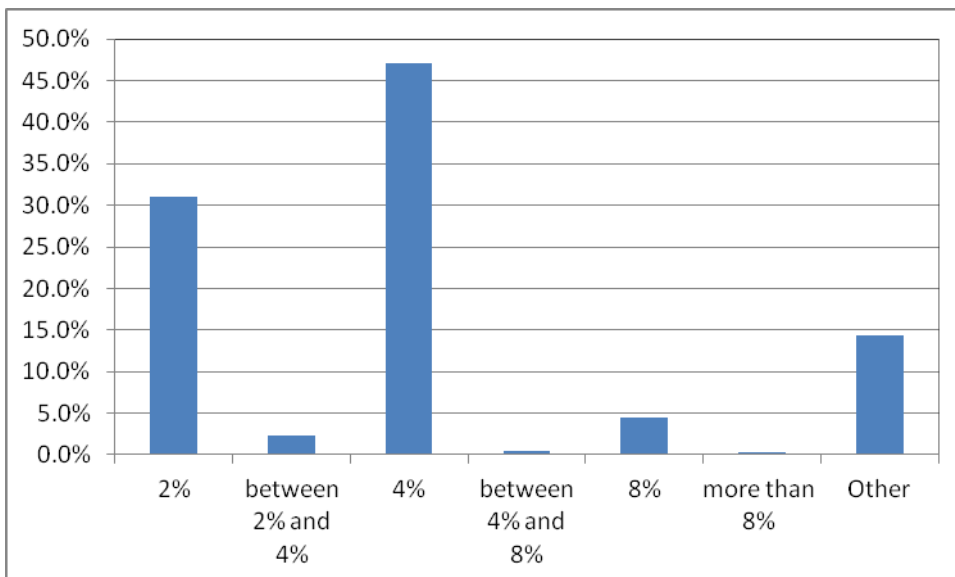
Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$391	\$925	\$1,718	\$1,338
7	\$843	\$1,514	\$2,341	\$1,899
8	\$646	\$1,273	\$2,124	\$1,708

SoFIE also reports relatively higher values of annual member contributions in wave 8, but the difference between SoFIE and the administrative measures is less pronounced. This could be due to the fact that individuals are better at reporting recent details and events. For members who joined KiwiSaver in waves 6 and 7, SoFIE measure of annual contributions in wave 8 could be closer to the IRD measure due to a better match between the reported and the administratively recorded contribution rates. For members who join the Scheme in wave 8, SoFIE measure of annual member contribution in wave 8 could better match the administrative measure due to a better match in the contribution rates and enrolment dates.

The remaining difference in wave 8 between the SoFIE and the administrative measures of annual contributions is potentially explained by the difference in membership levels between the two datasets. SoFIE appears to underreport new members relative to administrative data, particularly in wave 8. This could lead to fewer individuals at the lower end of the contributions distributions, since new members are less likely to have contributed for the entire year.

Figure 12 shows the distribution of member contribution rates reported by SoFIE respondents in wave 8 and can be compared to Figure 3 in Section 3. In general, the distribution appears to roughly match the distribution of contribution rates in the administrative data, with two major differences. First, the rates reported by the respondents represent their current contribution rate, which might not reflect the rate at which members contributed for the majority of their time in the Scheme. This can result in a sizeable mismatch between SoFIE and administrative measures of contributions, even if the individual correctly reported the rate at which they are currently contributing.

**Figure 12 – Annual member contribution rates (SoFIE)**



Another feature of SoFIE member contribution rates is the presence of the “Other” category. This category includes individuals who reported that they contribute at different

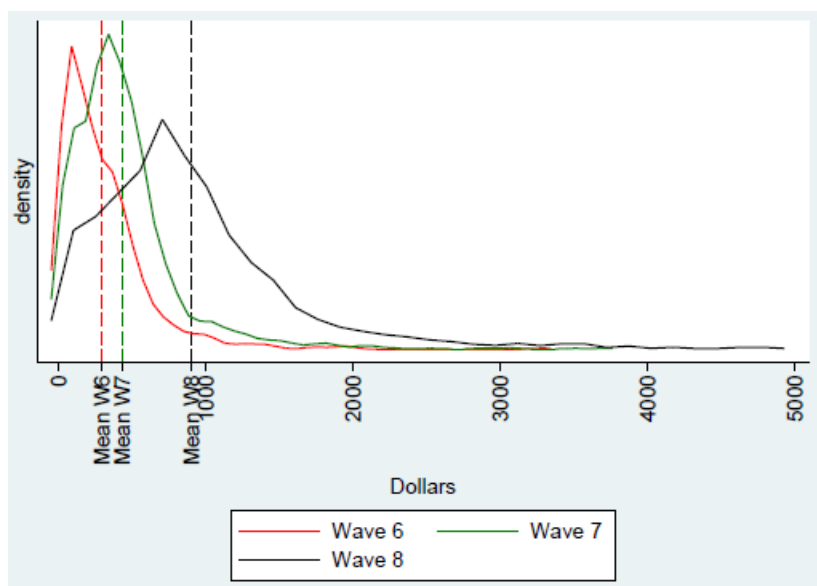
rates through more than one employer and individuals who only make irregular payments. For both groups, SoFIE does not collect any other information about their contribution rates, making it impossible to determine what proportion of income these individuals contributed to KiwiSaver. This is a significant weakness in the survey data.

#### 4.4 Annual employer contributions

Figure 13 shows the pattern of annual employer contributions and Table 14 provides a summary of the employer contribution distribution for each wave. The figure and the table can be directly compared to Figure 4 and Table 6 in Section 3.

Since all employers were assumed to contribute at the mandatory minimum rate in the calculation of annual employer contributions in each wave, Figure 13 shows how tenure and members' income translate into the distribution of employer contributions.

**Figure 13 – Distribution of annual employer contributions (SoFIE)**



**Table 14 – Summary of annual employer contributions (SoFIE)**

Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$96	\$221	\$409	\$295
7	\$209	\$370	\$547	\$431
8	\$459	\$773	\$1,136	\$903

Figure 13 and Table 14 show that SoFIE reports noticeably higher values of annual employer contributions than those observed in the administrative data particularly in waves 7 and 8. Similarly to the discussion in the previous sections, higher values in SoFIE can potentially be due to the noticeable difference in the number of new members between the two datasets. Since new members have shorter tenure in the Scheme, it is possible that SoFIE is missing members who would have been at the lower end of the employer contributions distribution.

