Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Non-response Bias in Household Inflation Expectations Surveys

CAMA Working Paper 28/2023
June 2023

Meltem Chadwick
Reserve Bank of New Zealand

Rennae Cherry
Reserve Bank of New Zealand

Jaqueson K. Galimberti
Asian Development Bank
KOF Swiss Economic Institute, ETH Zurich
Centre for Applied Macroeconomic Analysis, ANU

Abstract

This paper uses micro-data from the Reserve Bank of New Zealand's Household Inflation Expectations survey to obtain a more accurate read of households' true inflation expectations by understanding how different demographic groups respond (or do not respond) to specific questions in the survey. Using a Heckman selection model, we assess whether there is item non-response bias in the survey by comparing the demographic characteristics of responders and non-responders. We quantify and demonstrate how to adjust for bias in aggregate (mean) measures of inflation expectations caused by item non-response. We show that there is a positive bias, and the aggregate inflation expectation series shifts down after the adjustment.

Keywords

inflation expectations, household surveys, item non-response, demographic heterogeneity

JEL Classification

C83, D84, E31, E71

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

Non-response Bias in Household Inflation Expectations Surveys*

Meltem Chadwick[1], Rennae Cherry[1], and Jaqueson K. Galimberti[2,3,4]

[1]Reserve Bank of New Zealand
[2]Asian Development Bank
[3]Centre for Applied Macroeconomic Analysis, Australian National University
[4]KOF Swiss Economic Institute, ETH Zurich

This draft: June 9, 2023

Abstract

This paper uses micro-data from the Reserve Bank of New Zealand's Household Inflation Expectations survey to obtain a more accurate read of households' true inflation expectations by understanding how different demographic groups respond (or do not respond) to specific questions in the survey. Using a Heckman selection model, we assess whether there is item non-response bias in the survey by comparing the demographic characteristics of responders and non-responders. We quantify and demonstrate how to adjust for bias in aggregate (mean) measures of inflation expectations caused by item non-response. We show that there is a positive bias, and the aggregate inflation expectation series shifts down after the adjustment.

Keywords: inflation expectations, household surveys, item non-response, demographic heterogeneity.

JEL classifications: C83, D84, E31, E71.

*Paper presented at the 2023 BIS Annual Workshop of the Asian Research Network. We thank Seunghyeon Lee for an excellent discussion of our paper and other participants of the workshop for many insightful comments. The views expressed in this paper are those of the authors and do not necessarily represent the views of their corresponding institutional affiliations. Corresponding author: jgalimberti@adb.org (Jaqueson).

1 Introduction

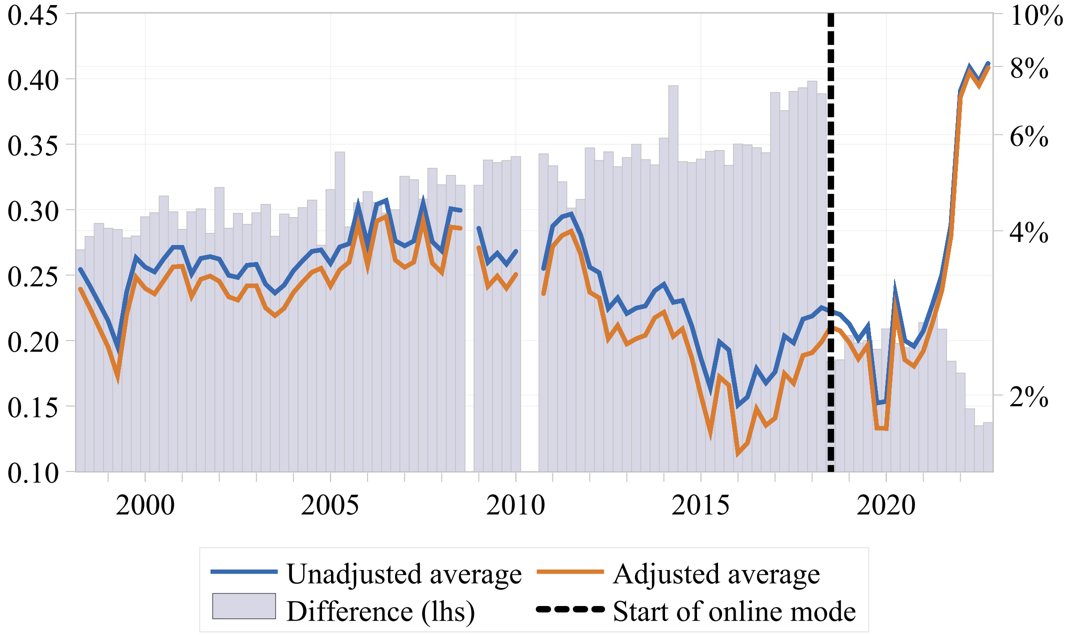
Measures of inflation expectations are of paramount importance to monetary policy. Inflation expectations are key determinants of prices through the forward-looking behaviour of households and firms while also underpinning whether beliefs are anchored to the central bank's inflation target. Surveys are the most common instrument used to measure inflation expectations. Whereas the focus is on aggregate measures, such as averages and medians, it is important to understand the accuracy of survey data in tracking the population's beliefs about inflation. In this paper, we use micro-data from the Reserve Bank of New Zealand's (RBNZ) Household Inflation Expectations survey to investigate how non-responses to inflation expectations questions can bias measurements obtained from such surveys.[1]

Our main contribution is to show that certain demographic groups tend to non-respond more frequently than the target population when asked about their inflation expectations. These non-responses amount to about 44% on average throughout our sample: young female respondents who have low income and come from ethnic minority groups end up underrepresented due to non-responses. Because these non-responses are not random, aggregate measures of inflation expectations derived from the sample of respondents can be biased. We quantify and show how to adjust for non-response bias in inflation expectations using a Heckman sample selection model. Figure 1 presents the evolution of average one-year-ahead inflation expectations throughout our sample period, from 1998Q2 to 2022Q4. We find that non-responses artificially raise average inflation expectations by about 0.30 percentage points.

Another important finding relates to the way the survey was conducted before 2018. Starting in 2018 Q3, the survey changed from being conducted by telephone to online mode. We find that this change significantly affected the incidence of non-response to the inflation expectations question. The average of non-responses decreased to about 24% since the survey moved to online mode. As evidenced in Figure 1, this change also significantly reduced the effect of non-response bias in estimating average inflation expectations. In other words, con-

[1] Throughout this paper, our focus is on so-called item non-responses to specific survey questions, as opposed to unit non-responses to the whole survey. The use of survey weights corrects the incidence of unit non-response (see Meyer et al., 2015, for further discussion).

Figure 1: *The RBNZ Survey of Household Inflation Expectations.*



Notes: The lines depict quarterly averages of one-year-ahead inflation expectations from the RBNZ household inflation expectations survey. The unadjusted average is a raw weighted average across respondents, while the adjusted average is calculated using our methodology to adjust for non-response bias. The gaps in 2008Q4 and 2010Q2/Q3 are due missing observations. The dashed line depicts the period at which the survey switched to online mode in 2018 Q3.

ducting the survey online seems to have made it more inclusive for previously underrepresented demographic groups.[2]

Our methodological approach is based on sample selection models. We first identify potential determinants of responses to the expectations question by estimating Probit regressions on several demographic variables collected with the survey. Probit regressions model the probability of an event, in our case response to the inflation expectations question, using a set of explanatory variables. These estimates help us define a selection equation, which determines when a respondent is likely to answer the inflation expectation question depending on their characteristics.[3]

We then study inflation expectations bias accounting for non-responses using a Heckman selection model (Heckman, 1974, 1979). The Heckman correction is based on the insight that sample selection can be viewed as a form of omitted variable bias – specifically, the method

[2] One possible contention to conducting the survey in online mode is that respondents may use online search tools to find, say, the latest forecast by the central bank, which could lead to less dispersion in expectations and perhaps less informative data. However, in our data we observe that the dispersion of inflation expectations as well as the number of outliers actually increased following the move to online mode.

[3] Our regressions include additional 'macro' variables, such as lagged inflation and lagged inflation squared, a yearly trend, seasonal dummy variables, and a dummy variable accounting for the change to online mode.

draws on Probit estimates of the selection equation to calculate the inverse Mills ratio, which is then used as an additional explanatory variable in the regression with missing observations.[4]

Our proposed adjustment to the calculation of average inflation expectations goes along similar lines: average indices can be easily obtained by running a regression of survey inflation expectations on quarter dummy variables; after including our baseline estimates of the Heckman correction term as an additional variable in this regression, we obtain average inflation expectations adjusted for non-response bias. Finally, our estimates use survey weights to account for unit non-response bias arising from difficulties in obtaining a representative survey sample of the population. Although these weights can not account for determinants of non-responses to the inflation expectations question, we also find that they are relevant for the analysis of inflation expectations bias.

Our findings about the under-representation of some demographic groups are consistent with previous studies in the literature. Exploring UK survey micro-data, an early study by Blanchflower and MacCoille (2009) also found significant non-response bias from young, female, and low-income respondents. Leung (2009) reported similar findings with a shorter sample from the RBNZ household survey. Looking at a sample from the US Michigan Survey of Consumers, Bruine de Bruin et al. (2010) corroborate findings that demographic variables play a significant role in determining inflation expectations. Our finding that online survey mode can attenuate non-response bias is consistent with previous studies. Bruine de Bruin et al. (2017), for example, find that online surveys achieve higher response rates to the inflation expectations question than face-to-face surveys.