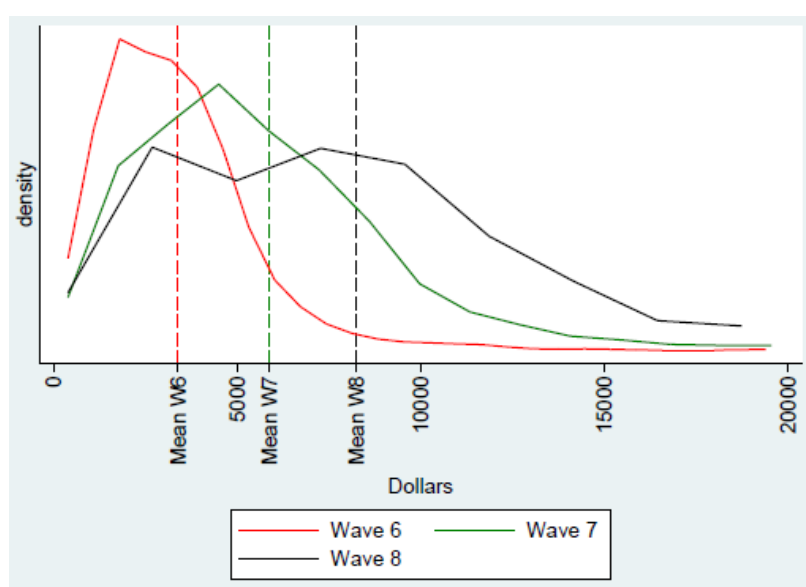
Even though SoFIE reports higher values for the lower quartile, median and upper quartile of the employer contributions, SoFIE reports a lower value for the mean in waves 6 and 7. This suggests that administrative data contain individuals whose employers contribute

above and beyond what would be expected if the employer contributed at the mandatory minimum rate of 1% during these waves.

## 4.5 Total cumulative contributions to KiwiSaver

Figure 14 and Table 15 summarise the distribution of the total cumulative value of all contributions to KiwiSaver derived using SoFIE information and can be compared to Figure 9 and Table 8 in Section 3. In addition to some of the features described in the previous sections, SoFIE does not contain any information on whether members made any extra contributions towards KiwiSaver directly through their provider. At an individual level, this should correspond to lower values of cumulative contributions in SoFIE for at least some of the individuals.

**Figure 14 – Distribution of cumulative value of KiwiSaver contributions (SoFIE)**



**Table 15 – Cumulative value of KiwiSaver contributions: summary statistics (SoFIE)**

Wave	Lower Quartile	Median	Upper Quartile	Mean
6	\$1,831	\$3,041	\$4,256	\$3,399
7	\$3,335	\$5,155	\$7,530	\$5,888
8	\$3,889	\$7,341	\$10,715	\$8,247

Results presented in Table 15 show that SoFIE values of cumulative contributions are consistently at a higher level across the measures presented in the table. The difference between SoFIE and administrative data is particularly noticeable in the values of lower quartiles. The value of the lower quartile in SoFIE in wave 8 is more than two times greater than the value of the lower quartile in the administrative data. These results, in conjunction with the discussion in from the previous sections, underscore the differences in the survey and administrative data at the aggregate level.

## 5 Key points of difference

This section summarises the main points of difference that exist in the KiwiSaver membership levels, membership tenure and member contribution rates between the SoFIE and the administrative datasets.

### 5.1 KiwiSaver membership

Table 16 summarises the overall KiwiSaver membership patterns as well as the number of automatically enrolled members in the SoFIE and the administrative datasets. From the table, it is clear to see that SoFIE closely matches the number of KiwiSaver members in the administrative data in wave 6, but begins to underreport the number of KiwiSaver members relative to administrative measure in subsequent waves.

**Table 16 – Summary table: KiwiSaver membership and automatically enrolled members**

<i>KiwiSaver Membership</i>	<i>Administrative Data</i>	<i>SoFIE</i>	<i>Difference (as % of the administrative measure)</i>
Wave 6	368,800	356,800	3.25%
Wave 7	636,600	507,800	20.23%
Wave 8	836,600	655,500	21.65%
<i>Automatically Enrolled Members</i>	<i>Administrative Data</i>	<i>SoFIE</i>	<i>Difference (as % of the administrative measure)</i>
Wave 6	106,600	69,800	34.52%
Wave 7	222,400	109,400	50.81%
Wave 8	306,100	147,700	51.75%

Difference in the number of automatically enrolled members between the two datasets appears to be the primary driver of the observed differences in the overall membership levels. In wave 6, the SoFIE measure is lower by almost 35% of the administrative measure. In waves 7 and 8, the difference between the two measures increase further, with the SoFIE measure being over 50% lower than the administrative measure of automatically enrolled members.

### 5.2 KiwiSaver membership tenure

Table 17 provides an illustrative example of the differences in membership tenure in the SoFIE and the administrative datasets. For the automatically enrolled members, lower quartile, median, and upper quartile of membership tenure are substantially higher in SoFIE relative to the corresponding measures in the administrative data. The most striking feature is the fact that SoFIE median membership tenure is almost as high as the upper quartile of membership tenure in the administrative data across all waves.

**Table 17 – Summary table: KiwiSaver membership tenure for automatically enrolled members**

<i>Membership Tenure for automatically enrolled</i>	<i>Administrative Data</i>			<i>SoFIE</i>		
	<i>Lower Quartile</i>	<i>Median</i>	<i>Upper Quartile</i>	<i>Lower Quartile</i>	<i>Median</i>	<i>Upper Quartile</i>
Wave 6	67	138	230	132	233	346
Wave 7	222	366	496	286	494	654
Wave 8	370	635	807	401	757	970

Recall error could be one of the factors driving the observed difference between the two datasets, since SoFIE respondents are asked to recall their enrolment date that can be up to two and a half years removed from their interview date in wave 8. In addition, the fact that SoFIE measures of KiwiSaver membership in wave 7 and 8 are more than 20% lower relative to the administrative measure suggests that SoFIE might not be as effective in capturing members who join during these waves. Smaller inflow of new members in SoFIE would result in longer tenure, particularly at the lower end of the tenure distribution.

### 5.3 KiwiSaver member contribution rates

Analysis of the SoFIE and administrative data on contribution rates highlighted two major differences between the two measures. First, the rates reported by the respondents in SoFIE represent their current contribution rate, which might not reflect the rate at which members contributed for the majority of their time in KiwiSaver. This can result in a sizeable mismatch between member contributions derived from SoFIE data and member contributions recorded in the administrative data, even if the individual correctly reported the rate at which they are currently contributing.

Another feature of SoFIE member contribution rates is the presence of the “Other” category. This category includes individuals who reported that they contribute at different rates through more than one employer and individuals who only make irregular payments. For both groups, SoFIE does not collect any other information about their contribution rates, making it impossible to determine what proportion of income these individuals contributed to KiwiSaver. This is a significant weakness in the survey data.

Overall, these key differences between the two datasets are likely to affect the differences observed in other variables such as member and employer contributions as well as total cumulative contributions. These differences are explored in greater detail in the next section.



## 6 Analysis of differences

This section provides a more detailed look at the differences that exist at the individual level in KiwiSaver membership and membership tenure as well as the differences in member, employer and total cumulative contributions. By examining these differences at the individual level, the section attempts to highlight the internal composition and dynamics of the observed differences. This can potentially identify patterns within the differences that could shed some light into which groups of KiwiSaver members have the greatest mismatch between their SoFIE responses and administrative records.

This section begins by covering the differences in membership levels and dates of enrolment first, before using simple econometric techniques to estimate the main determinants of differences in annual member and employer as well as total cumulative contributions to KiwiSaver.

### 6.1 KiwiSaver membership

Table 18 shows the mismatch in KiwiSaver membership data between SoFIE and the administrative data across waves. The table shows that only 12.3% of individuals who are eligible to be in KiwiSaver were recorded as KiwiSaver members in both SoFIE and the IRD data in wave 6. At the same time, over 10% were recorded as being a member in one source, but not the other. This highlights the fact that even though on aggregate the two data sources show similar membership levels in wave 6, significant differences are present at the individual level.

**Table 18 – Differences in membership**

KiwiSaver Membership (SoFIE)	KiwiSaver Membership (Admin Data)		
	Wave 6		
	Not a Member	Member	Total
Not a Member	77.6%	5.3%	82.9%
Member	4.8%	12.3%	17.1%
Total	82.3%	17.7%	100%

KiwiSaver Membership (SoFIE)	KiwiSaver Membership (Admin Data)					
	Wave 7			Wave 8		
	Not a Member	Member	Total	Not a Member	Member	Total
Not a Member	68.0%	8.3%	76.3%	60.6%	9.9%	70.5%
Member	2.3%	21.4%	23.7%	1.7%	27.7%	29.5%
Total	70.3%	29.7%	100%	62.4%	37.6%	100%

The table also shows that the proportion of individuals who reported being a member in SoFIE, but were not recorded as a member in the IRD data drops after wave 6. This pattern is possibly explained by the erroneous reporting of the enrolment date by some members. This would lead them to be identified as KiwiSaver members in wave 6, when in reality they enrolled into KiwiSaver a few months later in wave 7. This error appears to be mostly corrected after wave 6. Moreover, the table shows a large jump in the number of individuals who are members in the administrative data, but not in SoFIE. This is



possibly a result of questionnaire design in SoFIE. As noted earlier, individuals who indicated that they didn't have life insurance or a private superannuation scheme were not asked any questions about KiwiSaver due to questionnaire routing.

Table 19 presents a decomposition of KiwiSaver members who appear in the administrative data, but do not report being KiwiSaver members in wave 8 of SoFIE. The table shows that the greatest proportion of KiwiSaver members missing in SoFIE comes from those who automatically enrolled, followed by those who actively chose their provider. Most interestingly, members who actively enrolled in KiwiSaver either through their employer or a provider comprise the majority of individuals who appear in the IRD dataset, but not in SoFIE.

**Table 19 – KiwiSaver members who don't appear in SoFIE by wave 8, by enrolment type**

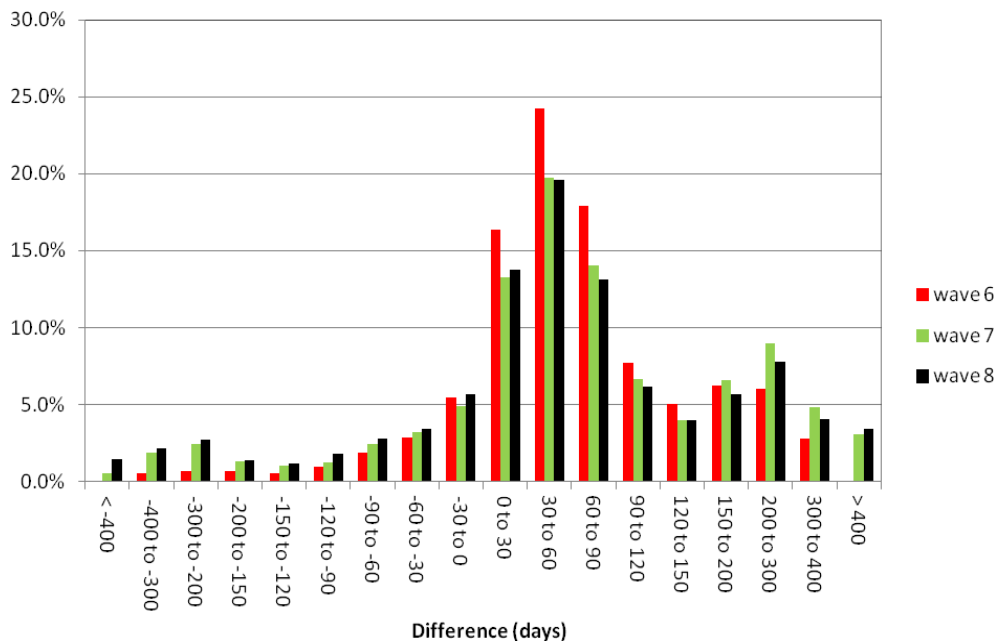
Opt in via Employer	16.0%
Chose provider	38.9%
Automatically enrolled	45.1%
<b>Total</b>	<b>100%</b>

The fact that both passive and active KiwiSaver members appear in the group who are underreported by SoFIE suggests that individuals may view KiwiSaver separately from private superannuation schemes.