None of the papers above provided an adjustment for the non-response bias in inflation expectations surveys. One standard approach to deal with missing observations is imputation. The U.S. Michigan Survey of Consumers (MSC), for example, uses distribution-based imputations to replace "Don't Know" responses with random draws from a distribution that matches the properties of observed data (Curtin, 1996). However, this imputation method does not consider the socio-demographic composition of the sample of respondents. It can, therefore, reinforce the effects of selection bias in the analysis of survey of expectations data.

[4]The Heckman sample selection model can be estimated using either a maximum likelihood approach or the original two-step approach described in the text (see, e.g., Puhani, 2000, for more details). Our estimates are robust to the choice of estimation method.

More broadly, the issue of item non-response has received increased attention in recent related studies. Focusing on a US longitudinal survey of professional forecasters, Bürgi (2023) compares methods for filling in missing observations due to survey attrition – naturally, this is a different problem than what we face with repeated cross-sectional surveys as the one we study here. An alternative approach for that case involves the use of survey design features. McGovern et al. (2018) explore HIV testing data to show that randomized incentives or survey interventions can provide ideal selection variables to correct for non-response bias. Comerford (2023) proposes using a verbal question to deal with non-response bias found in inflation expectations derived from density forecasts. Ex ante, these methods provide vital insights to survey design. However, the required survey features are rarely available for long-running surveys. Our approach offers a potential solution to these cases.

Finally, non-response bias can be important for other household surveys of inflation expectations. For example, in the U.S. MSC, a major survey in this area, non-responses to the inflation expectations question amounted to an average of 9% of the monthly samples collected between 1978 and 2022; non-responses also varied substantially over time, ranging from lows of about 3%, mostly observed in 1985 and 2022, to highs of about 25% observed in 1978.[5] Another example is the Bank of England's (BoE) Inflation Attitudes Survey, where non-responses to the inflation expectations question amounted to an average of about 15% of the quarterly samples collected between 2001 and 2022, and ranging from 8% to 25% over the period.[6,7] Our methodology can be easily applied to analysis and adjustment of these and other country surveys using their corresponding socio-demographic information on the surveyed households.

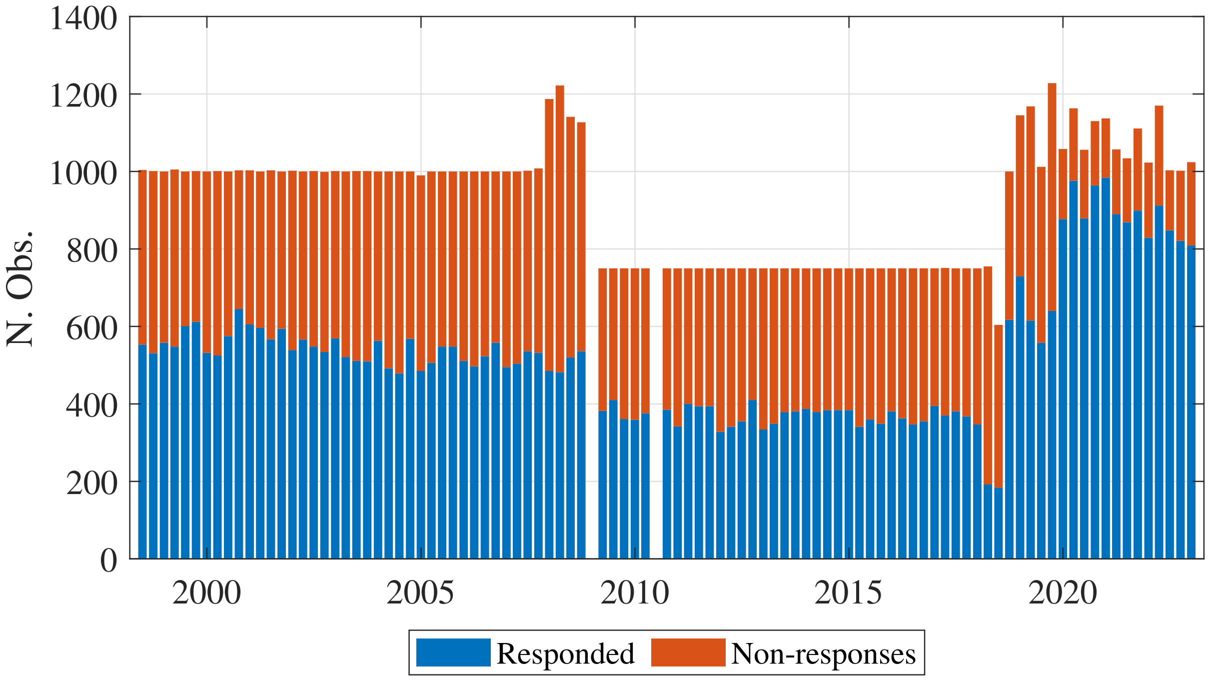
The remainder of this paper is organized as follows. Section 2 provides details about the survey and sample statistics. Section 3 analyzes potential determinants of non-responses to the inflation expectation question using Probit models. Section 4 focuses on estimates of inflation expectations bias and outlines our proposed approach to account for non-responses using a sample selection model. Section 5 shows how the correction for non-responses can be used

[5] The MSC also distinguishes between respondent's refusing to answer the inflation question (0.14% of the 1978-2022 sample, on average), those that don't know a point estimate neither the direction of change (0.84%, on average), and those that can answer about the direction but not about a point estimate (7.53% DK UP and 0.54% DK DOWN, on average), discarding the first two and imputing the last for index calculations.

[6] These statistics on the BoE survey exclude the data from 2020/Q2, when non-responses declined to 0.12% due to a design issue in the switch to online mode during the COVID-19 social distancing guidance. The issue was that the option of "Don't know/No idea" appeared only if the respondent tried to move on to the next question without answering the inflation expectations question. In subsequent surveys, the "Don't know/No idea" option was reintroduced with the other options for the question, and non-responses returned to usual levels.

[7] Intriguingly, the more recently launched ECB Consumer Expectations Survey does not allow the respondent to proceed with the survey without giving an answer to the inflation expectations question, which is a potential design flaw as discussed in the previous footnote.

Figure 2: *Responses and Non-responses to Inflation Expectation Question.*



Notes: The bars depict the number of survey responses collected across the quarters, decomposed by response/non-response to the inflation expectations question.

to adjust indices of average inflation expectations. Finally, section 6 concludes with some remarks.

2 Data

2.1 Survey Design

The Household Inflation Expectations survey is conducted every quarter and achieves approximately 1,000 household responses per quarter. Our sample covers the period from 1998 to 2022 and contains a total of 89,834 individual responses. The individual responses are anonymised every quarter. Hence survey waves are treated as repeated cross-sectional data. The survey goes into the field after the previous quarter's consumer price index inflation data have been released. The survey asks for households' perceptions of current inflation and expected inflation at varying horizons starting one year ahead. The inflation expectation question we focus on in this paper is formulated as follows: "As a percentage, what do you think will be the annual rate of inflation/deflation in the next 12 months?" Figure 2 illustrates that approximately half of the respondents choose not to answer the inflation expectation question.

Figure 2 also shows that the number of responses collected by the survey changed across time, ranging from a minimum of 604 responses in 2018 Q2 to a maximum of 1,228 in 2019 Q3. Another important change relates to the way the survey was conducted. Up until 2018 the survey was conducted by telephone. Starting in 2018 Q3, the survey changed to online mode. As evidenced in Figure 2, this change significantly affected the incidence of non-response to the inflation expectations question. The average of non-responses decreased to about 24% since the survey moved to online mode.