## 6.2 Enrolment dates

Figure 15 summarises the difference (in days) between IRD and SoFIE KiwiSaver enrolment dates. In the figure, a positive difference indicates that the SoFIE enrolment date precedes the IRD enrolment date. Conversely, a negative difference indicates that SoFIE enrolment date is after the IRD enrolment date.

**Figure 15 – Difference between the IRD and SoFIE enrolment dates in days**



From Figure 15, it can be seen that in the majority of cases, individuals report an enrolment date in SoFIE that precedes their enrolment date in the administrative data. This is most apparent in wave 6, with only approximately 10% of members reporting an enrolment date that was after their administrative enrolment date.

The figure also shows that only a small group of KiwiSaver members in SoFIE indicated an enrolment date that is the same as their enrolment date in the administrative data. The 0 to 30 day difference group shows that roughly 15% of SoFIE respondents provided a correct month of enrolment in any given wave. More often, the respondents report an enrolment date that is up to 3 months before the administrative enrolment date. This is most evident in wave 6, with over 40% of the respondents indicating an enrolment date that is between 30 and 90 days prior to the administrative date.

This result indicates that even if there were no other differences between the SoFIE and administrative data apart from the date of enrolment, SoFIE will overestimate the value of annual member contributions in the year of enrolment by 8-25% (1 to 3 month difference in dates in that year) for a substantial proportion of KiwiSaver members.

Finally, Figure 15 shows a distinct group of individuals whose IRD enrolment date precedes the date they reported in SoFIE. This unexpected result shows that for some reason, individuals were not aware of their membership. This result can indicate that some members are not actively monitoring their KiwiSaver contributions and have issues recollecting when they first joined the Scheme. Alternatively, this can indicate that some members did not receive much information about their KiwiSaver when they joined and instead reported the date when they received their first annual statement from their provider. Without further information, it is impossible to truly understand the reason behind the observed pattern.

### 6.3 Regression results

This section of the paper specifically looks at how well SoFIE predicts values of annual member and employer contributions, as well as total cumulative contributions at the individual level. In order to achieve this, Ordinary Least Squares (OLS) regression is used to estimate the values of contributions from the IRD using the values that were derived using SoFIE information. Due to the differences that exist in KiwiSaver membership between the two dataset within a given wave, this section only covers cases where individuals appear in both datasets.

Two models are used to study the relationship between SoFIE and administrative values. Linear models test a simple 1-to-1 relationship between the administrative and SoFIE values, while the quadratic models accounts for some non-linear relationship between SoFIE and IRD at specific levels of contributions.

There are three elements of the results that indicate how well SoFIE predicts administrative values of contributions.

The estimated coefficient indicates what value, on average, 1 dollar in SoFIE corresponds to in the administrative data. A coefficient of 1 means that 1 dollar of contributions in SoFIE corresponds to 1 dollar of contributions in the administrative data. A coefficient greater than 1 indicates that SoFIE underreports the value of contributions compared to the administrative measure, while a coefficient smaller than 1 indicates that SoFIE reports a higher value of contributions than what is observed in the administrative data.

The reported constant also contains important information about the relationship between the IRD and SoFIE values. A large positive constant indicates that regardless of the relationship predicted by the coefficient, the SoFIE measure is lower than the administrative measure of KiwiSaver contributions at the lower end of the contributions distribution. A negative constant indicates that the SoFIE measure exceeds the administrative measure at the lower end of the contributions distribution. For the rest of the value ranges, the constant adjusts the relationship predicted by the regression coefficient, which makes it necessary to interpret the two values together.

Finally, the R-squared statistic shows how well SoFIE values explain the variation within the IRD values. R-squared values close to 1 indicate a high level of fit between the values predicted by the model and the actual values. Low values indicate that there is a lot of variation in the actual values around the values predicted by the model.

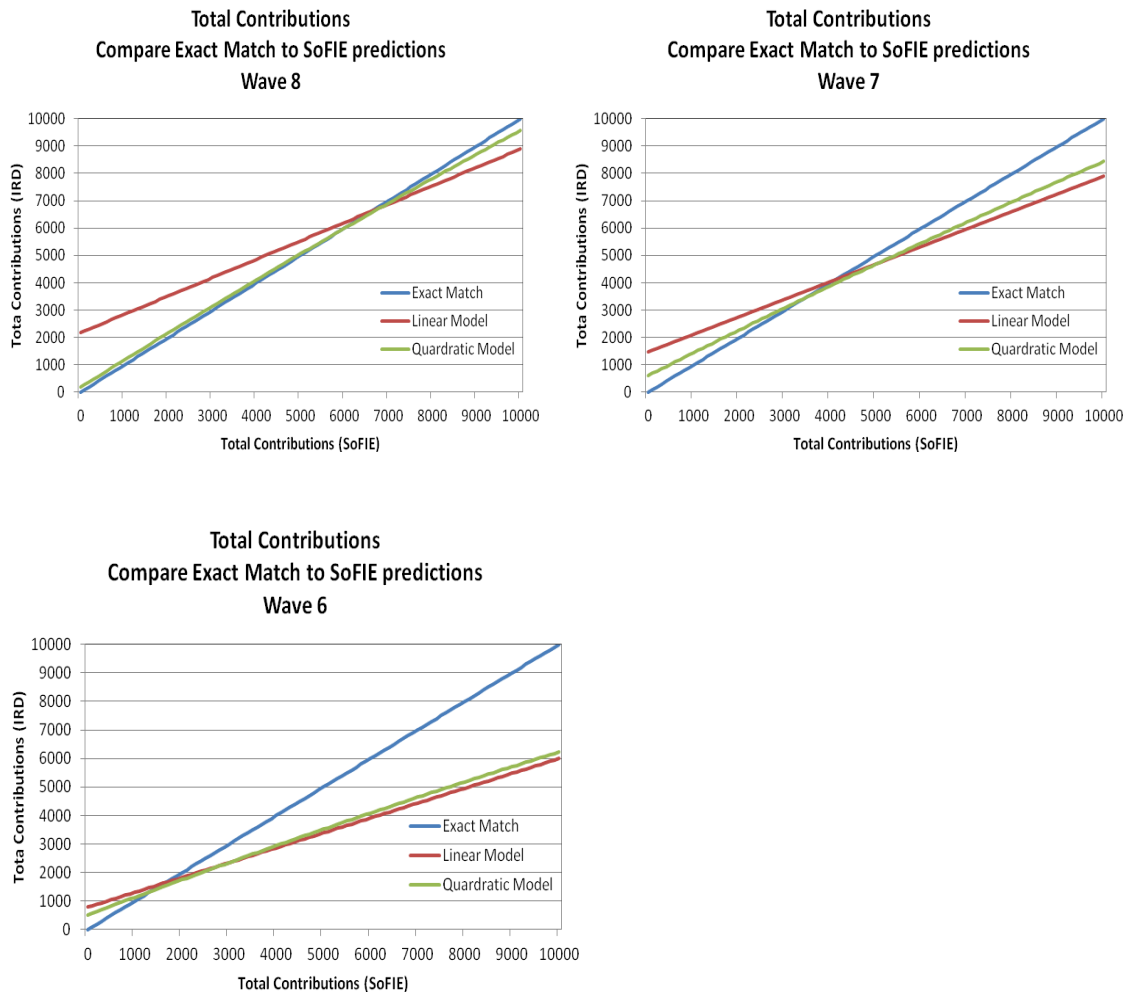
### 6.3.1 Total cumulative contributions to KiwiSaver