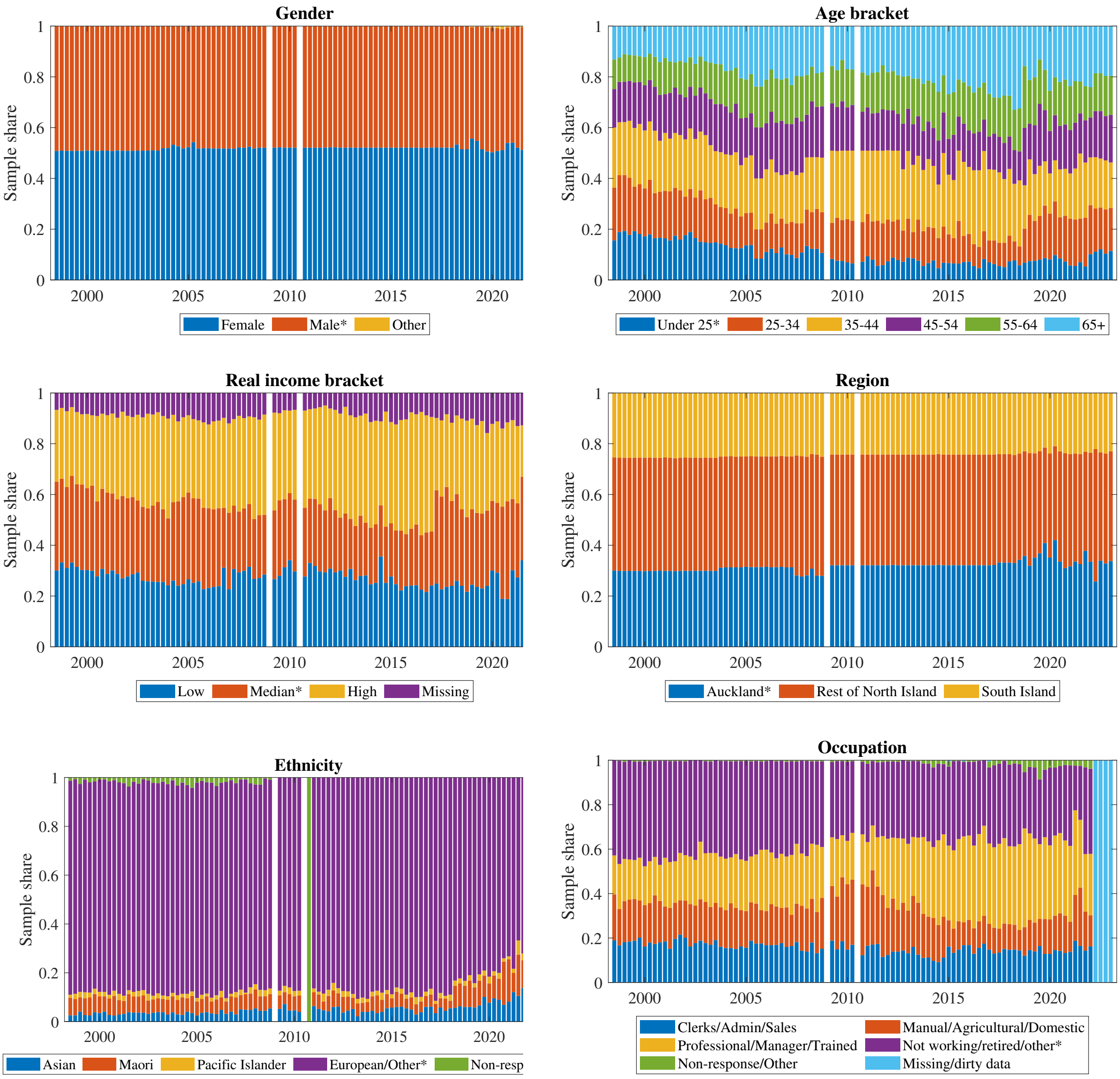
The dataset contains a variety of demographic information about survey respondents including, age, gender, ethnicity, income, employment, occupation, region, children in the household, number of adults in the household, marital status, home ownership status, and grocery shopping.[8] While the survey has been redeveloped over time to improve data quality and better align with international best practice for capturing household inflation expectations, the data are limited by changes in collection and measurement. The most common evolution in the measurement of variables over time is the move from granular to less granular levels. In such cases, we merge granular observations into less granular categories to provide consistency.

For example, from 1998 Q1 to 2008 Q3, Pacific Island respondents could identify as one of six Pacific Island ethnic groups: Cook Island, Niuean, Fijian, Tongan, Samoan, or other Pacific Islands. In 2008 Q4, the six Pacific Island ethnic groups were aggregated into a single ethnic umbrella group, leaving Pacific Island respondents with one response option, specifically, Pacific Islanders. To maintain consistency, respondents who identified with one of the six Pacific Island ethnic groups from 1998 Q1 to 2008 Q3 were grouped and categorised following the umbrella Pacific Islander ethnic group.

As another example, we construct a real household income variable based on the availability of different nominal income variables across the sample. Subject to availability, we calculate household income using the median values of the more granular household income bracket intervals. We then adjust the nominal household incomes to real household incomes using the 2022 Consumer Price Index as the base level. Following this, we classify real household incomes into one of 3 categories: under $50,000 (low), $50,000 - $100,000 (median), and over $100,000 (high). In the absence of nominal household income, real personal income is used to

[8] See Table A.1 in Appendix A for tabulations of these data and further discussion in the next subsection.

Figure 3: *Evolution of Sample Compositions by Selected Variables.*



Notes: The bars depict the survey sample composition by selected variables across the quarters. Categories with a * indicate the base categories used in the regression analysis.

fill in the missing values (which are calculated similarly). This method assumes that if nominal household income was missing, but real personal income was low/median/high, real household income would have been low/median/high, respectively.[9] Figure 3 presents the evolution of our sample compositions according to different demographic variables.

[9]Upon further inspection of the observations with available data on both household and personal income we find that these two are highly correlated: the Phi coefficient of correlation based on the contingency table between these two variables equals to 0.702 and the hypothesis of independence is strongly rejected (p-value=0).

2.2 Sample Statistics

Over the period that the survey was conducted by telephone, 1998 Q2 to 2018 Q2, 51.75% of respondents were female, 5.8% were Māori, 4.07% Asian, 1.87% Pacific Islander, and 88.26% identified with 'other' ethnic groups. Among the survey respondents, 14.8% were aged 25-34, 23.54% were 35-44, 16.95% were 45-54, 14.68% were 55-64, 18.62% were 65+, and 11.41% were under 25.

Following the move to online mode in quarter 3 of 2018, we observe a redistribution of demographics among respondents. Respondents who identified with Asian ethnicity increased by 117.19%, Māori respondents by 70.17%, Pacific Islanders by 33.69%, and respondents who identified with other ethnicity groups decreased by 10.73%. Survey respondents between the ages of 25-34, 45-54, 55-64, and 65+ increased by 18.04%, 7.14%, 10.76%, and 13.64%, respectively. Contrary to this, respondents under 25 decreased by 31.55%, while those aged 35-44 decreased by 18.69%. Furthermore, female respondents increased marginally by 0.89%.

Over the whole sample, 1998 Q2 to 2022 Q4, the median age range of survey respondents was 45 to 54. In total, 51.85% of respondents were female, 6.68% were Māori, 5.1% Asian, 2% Pacific Islander, and 86.21% identified with 'other' ethnic groups. Over $100,000 was the most frequently reported real household income range, followed by $50,000 to $100,000, with 34.37% and 27.68% of respondents reporting real household incomes within those ranges, respectively. Regarding the real personal income of survey respondents, 37.99% reported incomes under $40,000, 31.99% reported incomes between $40,000 to $80,000, and 19.75% reported incomes over $80,000.

The (measured) average one-year ahead inflation expectation of female respondents has historically been higher than that of male respondents. However, since 2021, the (measured) average one-year-ahead inflation expectation of males has marginally surpassed that of females. Similarly, respondents aged 35 and under have traditionally had higher (measured) average one-year-ahead inflation expectations compared to respondents aged 35 and over. In recent years, however, this trend has inverted, resulting in lower inflation expectations among respondents under 35 compared with those over 35. Likewise, the (measured) average one-year-ahead inflation expectation for Pacific Islander respondents is higher than for Māori, Asian and other ethnic groups.

Figure 4 shows how the composition of responses and non-responses to the inflation expectation question compares across some selected population groups. There is a case for sample selection bias by variables such as age, gender and ethnicity. Note how the group compositions change across participants that responded versus those that non-responded to the inflation question. This non-random incidence of item non-responses undermines the accuracy of the survey as a representative depiction of the inflation beliefs of New Zealand's population.

2.3 Sample Availability and Outliers

Figure 5 illustrates changes to survey data collection over time, with the green shaded areas reflecting available observations by variable. The most notable omission is the data for the fourth quarter of 2008 and the second quarter of 2010, which was missing from the data provider database. Some variables are only available for portions of our sample. This is important for our regression analysis in the following sections, as it will determine the number of observations available for each specification. E.g., groceries shopping was only available up to 2008, while home ownership entered the survey after that point. Hence, we can only evaluate their effects separately according to the corresponding sub-samples.

Finally, outlier responses are commonplace in survey instruments and can have a significant impact on estimates derived on the basis of this type of data. The method for outlier detection used in the construction of aggregate indices from the RBNZ household survey of inflation expectations also changed across time. Particularly, starting from 2022 the methodology shifted from a fixed (-2,15) rule, which discarded inflation expectations lower than -2% and higher than 15%, to an interquartile range (IQR) method based on a whisker equal to 1.5.[10]

Table 1 shows how different outlier detection rules compare for our sample. There is a clear issue with consistency of the previous rules between the two sub-periods divided by the 2022 change. While the (-2,15) rule turned too restrictive from 2022, the new 1.5×IQR rule implied a higher number of outliers in the pre-2022 period, and still discards a large number of observations post-2022. Hence, these rules seem too restrictive, especially considering that

[10] The IQR method excludes observations falling outside the following limits:

$$Lower\, limit = Q1 - whisker \times IQR,$$
$$Upper\, limit = Q3 + whisker \times IQR,$$

where $Q1$ refers to the 25th percentile, $Q3$ refers to the 75th percentile, and IQR is the difference between $Q1$ and $Q3$ in the data series.