Table 20 contains the results for the regression of SoFIE values of total cumulative contributions on the corresponding values in the administrative data. Figure 16 shows the graphical representation of these results. In each graph, the blue line represents the scenario where 1 dollar of total cumulative contributions in SoFIE corresponds to 1 dollar of total cumulative contributions in the administrative data. Values to the right of the blue line indicate that SoFIE values of cumulative contributions are higher than the values in the administrative data. Conversely, values to the left of the blue line indicate that SoFIE values of cumulative contributions are lower than the corresponding values in the administrative data.

**Table 20 – Regression results (total cumulative contributions)**

	Wave 8		Wave 7		Wave 6	
	<i>Linear Model</i>	<i>Quadratic Model</i>	<i>Linear Model</i>	<i>Quadratic Model</i>	<i>Linear Model</i>	<i>Quadratic Model</i>
SoFIE Total Contributions	0.6712*** (0.0757)	1.0154*** (0.0414)	0.6449*** (0.0498)	0.8469*** (0.0594)	0.5214*** (0.0573)	0.6328*** (0.0779)
Total Contributions Squared	-	-7.47e-06*** (1.17e-06)	-	-6.40e-06*** (2.45e-06)	-	-6.09e-06 (6.22e-06)
Constant	2179.38***	174.18	1461.44***	605.6491***	785.76***	509.26***
Number of Observations	2603	2603	2053	2053	1390	1390
R-squared	0.6020	0.6939	0.5421	0.5686	0.4305	0.4390

**Figure 16 – Total cumulative contributions: regression graphs**



Results from the linear model show that SoFIE underestimates the value of total cumulative contributions relative to the administrative data at the lower level of contributions, while overestimating the values of total cumulative contributions at the higher level of contributions. The estimated coefficient shows that, on average, 1 dollar of total cumulative contributions in SoFIE corresponds to about 67 cents in administrative data in wave 8, 64 cents in wave 7 and 52 cents in wave 6. The observed increase in the constant from wave 6 to wave 8 can be interpreted as the growth in the total cumulative value of contributions over time.

The R-squared values suggest a large degree of variation around the values predicted by the linear model. Increase in R-squared values over the course of three waves suggests that SoFIE is more precise at predicting the values of total cumulative contributions in later waves.

The results from the quadratic model show an unambiguous improvement in the overall fit. For wave 8, the quadratic model shows that on average SoFIE does a relatively good job of predicting the values of total cumulative contributions up to about \$7,000. Statistical insignificance of the constant and a significant estimated coefficient of 1.0154 means that on average SoFIE slightly underestimates the value of cumulative contributions relative to administrative data at lower levels of contributions. For values above \$7,000, SoFIE overestimates the value of total cumulative contributions relative to administrative data, which is consistent with what was observed from the linear model.

The quadratic model also achieves better results in waves 7 and 6. For both waves, the improvement in the estimated coefficient and the reduction in the constant suggest that SoFIE underestimates the values of cumulative contributions relative to administrative data at lower levels of contributions by less than what is suggested by the linear model. However, these improvements do not change the overall conclusion reached from the results of the linear model.

### 6.3.2 Annual member contributions

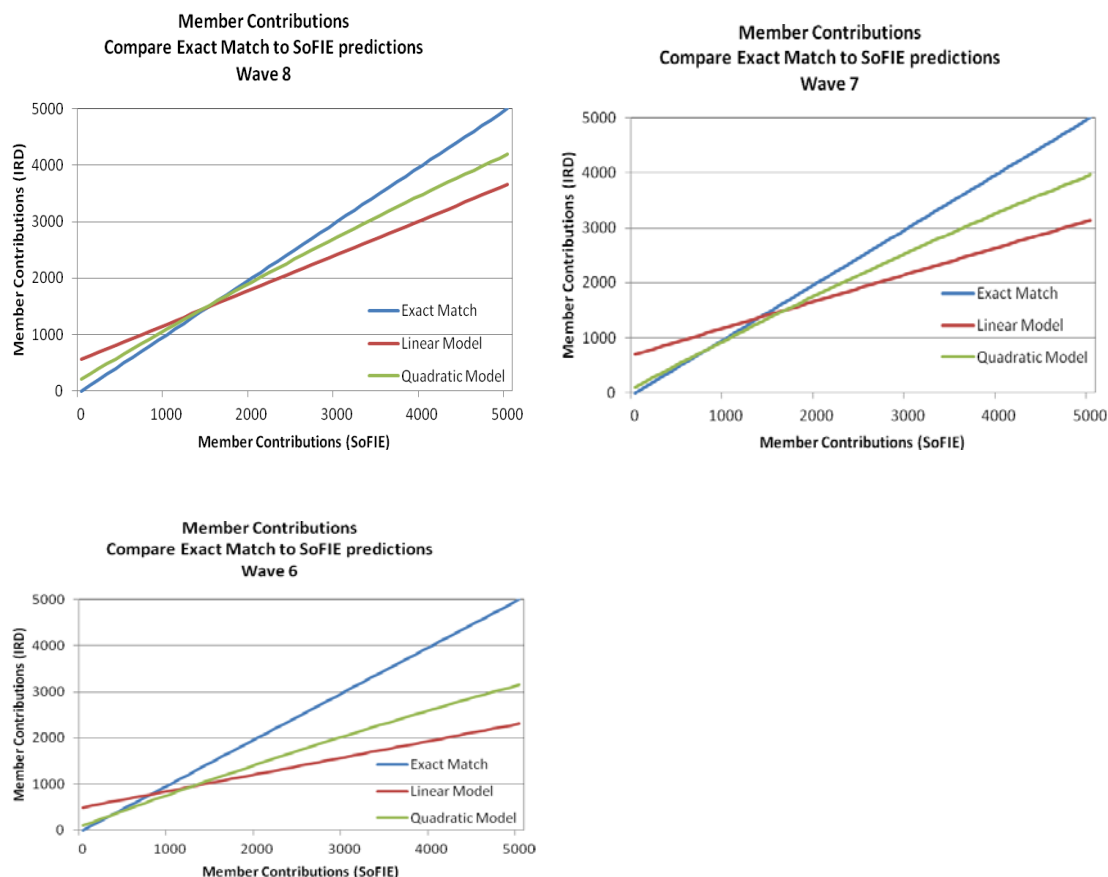
Similarly to the results from total cumulative contributions, the results from the regression of SoFIE values of annual member contributions on administrative values suggest that SoFIE underestimates the value of annual member contributions relative to administrative values at the low levels of contributions, while overestimating the values of contributions at the higher levels of contributions.