Figure 4: *Responses and Non-responses to Inflation Expectation Question by Group.*

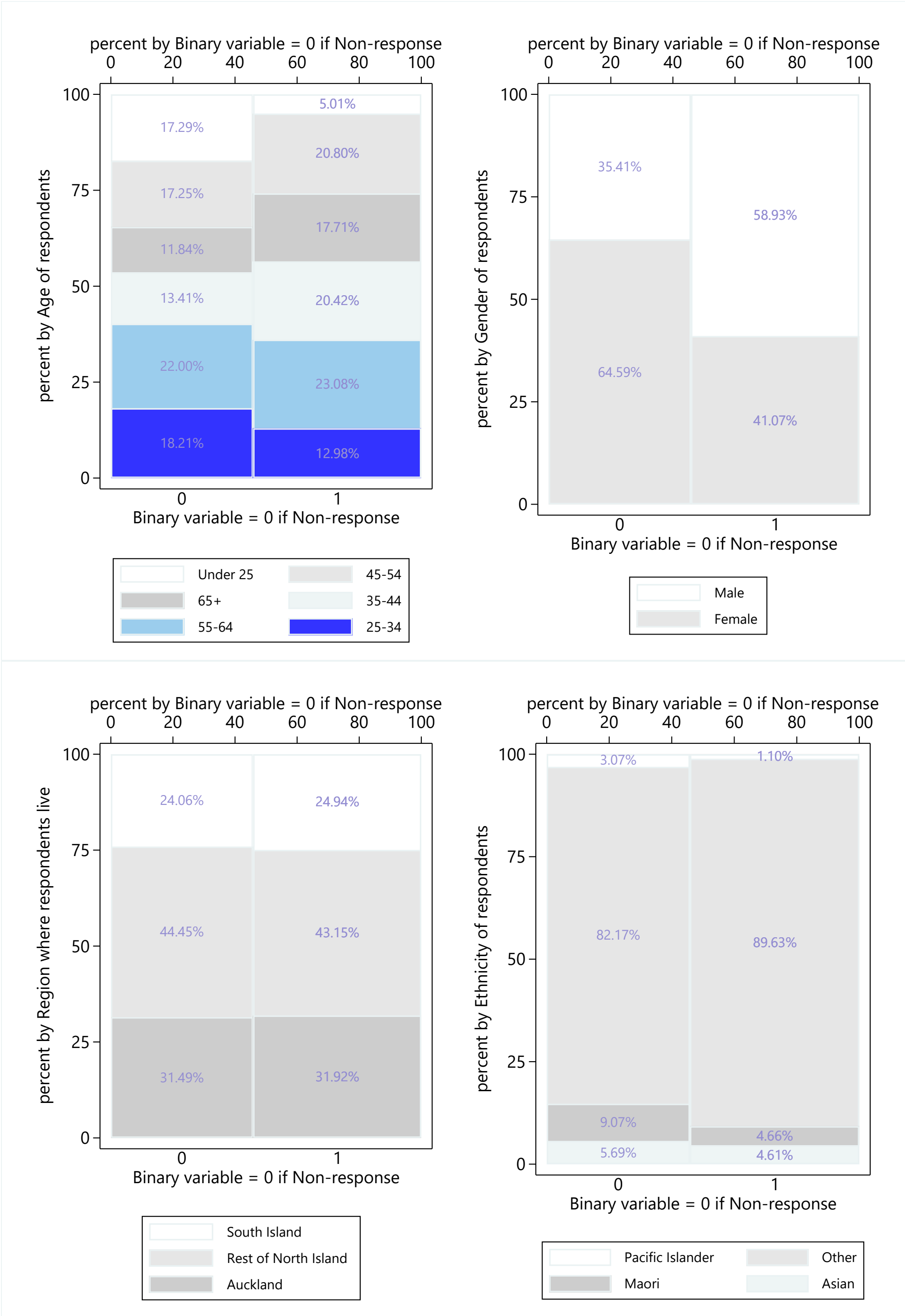
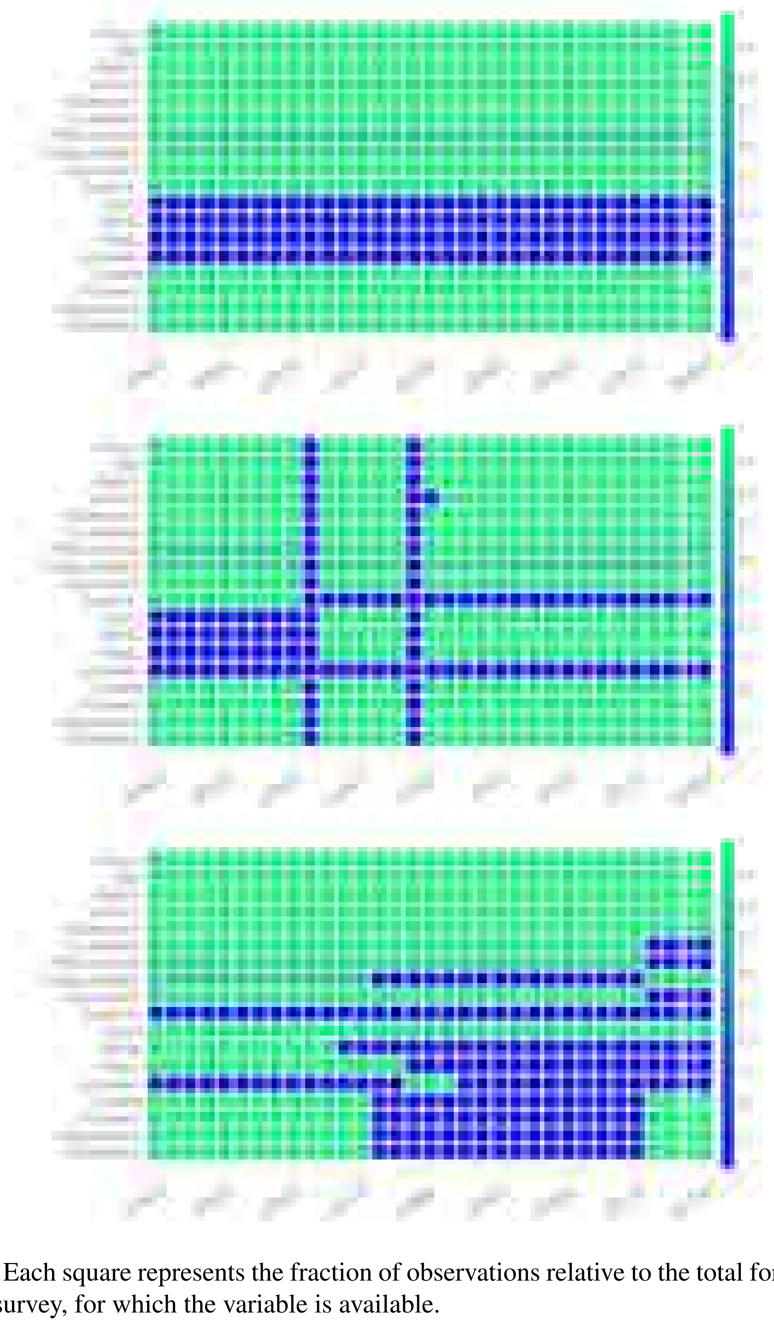


Figure 5: *Available Sample Observations by Variable.*



Notes: Each square represents the fraction of observations relative to the total for the corresponding quarter's survey, for which the variable is available.

Table 1: *Frequency of Outliers by Detection Rule.*

Sample	Rule			
	(-2,15)	**±1.5×IQR**	**±3×IQR**	**Hybrid IQR**
1998Q2-2021Q4	1.84%	4.80%	2.48%	4.80%
2022Q1-2022Q4	18.55%	16.19%	10.19%	10.19%

Notes: The statistics are average frequency of outliers detected according to the different rules for the samples indicated in the first column. (-2,15) discards inflation expectations lower than -2% and higher than 15%. IQR stands for the interquartile range, and these rules discard inflation expectations lower than $Q1 - whisker \times IQR$ and higher than $Q2 + whisker \times IQR$, with varying *whisker* values. The hybrid IQR rule uses a $whisker = 1.5$ for the sample from 1998Q2 to 2021Q4 and a $whisker = 3$ for the sample from 2022Q1 to 2022Q4 in order to attenuate the difference in outlier frequency across these subperiods.

inflation and inflation expectations have been relatively high over the end-of-sample period. To address this issue we applied a hybrid rule, adjusting the *whisker* to equal 3 from 2022, while keeping the current IQR rule for the pre-2022 period. By using the hybrid IQR method, a greater degree of consistency in the number of outliers excluded across the sample is achieved. In total, 4,535 observations were excluded from the analysis.

3 Determinants of Responses to the Inflation Question

As a first step in finding the non-response bias and correcting for it, we study the determinants of each household's non-response behaviour. To do so, we estimate Probit models that relate the choice of response to the household's demographic characteristics. This part of the analysis relates to Blanchflower and Kelly (2008), Blanchflower and MacCoille (2009) and Leung (2009). We evaluate the effects of socio-demographic and personal characteristics from Probit models on the likelihood of response to the quantitative question of inflation expectations for the next 12 months. Equation 1 represents the probit models we estimate.

$$Pr(Response_i = 1 | X_i, Z_i) = \Phi(\beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 D_{Season} + \beta_4 T_{Year} + \beta_5 D_{Online}). \quad (1)$$

Demographic variables are represented by X_i, and macroeconomic variables are represented by Z_i. The dummy variable D_{online} indicates the respondent i took the survey after 2018 Q2, T_{Year} is the year trend and D_{Season} is the quarter seasonal dummy. $Response_i$ is a dummy variable equal to 1 if respondent i answered the question related to inflation expectation. Φ is the standard Normal cumulative distribution function.

We estimate four specifications of such Probit models with different sets of explanatory variables to maximize our sample coverage. Table 2 lists estimates of the average partial effects associated with each of these Probit specifications.[11] Women seem to be less likely to respond to the inflation expectations question, with an average probability of responding about 20% lower than men. Maori and Pacific Islanders are more likely to be in the category of non-response too. In contrast, older individuals, who are employed, and who have higher incomes are more likely to respond. Additionally, conducting the survey in online mode significantly increases the response rates, increasing the probability of response by about 33% – we discuss further about the effects of conducting the survey in online mode below. We also find a significant downward trend in the responses to the inflation question. In spite of the recent increase in responses due to the shift to online mode, and after accounting for that, the long run trend estimates indicate an increase of item non-responses of about 1% per year. This is consistent with more broad evidence of increased rates of both unit and item non-responses in household surveys (see Meyer et al., 2015). These findings are robust across the different specifications.

[11] Estimates of the coefficients of equation 1 are presented in Table A.2 in Appendix A.

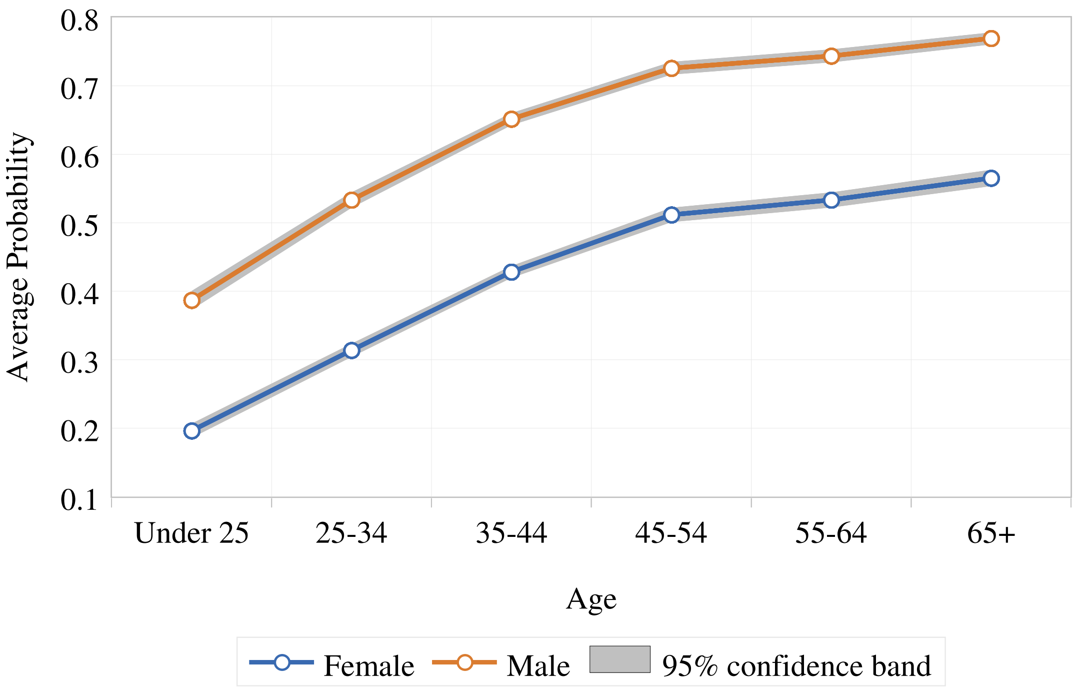
Table 2: *Average Partial Effects on Responses to the Inflation Expectation Question.*

Variables	Categories	(1)	(2)	(3)	(4)
Gender	Female	-0.211***	-0.223***	-0.235***	-0.199***
		(0.004)	(0.004)	(0.007)	(0.005)
Age	25-34	0.131***	0.131***	0.168***	0.058***
		(0.008)	(0.008)	(0.011)	(0.013)
	35-44	0.246***	0.255***	0.321***	0.134***
		(0.007)	(0.008)	(0.011)	(0.013)
	45-54	0.324***	0.327***	0.393***	0.194***
		(0.008)	(0.008)	(0.011)	(0.013)
	55-64	0.343***	0.337***	0.405***	0.186***
		(0.008)	(0.008)	(0.011)	(0.013)
	65+	0.372***	0.353***	0.371***	0.207***
		(0.007)	(0.008)	(0.012)	(0.013)
Region	Rest of North Island	-0.012***	-0.012***	-0.020***	-0.010*
		(0.004)	(0.005)	(0.006)	(0.006)
	South Island	-0.001	-0.000	-0.001	-0.008
		(0.005)	(0.005)	(0.007)	(0.007)
Ethnicity	Asian	-0.053***	-0.055***	-0.035**	-0.060***
		(0.009)	(0.010)	(0.016)	(0.012)
	Maori	-0.152***	-0.149***	-0.185***	-0.118***
		(0.007)	(0.008)	(0.012)	(0.010)
	Pacific Islander	-0.192***	-0.192***	-0.222***	-0.166***
		(0.013)	(0.014)	(0.022)	(0.018)
Real Income	High	0.104***	0.096***	0.108***	0.096***
		(0.004)	(0.005)	(0.007)	(0.006)
	Low	-0.089***	-0.092***	-0.096***	-0.089***
		(0.005)	(0.005)	(0.008)	(0.007)
Employment	Employed	0.026***			0.017**
		(0.005)			(0.007)
Occupation	Clerks/Admin/Sales		-0.004	-0.041***	
			(0.006)	(0.008)	
	Manual/Agricultural/Domestic		-0.017***	-0.034***	
			(0.006)	(0.008)	
	Professional/Manager/Trained		0.055***	0.050***	
			(0.005)	(0.008)	
Dependent Children	Yes		-0.031***	-0.032***	-0.019***
			(0.004)	(0.006)	(0.007)
Groceries Shopping	Yes - jointly/shared			0.013	
				(0.010)	
	Yes - main			-0.025***	
				(0.008)	
Home Ownership	Living with parents				-0.224***
					(0.013)
	Mortgage				-0.077***
					(0.007)
	Other				-0.114***
					(0.016)
	Renting				-0.129***
					(0.008)
Online	Yes	0.335***	0.338***		0.336***
		(0.001)	(0.001)		(0.001)
Year Trend		-0.009***	-0.010***	-0.014***	-0.004***
		(0.000)	(0.000)	(0.000)	(0.001)
Lagged Inflation		0.002	0.001	0.008***	0.006*
		(0.001)	(0.002)	(0.003)	(0.003)
N.Obs.		75,400	71,799	36,011	35,993
Sample		98Q2-22Q4	98Q2-21Q4	98Q2-08Q3	09Q1-21Q4
McFadden R^2		0.170	0.170	0.185	0.171

Notes: Average partial effects are calculated using the delta method and averaging over the sample observations holding other variables constant at their sample values. Numbers in parenthesis are robust standard errors. ***,**,* stand for 1%, 5%, 10% statistical significance, respectively.

Figures 6 and 7 summarize some of the results illustrated in the first column of Table 2. Accordingly, gender is an important determinant of the response probability: males are more likely to respond and this probability increases with age. Ethnicity is also an important determinant of response probability: a significant lower response rate is attached to being a Maori or a Pacific Islander.

Figure 6: *Predicted Probabilities of Response by Age and Gender.*

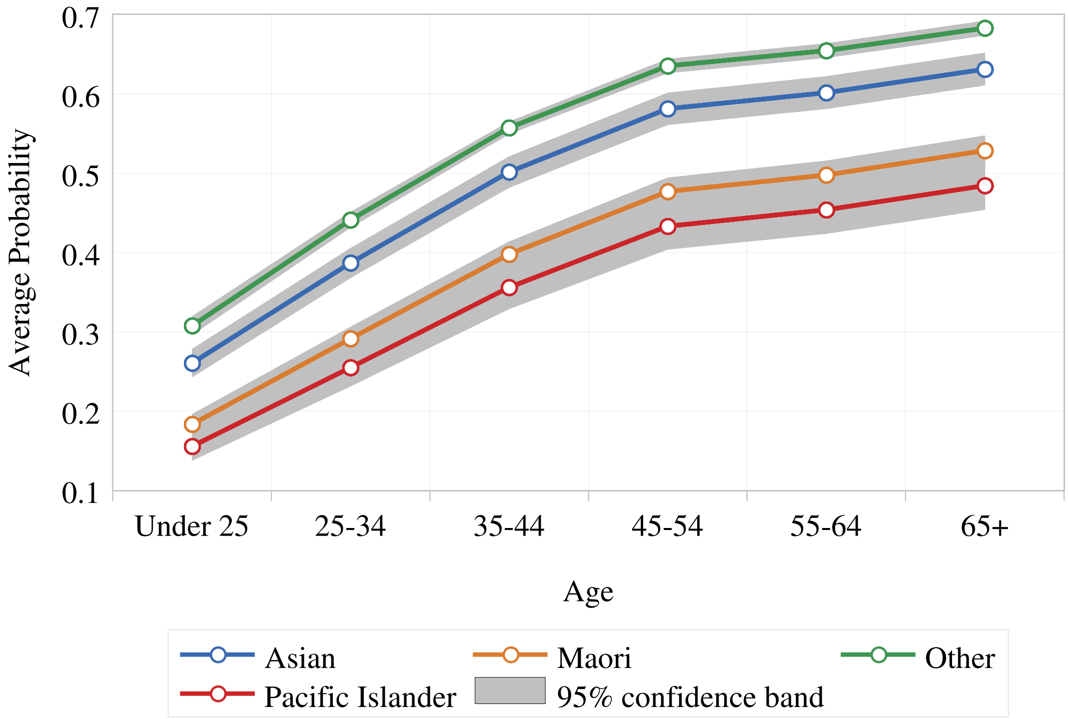


Notes: Response probabilities are based on the baseline Probit model estimates, column (1) from Table A.2, and are obtained as the average of predicted probabilities holding other variables constant at their sample values.

Figure 8 illustrates the link between lagged inflation rates and the predicted probabilities of response. In our Probit estimations we have found evidence of a quadratic relationship between responses and lagged inflation rates.[12] Intriguingly, the averaged nonlinear effects of lagged inflation over our sample turn out mostly insignificant, as reported in Table 2. However, this is a misleading artifact of averaging. As we can see from Figure 8, lagged inflation has an interesting profile of regime-dependent effects on responses. Particularly, when the lagged inflation hits the range of 5-7% the slope of the response probability turns steeply positive. This may imply that when inflation moves out of the zone of "rational inattention", where inflation

[12]We used lagged values of the inflation rate because these are the latest available information to respondents at the time the surveys are conducted.