In wave 7 and 8, SoFIE appears to begin overestimating the values of annual member contributions relative to administrative data at around \$1,700. In wave 6, SoFIE begins to overestimate the values of member contributions at a much lower level, reflecting the lower overall levels of contributions in that wave. The results show that 1 dollar in SoFIE corresponds to 62 cents in administrative data in wave 8, 49 cents in wave 7 and 36 cents in wave 6.

**Table 21 – Regressions results (annual member contributions)**

	<b>Wave 8</b>		<b>Wave 7</b>		<b>Wave 6</b>	
	<i>Linear Model</i>	<i>Quadratic Model</i>	<i>Linear Model</i>	<i>Quadratic Model</i>	<i>Linear Model</i>	<i>Quadratic Model</i>
SoFIE Member Contributions	0.6175*** (0.0571)	0.9043*** (0.0392)	0.4873*** (0.0674)	0.8879*** (0.0459)	0.3625*** (0.0739)	0.7023*** (0.0486)
Member Contributions Squared	-	-0.0000217*** (3.66e-06)	-	-0.0000234*** (3.39e-06)	-	-0.0000184*** (2.88e-06)
Constant	566.816***	211.4549***	704.5261***	106.0694*	494.8878***	100.4454**
Number of Observations	2424	2424	1876	1876	1091	1091
R-squared	0.6036	0.6762	0.4690	0.6050	0.4195	0.5663

**Figure 17 – Annual member contributions: regression graphs**



Similarly to what was observed in the results for total cumulative contributions, the quadratic model produces a substantially better fit than the linear model. In wave 8, the results from the quadratic model indicate that SoFIE slightly underestimates annual member contributions relative to administrative data for values under \$2,500 and overestimates annual member contributions for values greater than \$2,500.

In relation to waves 7 and 6, the results from the quadratic model show a substantial reduction in the constant. This means SoFIE starts to overestimate values of member contributions at a much lower level than what is predicted by the linear model. In wave 7, SoFIE appears to overestimate the values of member contributions for contributions greater than \$1,500. For wave 6, SoFIE begins to overestimate the value of contributions for values greater than \$700.

### 6.3.3 Annual employer contributions

The results from the regression on annual employer contributions are striking, particularly in relation to the difference that is observed between the results for wave 8 and earlier waves.

In wave 8, the results from the linear model suggest that SoFIE slightly overestimates the value of annual employer contributions relative to administrative data across all levels of contributions. Insignificant constant and an estimated coefficient of 0.9473 show that 1 dollar predicted by SoFIE corresponds to 95 cents of annual employer contributions in the administrative data. Similar results are observed from the quadratic model.

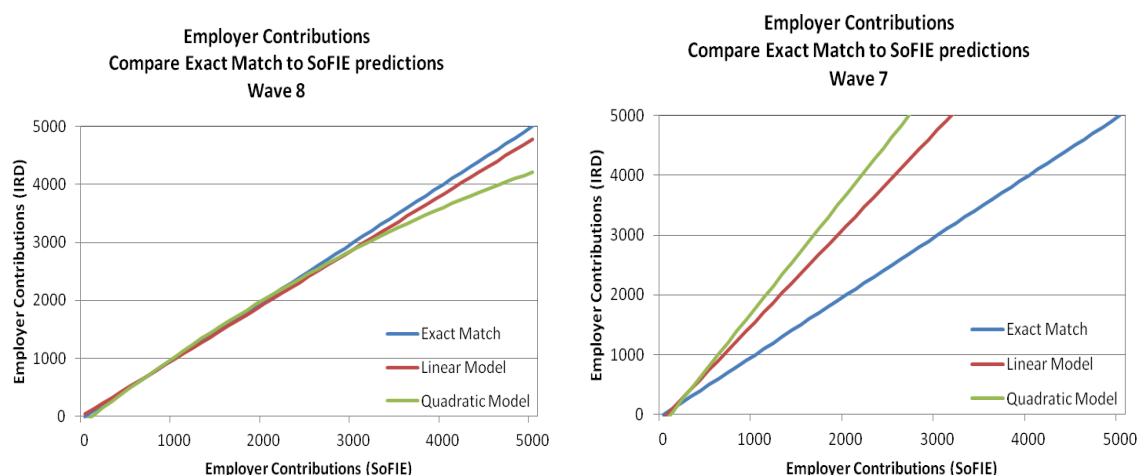
Results from waves 6 and 7 are substantially different. In wave 7, the results indicate that SoFIE overestimates the values of annual employer contributions relative to administrative data at very low levels, but generally underestimates the values of annual employer contributions for higher levels of contributions.