Figure 7: *Predicted Probabilities of Response by Age and Ethnicity.*



Notes: Response probabilities are based on the baseline Probit model estimates, column (1) from Table A.2, and are obtained as the average of predicted probabilities holding other variables constant at their sample values.

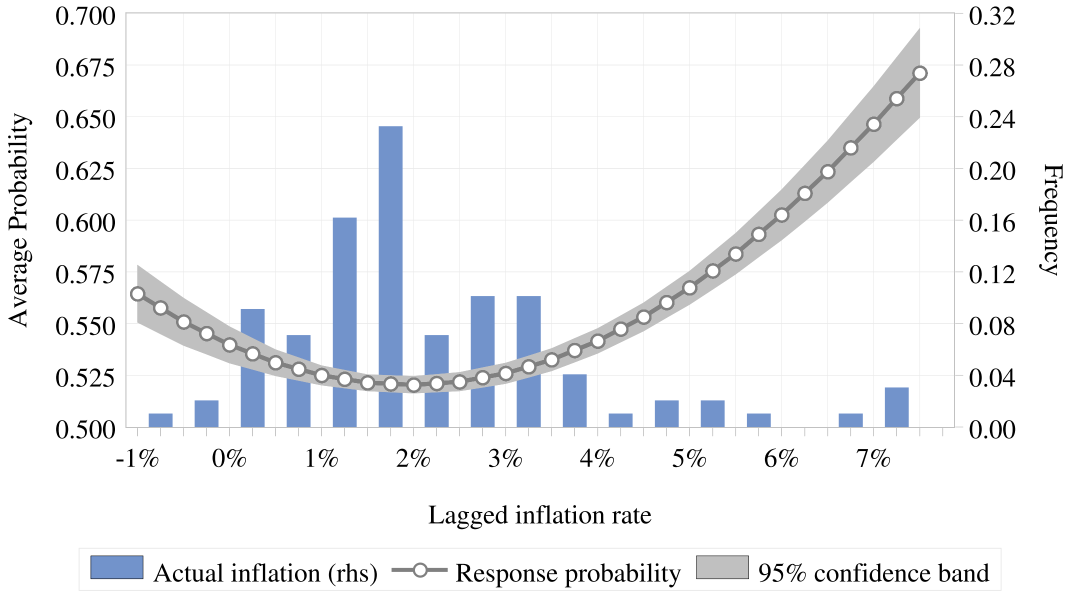
is hardly noticed by economic agents, price changes snap into sharp focus and increase agents' ability/willingness to respond to the inflation expectations question.[13]

The estimates using the extended specifications (columns 2-4 in Table 2) provide additional insights into the causes of (non-)responses to the inflation expectations question. First, only the more skilled occupation, covering professionals, managers and trained service workers, showed probability of response greater than the base group of unemployed/retired/others. As depicted in Figure 3, that occupation also showed the greatest increase in the survey sample composition, which explains why the employment variable (roughly averaging over occupations) showed a positive effect in columns (1) and (4) of Table 2.

Households with dependent children and those not owning a freehold house (base category for home ownership) tended to respond less to the inflation question, while the effect of groceries shopping was only significant, and negative, when the respondent was the main responsible for that chore. The latter effect, though of a small magnitude of -2.5%, can be an important determinant of the relevance of the beliefs elicited by the survey. Previous research

[13] For related research on this see, e.g., Borio et al. (2023).

Figure 8: *Predicted Probabilities of Response by Lagged Inflation Rate.*



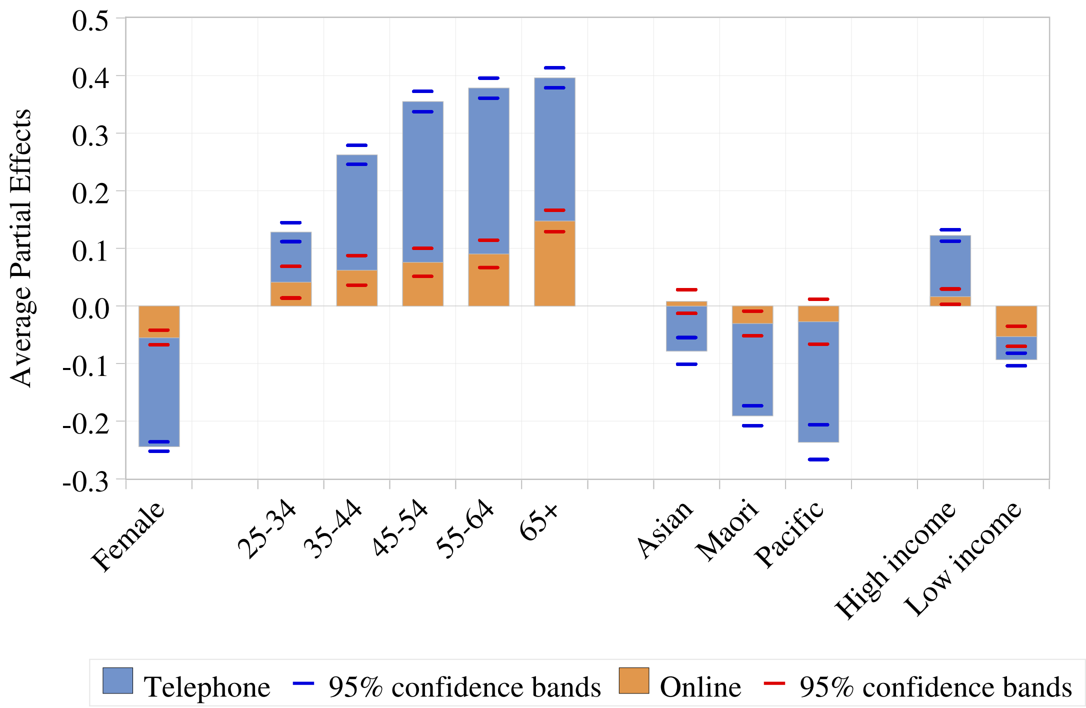
Notes: The blue bars depict the distribution of actual inflation rates over the sample period. Response probabilities are based on the baseline Probit model estimates, column (1) from Table A.2. They are obtained as the average of predicted probabilities holding other variables constant at their sample values.

suggests exposure to grocery prices as an important determinant of consumers' belief formation (D'Acunto et al., 2021; D'Acunto et al., 2023).

Finally, we also look at how conducting the survey in online mode can change the effects of the demographic characteristics on response probabilities. In order to do that, we re-estimate our Probit model specifications with additional terms interacting the online dummy variable with the demographic variables. A summary of these estimates is presented in Figure 9, focusing on the case extending our baseline specification.

Overall, we find that conducting the survey in online mode reduced the effects of demographic characteristics on the probabilities of response. For example, prior to the switch to online mode, women were 24.4% less likely to respond to the inflation question than men; this difference decreases to only 5.5% since the survey moved to online mode. Similarly, most of the differences by ethnicity turn insignificant after the survey moved online. This evidence indicates that conducting the survey online made it more inclusive for previously underrepresented demographic groups.

Figure 9: *Average Partial Effects of Online Mode for Selected Variables.*



Notes: Average partial effects are obtained holding other variables constant at their sample values. The estimates presented here are based on a re-estimation of the baseline Probit specification, column (1) from Table A.2, splitting the effects by survey mode. See Table A.3 in Appendix A for the corresponding numerical estimates and other specifications.

4 Estimates of Inflation Expectations Bias

In this section, we are interested in evaluating how biased are household inflation expectations. Bias is defined as the average of inflation expectations errors, taking into account the timing of the forecasts and their target realization. More formally,

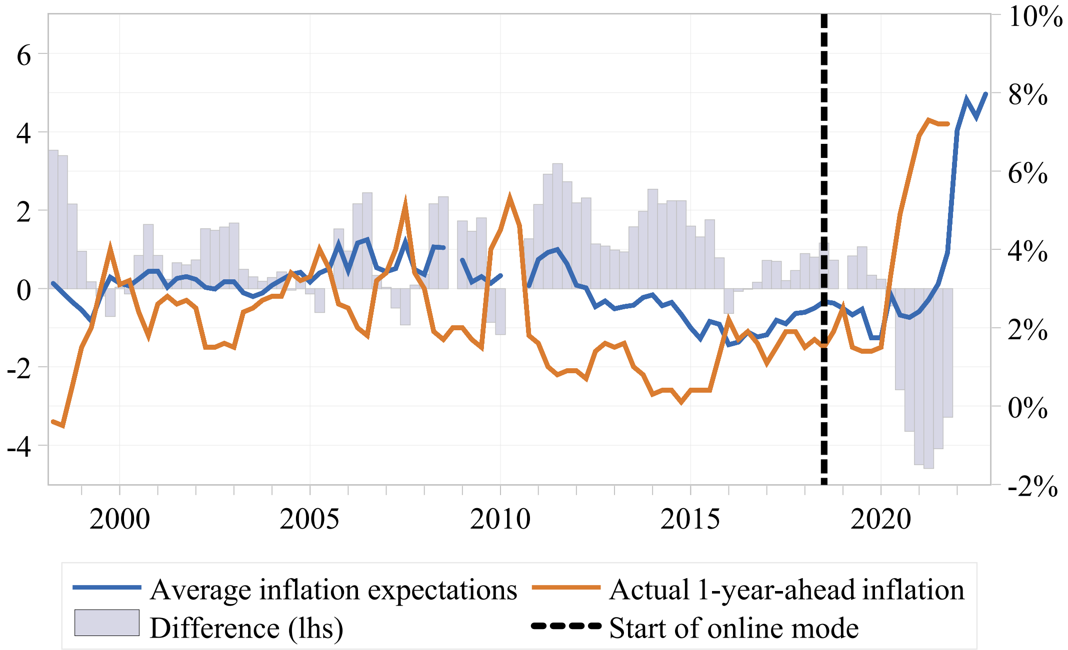
$$Bias \equiv E\left[\pi_i^e\left(\underbrace{t}_{base},\overbrace{t+h}^{target}\right) - \pi_{t+h}\right], \tag{2}$$

where $\pi_i^e(t, t+h)$ is respondent i period t forecast of inflation between t and $t+h$, and π_{t+h} is the actual inflation rate over that same period. In our case, h is 4 quarters. Figure 10 illustrates how the average bias evolved over our sample.