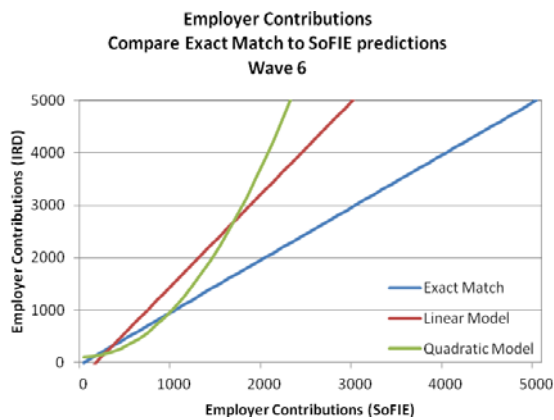
Wave 6 linear model results are similar to the results from the linear model in wave 7. However, wave 6 results from the quadratic model suggest that SoFIE slightly overestimates the value of annual employer contributions up to around \$1,000. Past this point, SoFIE appears to dramatically underestimate the value of annual employer contributions relative to administrative data.

**Table 22 –Regression results (annual employer contributions)**

	Wave 8		Wave 7		Wave 6	
	Linear Model	Quadratic Model	Linear Model	Quadratic Model	Linear Model	Quadratic Model
SoFIE Employer Contributions	0.9473*** (0.0443)	1.1765*** (0.0631)	1.6067*** (0.1843)	1.9102*** (0.2865)	1.7610*** (0.4006)	0.2930 (0.2740)
Employer Contributions Squared	-	-0.0000636*** (0.0000208)	-	-0.0001611 (0.0002014)	-	0.0009395*** (0.0001629)
Constant	42.2924	-83.3381***	-59.5541	-142.2445**	-233.3647*	109.3419
Number of Observations	2185	2185	1677	1677	471	471
R-squared	0.6424	0.6524	0.3608	0.3647	0.3576	0.4538

**Figure 18 – Annual employer contributions: regression graphs**





The results for all waves can be potentially explained by the changes in the employer contribution rates across time. In wave 8, the administrative data indicate that almost all employers contributed at 2%, which closely matches the contribution rate that was imputed for SoFIE respondents in order to derive the value of annual employer contributions.

In waves 7 and 6, administrative data show that employers contributed at various rates, with a distinct group contributing above the 1% mandatory minimum. As noted earlier, in the absence of information on employer contributions in SoFIE, 1% contribution rate was used to derive the values of annual employer contributions. At the individual level, this creates a mismatch between the actual and the imputed rate of employer contributions, leading to SoFIE underestimating employer contributions relative to administrative data.

Coincidentally, the large difference in the R-squared value in wave 8 compared to waves 7 and 6 is consistent with the conclusions reached earlier in this section. Lower R-squared values in waves 7 and 6 suggest a large degree of variation around the values predicted by SoFIE, indicating strong heterogeneity in the contribution behaviour among employers.

These results demonstrate some of the shortcomings of the survey data. Without additional information on annual employer contribution rates, it is impossible to determine when a higher rate of contributions should be used. This results in noticeable differences between values predicted by SoFIE and those observed in the administrative data.



## 6.4 Main determinants

This section provides a detailed analysis of the main determinants of the difference observed between SoFIE and the administrative data at the individual level. The results in this section are produced using pooled regression on the difference between the IRD variables and the SoFIE variables.

Table 23 summarises the list of explanatory variables that were used to explain the difference in the IRD and SoFIE variables. Table 24 summarises the results from three different models of differences between SoFIE and the administrative data. Model 1 examines the main determinants of the difference in total cumulative contributions between SoFIE and administrative data. Model 2 and Model 3 look at difference in annual member and employer contributions respectively.

**Table 23 – Main Determinants: summary of regression variables**

<b>Variable name</b>	<b>Definition</b>
Difference in Dates	Difference between IRD KiwiSaver enrolment date and the SoFIE enrolment date measured in days. Positive value indicates that the IRD enrolment date is after the SoFIE enrolment date.
Difference in Member Rates	Difference between IRD member contribution rate and SoFIE contribution rate measured in percentage points. Positive value indicates that IRD rate is greater than the SoFIE rate.
Difference in Earnings	Difference between IRD earnings from salary and wages and SoFIE earnings measured in dollars. Positive value indicates that IRD earnings are greater than SoFIE earnings.
Squared Difference in Earnings	Squared value of the difference in earnings from salary and wages
Difference in Employer Rates	Difference between IRD employer contribution rate and SoFIE contribution rate measured in percentage points. Positive value indicates that IRD rate is greater than SoFIE rate
Income Flag	Dummy variable that identifies cases where individuals did not report some element of the income from salary or wages. The variable takes a value of one if there is some element missing.
Extra Flag	Dummy variable that identifies instances where individuals made extra contributions to their KiwiSaver account through a provider
Wave	Dummy variables representing survey waves (wave 6 and 7)

The results show that the difference in enrolment dates has a strong effect on the observed difference between SoFIE and the administrative data in total cumulative, annual member and employer contributions. The effect is particularly strong on the difference in total cumulative contributions to KiwiSaver and indicates that, on average, a one day difference in enrolment dates corresponds to a \$6.65 difference in total cumulative contributions to KiwiSaver. For annual member and employer contributions, the average effect of a day difference between enrolment dates is 68 and 27 cents respectively.

Earlier analysis shows that the majority of individuals reported an enrolment date that preceded their administrative enrolment date by up to 90 days. Combining this information with the results from regression analysis suggests that for the bulk of individuals the difference in dates will on average result in SoFIE overestimating total cumulative contributions by up to \$598 relative to administrative data. For the difference in annual member and employer contributions, 90 day difference in enrolment dates will result in SoFIE overestimating the administrative measures by \$61 for annual member contributions and \$24 annual employer contributions.

**Table 24 – Regressions on the difference in total cumulative, member and employer contributions**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
Difference in Dates	-6.6464*** (0.2898)	-0.6824*** (0.0742)	-0.2703*** (0.0359)
Difference in Member Rates	470.3827*** (41.5039)	267.1301*** (21.7641)	- -
Difference in Earnings	0.04884*** (0.0072)	0.0246549*** (0.0029)	0.0125991*** (0.0011)
Squared Difference in Earnings	-1.50e-07*** (3.99e-08)	-9.57e-08*** (1.87e-08)	-1.49e-08 (9.97e-09)
Difference in Employer Rates	492.286*** (63.6466)	- -	469.5429*** (19.1285)
Income Flag	499.4821*** (190.2184)	201.6086*** (54.9819)	-21.10931 (31.1460)
Extra Flag	1487.419*** (199.5806)	- -	- -
Wave 6	513.9434*** (108.4972)	75.6598 (57.9208)	95.37605*** (22.1684)
Wave 7	34.0359 (59.3401)	-18.2264 (21.7399)	8.416754 (12.5488)
<b>Number of Observations</b>	4948	5902	4931
<b>R-squared</b>	0.4063	0.3368	0.6221

Regression results highlight the importance the differences in member and employer contribution rates play in explaining the observed difference between SoFIE and administrative measures of contributions. With regards to member contribution rates, for

every percentage point by which SoFIE contribution rate exceeds the administrative rate SoFIE, on average, overestimates the value of total cumulative contributions by \$470 and the value of annual member contributions by \$267.