The richness of information available in the micro-data from the survey of households' expectations allows us to go one step further and attempt to understand how inflation expectations bias depends on household characteristics. For that purpose, we run regressions of the following form:

$$\pi_i^e - \pi_{t+1} = \alpha_0 + \alpha_1 X_i + \alpha_2 Z_i + \alpha_3 D_i + u_i, \tag{3}$$

Figure 10: *Average Inflation Expectations and Actual Inflation Rates.*



Notes: The lines depict quarterly averages of one-year-ahead inflation expectations from the RBNZ household inflation expectations survey and the corresponding one-year-ahead actual inflation rate. The gaps in 2008Q4 and 2010Q2/Q3 are due missing observations in the micro dataset. The dashed line depicts the period at which the survey switched to online mode in 2018 Q3.

where, as in our previous notation, X_i contains household characteristics, Z_i contains macroeconomic variables at the time the respondent answered the survey, and D_i contains other controls for trend, seasonality and survey mode.

If the survey data were not affected by sample selection, Equation 3 could be estimated by Ordinary Least Squares (OLS), or Weighted Least Squares (WLS) to account for survey weights. However, as our results from the previous section indicated, the missing responses to the inflation expectation question are not random. In order to account for such selection bias, we estimate inflation bias using the Heckman selection model (Heckman, 1974, 1979). The Heckman selection model is a statistical model that addresses the issue of selection bias in econometric analyses. Selection bias arises when a subset of observations is systematically different from the rest of the sample, leading to biased estimates of model parameters. The Heckman selection model consists of two equations: the selection equation and the outcome equation.

The selection equation models the probability of selection into the sample, while the outcome equation models the relationship between the outcome variable and the explanatory variables. The selection equation is typically a Probit model that relates the probability of being included in the sample to a set of variables that are correlated with the selection process. The selection equation for our case is defined by the extended baseline specification of Equation 1, including the variables in column 1 of Table 2 plus the interaction terms with the online dummy variable. We choose this specification as a selection equation because most of the other variables we used for robustness usually restrict our sample size, whereas inclusion of the online interaction terms provide a more accurate identification of the heterogeneity of response rates.

The outcome equation can include the same set of explanatory variables as the selection equation, as well as an additional error term that captures the unobserved factors that affect the outcome variable. For identification purposes, it is often recommended that the selection equation includes additional variables, also known as exclusion restrictions, that are correlated with selection but not with the outcome (see Puhani, 2000, for further discussion). In our estimations, we drop four variables from the outcome equation for identification purposes: region, employment status, year trend, and lagged inflation squared. Although these variables can arguably be excluded *a priori* for not being expected to be related to bias, in our experimental estimations, these variables were indeed found to be either statistically insignificant (region and employment) or leading to variance inflation due to multicollinearity (year trend and lagged inflation squared).

Table 3 presents a comparison of estimates of inflation expectations bias for different groups of the population and across different estimation methods. There are several interesting findings. First, in contrast to the Probit regressions, using survey weights seems relevant for estimates of expectations bias. Comparing the estimates without selection, in columns (1) and (2), we note important differences on the effects of age and ethnicity; namely, accounting for the survey weights increases the magnitudes and significance of these variables' effects. Respondents older than 25 years old are found to have an average bias of more than 0.23 p.p. higher than respondents from the younger (<25 years) base age bracket. Pacific Islanders have an average bias of 0.34 p.p. higher than the base ethnic group (Others/NZ Europeans), while Maori people and Asians have an average bias of 0.16 and 0.13 p.p. higher than the base group, respectively. Of course, these estimates reflect only the sample of respondents that provided a response to the inflation expectations question.

To account for selection, we estimate Heckman selection models using the baseline Probit estimates from the previous section for the selection equation. We also explore two alternative estimation approaches: the two-step estimator, and the maximum likelihood (ML) estimator, presented in columns (3) and (4) of Table 3, respectively. The Heckman model estimates are mostly consistent across the two estimation methods; the only prominent difference relates to the statistical significance of gender, which is insignificant under the two-step estimator. Estimates of the significance of correction for selection, as captured by the lambda and rho (correlation between error terms) parameters, indicate the appropriateness of the selection model.

Perhaps more importantly, the estimates accounting for selection are in stark contrast to those obtained without the correction. First, after accounting for selection bias, the difference in inflation expectations between women and men decreases by more than a half. In fact, under the two-step estimator (column 3), gender does not seem to have a statistically significant effect on inflation expectations bias. Second, the differences in expectations bias by ethnicity and income bracket also turn insignificant after accounting for selection. Third, the effects of age increase in magnitude after accounting for selection, particularly steepening the profile of higher over-predictions for older respondents. Finally, some macro effects seem robust to selection bias and weighting: the higher the previous quarter's inflation, the higher respondents' over-prediction of inflation, and the switch to online mode decreased inflation predictions by more than 2.29 p.p., after controlling for the effects of the other variables.[14]

[14] The large magnitude of the estimate on the online mode dummy variable should be interpreted with caution, as it is also capturing the large negative biases that emerged from 2020Q3 onwards due to the sluggish adjustment of expectations to the increase of actual inflation rates 4 quarters ahead starting from 2021Q3 (see Figure 10).

Table 3: *Estimates of Inflation Expectations Bias With and Without Selection.*

Variables	Categories	(1) No Selection OLS	(2) No Selection WLS	(3) Heckman Selection Two Step	(4) Heckman Selection Max. Likelihood
Gender	Female	0.240*** (0.023)	0.276*** (0.023)	-0.005 (0.040)	0.105* (0.063)
Age	25-34	0.106* (0.063)	0.241*** (0.066)	0.462*** (0.071)	0.390*** (0.084)
	35-44	0.067 (0.058)	0.247*** (0.061)	0.609*** (0.073)	0.486*** (0.103)
	45-54	0.104* (0.059)	0.235*** (0.062)	0.689*** (0.080)	0.528*** (0.119)
	55-64	0.165*** (0.059)	0.317*** (0.062)	0.781*** (0.082)	0.620*** (0.123)
	65+	0.131** (0.060)	0.288*** (0.062)	0.773*** (0.083)	0.609*** (0.127)
Ethnicity	Asian	-0.093 (0.065)	0.130* (0.066)	0.045 (0.067)	0.082 (0.069)
	Maori	0.007 (0.063)	0.160** (0.066)	-0.052 (0.071)	0.026 (0.081)
	Pacific Islander	0.053 (0.141)	0.342** (0.158)	0.060 (0.157)	0.152 (0.170)
Real Income	High	-0.044* (0.025)	-0.087*** (0.026)	0.037 (0.030)	-0.010 (0.038)
	Low	0.121*** (0.033)	0.151*** (0.034)	0.022 (0.037)	0.065 (0.045)
Online	Yes	-2.766*** (0.032)	-2.587*** (0.035)	-2.294*** (0.046)	-2.481*** (0.046)
Lagged Inflation		0.155*** (0.009)	0.203*** (0.009)	0.215*** (0.010)	0.201*** (0.010)
Heckman	λ (lambda)			0.760*** (0.091)	
	ρ (rho)				0.284*** (0.094)
N.Obs.		39,312	39,312	39,312	39,312
R^2 (unweighted)		0.238	0.233	0.236	
Root MSE		2.136	2.143	2.139	2.148

Notes: Regressions (2) to (4) are weighted using survey weights. All regressions include quarter dummies. Estimates for the selection equation under the Heckman selection models are not presented for succinctness – these are based on the Probit baseline specification from Section 3 extended with interaction terms for online mode, i.e., the estimates underlying Figure 9. The R^2 statistics refer to the outcome equation. Numbers in parenthesis are robust standard errors. ***,**,* stand for 1%, 5%, 10% statistical significance, respectively.

Our estimates indicate that inflation biases calculated directly from the sample of observations, i.e., without accounting for selection, give a distorted picture of the socio-demographic differences in the population's inflation expectations. Many differences across groups, such as by gender, income and ethnicity, turn statistically insignificant after accounting for selection bias.

5 Adjusted Indices

In this section, we show how the correction for non-responses developed in the previous section can be used to adjust indices of average inflation expectations. The adjustment is again based on the Heckman correction for regressions, except that here the regression is designed to provide estimates of average inflation expectations.

A simple approach to obtain average indices of inflation expectations is to run a linear regression of the micro survey of inflation expectations data on quarter dummy variables, i.e.,

$$\pi_i^e = \hat{\delta}_1 I\left(t = 1998q2\right) + \ldots + \hat{\delta}_{99} I\left(t = 2022q4\right), \tag{4}$$

where $I\left(\bullet\right) = 1$ when the condition between brackets is true, and $\left\{\hat{\delta}_t\right\}$ are estimates of average inflation expectations for each quarter. This regression can be estimated using WLS to account for survey weights in constructing the expectations index.