The size of the estimated effect of the difference in member contribution rates on the value of total cumulative contributions is roughly double the estimated effect on the value of annual member contributions. This can be explained by the fact that the value of the estimated effect on total cumulative contributions implicitly includes the value of the Member Tax Credit paid to the member as the result of the higher reported contribution rate. Since the Member Tax Credit matches the value of member contributions up to \$1042.86, it is not surprising that the estimated effect on the difference in total cumulative contributions is roughly twice the size of the effect on the difference in member contributions.

The difference in employer contribution rate appears to have a consistent effect on both total cumulative contributions and employer contributions. The results show that, on average, a 1 percentage point difference in employer contribution rate will result in SoFIE overestimating the value of total cumulative contributions by \$492.29 and overestimating the value of annual employer contributions by \$469.54. The similarity in the estimated coefficients is not surprising, since employer contributions do not carry any additional benefits like the Member Tax Credit.

Regression results also indicate the strong impact extra contributions have on the observed difference in total cumulative contributions. The results show that, on average, members who contribute extra to KiwiSaver have \$1487.42 more in total cumulative contributions than what is predicted by SoFIE.

These results highlight that minor differences between the information reported by survey respondents and the administrative data can result in dramatic differences in annual member and employer contributions as well as in total cumulative contributions. This is particularly true with regards to the differences in member and employer contribution rates and the date of enrolment.

## 7 Discussion and conclusion

---

This paper explored the KiwiSaver information from SoFIE and the IRD datasets. In particular, the paper explored how these datasets differ at aggregate and individual levels. Simple descriptive techniques as well as regression analysis were used to determine the size and the direction of the differences between the two datasets.

The results from descriptive analysis show that SoFIE generally mirrors the patterns observed in the administrative data. However, differences exist in all variables covered by the paper.

In particular, the paper finds that SoFIE membership levels in waves 7 and 8 are lower relative to the levels observed in the administrative data. More detailed analysis reveals that automatically enrolled members are the biggest group among those who appear as members in the administrative data, but do not indicate being a KiwiSaver member in SoFIE. The results also show that active members (those who opted-in via their employer or by choosing a provider) comprise the majority of those who appear as members in the administrative data, but not in SoFIE. Questionnaire routing in SoFIE could have contributed to the observed difference in membership between the two datasets. The fact

that some individuals indicated that they do not have a private superannuation scheme in SoFIE, thus skipping the questions on KiwiSaver, also suggests that members might view KiwiSaver separately from other private superannuation schemes.

The paper also notes the differences that exist in the enrolment dates reported in SoFIE and the administrative data. The results show that only a small proportion of members report an enrolment date that matches the enrolment date in the administrative data. For the majority of cases, individuals reported an enrolment date in SoFIE that precedes their administrative enrolment date by up to 90 days.

Differences in membership levels and enrolment dates between SoFIE and the administrative data appear to be primarily responsible for the observed differences in the aggregate measures of annual member and employer contributions, as well as in total cumulative contributions. Differences in enrolment dates leads to longer reported KiwiSaver membership tenure in SoFIE relative to the administrative data. At the same time, lower KiwiSaver membership in wave 7 and 8 in SoFIE could lead to SoFIE having fewer individuals at the lower end of the contributions distributions. Altogether, these differences appear to contribute to higher median, upper and lower quartile values of member, employer and total cumulative contributions in SoFIE than the corresponding values in the administrative data.

At the individual level, regression results suggest that SoFIE underestimates the values of annual member contributions and total cumulative contributions at the lower end of the contributions distribution relative to administrative data. At the same time, these results also show that SoFIE tends to overestimate the values of annual member and total cumulative contributions at the higher end of the contributions distribution.

For annual employer contributions, regression results for wave 8 indicate the same pattern that is observed with member and total cumulative contributions. However, in waves 6 and 7, regression results indicate that SoFIE overestimates the values of annual employer contributions relative to administrative data at the lower end of the contributions distribution, while considerably underestimating contributions at the higher end.

This is a reverse of the pattern that is observed from the results for annual member and total cumulative contributions. One of the possible explanations for this result could be the fact that some employers were contributing more than the mandatory minimum contribution rate of 1% during waves 6 and 7. In such cases, the value of annual employer contributions in the administrative data will be higher than in SoFIE, since the calculations of employer contributions in SoFIE were made using a 1% assumed employer contribution rate.

The results from this paper highlight some of the challenges and opportunities in the use of administratively linked survey data. Many challenges come from attempting to reconcile what individuals reported about their KiwiSaver in SoFIE with the information in the administrative data. An obvious example comes from the difference in enrolment dates. The administrative enrolment date can be considered to be the correct date at which the individual joined the Scheme, but might not accurately reflect the way individuals view their membership tenure.

On the other hand, the combination of survey and administrative data provide a number of opportunities. SoFIE data contains information on individuals' assets and liabilities as well as a variety of socio-demographic data. This information greatly enhances the descriptive power of the administrative data and can enable a wider range of analysis to be conducted to evaluate the effects of KiwiSaver on savings. At the same time, individually

linked administrative data can offer greater precision and choice for variables that are available in both data sets. It can also help to fill in any missing information for cases where individuals might have forgotten or refused to provide certain details.

Overall the advantages offered by the combination of survey and administrative data could lead to better-quality analyses of the effects of KiwiSaver on savings than are possible if the evaluation relies exclusively on survey data.

## References

---

Kim, C., & Tamborini, C. (2009). *Earnings Differences between Survey and Administrative Data: Nonlinear Measurement Error*. Paper presented at the American Sociological Association Annual Meeting (2009).

Law, D., Meehan, L., & Scobie, G. M. (2011). *KiwiSaver: An Initial Evaluation of the Impact on Retirement Saving*. *New Zealand Treasury Working Papers*.