The Heckman correction for non-response bias can be applied to equation 4 by including the inverse Mills Ratio, $\tilde{\lambda}_i$, obtained from the baseline Probit model, as an additional explanatory variable,

$$\pi_i^e = \hat{\theta}\tilde{\lambda}_i + \hat{\delta}_1' I\left(t = 1998q2\right) + \ldots + \hat{\delta}_{99}' I\left(t = 2022q4\right). \tag{5}$$

Estimates of equations 4 and 5 are presented in Figure 1 as the unadjusted and adjusted averages of inflation expectations, respectively. The average adjustment for selection bias amounts to -0.297 over the full sample, and ranges from a minimum of -0.398 (2018 Q1) to a maximum of -0.135 (2022 Q3).[15]

6 Concluding Remarks

In this paper, we provided an attempt to understand the demographic determinants of non-responses to inflation expectations questions in the RBNZ's Household Inflation Expectations survey. To address that issue, we use a Probit modelling approach. We find significant item non-response bias in this survey. Non-respondents to the one-year ahead inflation question are especially likely to be aged under 25, female, from a minority ethnic group (Māori, Pacific, Asian), unemployed, and from a low income household. A switch in the conduct of the survey

[15] Annual averages of these estimates are also presented in Table A.4 in Appendix A, where we also compare our estimates to published average inflation expectations – those can differ due to the method used to detect and exclude outliers from computations.

to online mode is found to substantially decrease non-responses to the inflation question, as well as decreasing the effects of socio-demographic characteristics on response rates.

We also identify consistent differences in inflation expectations according to the age of the respondent, but find that observed differences in expectations by gender, ethnicity and income are mostly due to sample selection bias. Namely, after accounting for sample selection, most of the differences in inflation expectations by socio-demographic characteristics turn insignificant or decrease substantially. The only exception is age, where we find that older individuals tend to over-predict one-year-ahead inflation more than the young.

These findings have important implications for how central banks use household inflation expectations measures. Even though the survey assigns weights based on the population distribution and thus attempts to correct for unit non-response bias, it does not correct for item non-response bias. Because both inflation perceptions and non-responses can differ across demographic groups, the weights allocated to the individual responses are likely to misrepresent the population. To address this issue, we propose an adjustment to the calculation of mean inflation expectations estimates using a sample selection correction model. We find that the unadjusted aggregate measure commonly used to gauge households' expectations are, on average, about 0.30 percentage points higher than a measure that accounts for item non-response bias.

Our findings also allow drawing important recommendations for how policymakers communicate with the population. Most households are likely to rely on the guidance provided by policymakers when forming their expectations about future inflation. The socio-demographic differences we identify in this paper suggest that some groups of the population may be less confident in providing a response to the inflation expectations question. A potential way to address these gaps would be to improve the outreach of policy with more targeted communications. We hope our research will help inform the development of such policies and lead to more accurate measures of inflation expectations.

A Data Appendix

Table A.1: *Data tabulations.*

(a) Age of Respondents

	Freq.	Percent	Cum.
15-17	2,283	2.74	2.74
18-24	6,564	7.89	10.63
25-34	12,796	15.38	26.01
35-44	18,792	22.59	48.6
45-54	14,320	17.21	65.81
55-64	12,498	15.02	80.83
65+	15,952	19.17	100
Total	83,205	100	

(b) Gender of Respondents

	Freq.	Percent	Cum.
Female	43,140	51.85	51.85
Male	40,065	48.15	100
Total	83,205	100	

(c) Ethnicity of Respondents

	Freq.	Percent	Cum.
Asian	4,246	5.1	5.1
Maori	5,557	6.68	11.78
Other	71,734	86.21	98
Pacific Islander	1,668	2	100
Total		83,205	100

(d) Regions where Respondents Live

	Freq.	Percent	Cum.
Auckland	26,392	31.72	31.72
Rest of North Island	36,398	43.74	75.46
South Island	20,415	24.54	100
Total	83,205	100	

(e) Real Income of Respondents

	Freq.	Percent	Cum.
High	28,759	38.14	38.14
Low	22,694	30.1	68.24
Median	23,947	31.76	100
Total	75,400	100	

(f) Employment Status of Respondents

	Freq.	Percent	Cum.
Employed	56,908	68.39	68.39
Unemployed/Other	26,297	31.61	100
Total	83,205	100	

Table A.2: *Probit Model Estimates on Responses to the Inflation Expectation Question.*

Variables	Categories	(1)	(2)	(3)	(4)
Gender	Female	-0.625*** (0.011)	-0.657*** (0.012)	-0.702*** (0.020)	-0.593*** (0.016)
Age	25-34	0.407*** (0.023)	0.412*** (0.024)	0.535*** (0.034)	0.174*** (0.039)
	35-44	0.742*** (0.022)	0.771*** (0.023)	0.976*** (0.032)	0.401*** (0.038)
	45-54	0.973*** (0.023)	0.983*** (0.024)	1.190*** (0.034)	0.580*** (0.039)
	55-64	1.032*** (0.024)	1.013*** (0.025)	1.226*** (0.036)	0.557*** (0.039)
	65+	1.121*** (0.024)	1.060*** (0.025)	1.124*** (0.038)	0.620*** (0.041)
Region	Rest of North Island	-0.037*** (0.013)	-0.036** (0.015)	-0.063*** (0.020)	-0.032* (0.018)
	South Island	-0.003 (0.015)	0.000 (0.021)	-0.002 (0.024)	-0.024 (0.021)
Ethnicity	Asian	-0.161*** (0.028)	-0.167*** (0.029)	-0.108** (0.050)	-0.180*** (0.036)
	Maori	-0.462*** (0.023)	-0.453*** (0.024)	-0.576*** (0.039)	-0.358*** (0.031)
	Pacific Islander	-0.589*** (0.042)	-0.588*** (0.045)	-0.699*** (0.075)	-0.508*** (0.057)
Real Income	High	0.315*** (0.013)	0.290*** (0.014)	0.329*** (0.020)	0.293*** (0.019)
	Low	-0.266*** (0.015)	-0.275*** (0.015)	-0.291*** (0.023)	-0.267*** (0.021)
Employment	Employed	0.080*** (0.015)			0.052** (0.021)
Occupation	Clerks/Admin/Sales		-0.011 (0.017)	-0.127*** (0.026)	
	Manual/Agricultural/Domestic		-0.050*** (0.018)	-0.105*** (0.026)	
	Professional/Manager/Trained		0.168*** (0.016)	0.156*** (0.025)	
Dependent Children	Yes		-0.094*** (0.014)	-0.100*** (0.020)	-0.059*** (0.020)
Groceries Shopping	Yes - jointly/shared			0.039 (0.030)	
	Yes - main shopper			-0.078*** (0.024)	
Home Ownership	Living with parents				-0.677*** (0.039)
	Mortgage				-0.233*** (0.022)
	Other				-0.344*** (0.047)
	Renting				-0.389*** (0.025)
Online Mode	Yes	1.103*** (0.018)	1.128*** (0.019)		1.049*** (0.025)
Year Trend		-0.027*** (0.001)	-0.030*** (0.001)	-0.044*** (0.000)	-0.014*** (0.002)
Lagged Inflation	Linear	-0.060*** (0.010)	-0.074*** (0.014)	-0.060*** (0.021)	-0.021 (0.025)
	Squared	0.015*** (0.002)	0.020*** (0.003)	0.019*** (0.005)	0.011** (0.005)
N.Obs.		75,400	71,799	36,011	35,993
Sample		98Q2-22Q4	98Q2-21Q4	98Q2-08Q3	09Q1-21Q4
McFadden R2		0.170	0.170	0.185	0.171

Notes: All regressions are weighted using survey weights. All regressions include year trend and quarter dummies. Numbers in parenthesis are robust standard errors. ***,**,* stand for 1%, 5%, 10% statistical significance, respectively.

Table A.3: *Average Partial Effects Split by Survey Mode.*

Variables	Categories	(1) Online=0	(1) Online=1	(2) Online=0	(2) Online=1	(4) Online=0	(4) Online=1
Gender	Female	-0.244***	-0.055***	-0.252***	-0.043***	-0.250***	-0.048***
		(0.004)	(0.006)	(0.004)	(0.007)	(0.006)	(0.008)
Age	25-34	0.128***	0.041***	0.127***	0.042***	0.083***	0.059***
		(0.008)	(0.014)	(0.009)	(0.016)	(0.017)	(0.017)
	35-44	0.262***	0.062***	0.268***	0.059***	0.197***	0.081***
		(0.008)	(0.013)	(0.008)	(0.015)	(0.019)	(0.017)
	45-54	0.355***	0.076***	0.352***	0.061***	0.283***	0.078***
		(0.009)	(0.012)	(0.009)	(0.015)	(0.020)	(0.017)
	55-64	0.378***	0.090***	0.363***	0.080***	0.277***	0.093***
		(0.009)	(0.012)	(0.009)	(0.014)	(0.020)	(0.016)
	65+	0.396***	0.148***	0.367***	0.137***	0.302***	0.133***
		(0.009)	(0.009)	(0.009)	(0.011)	(0.020)	(0.015)
Region	Rest of North Island	-0.017***	0.010	-0.015***	0.009	-0.016**	0.011
		(0.005)	(0.007)	(0.005)	(0.008)	(0.007)	(0.009)
	South Island	-0.007	0.020**	-0.003	0.017**	-0.015*	0.016
		(0.006)	(0.008)	(0.005)	(0.009)	(0.008)	(0.010)
Ethnicity	Asian	-0.078***	0.008	-0.080***	0.007	-0.104***	0.004
		(0.012)	(0.010)	(0.012)	(0.012)	(0.016)	(0.014)
	Maori	-0.190***	-0.030***	-0.184***	-0.020*	-0.180***	-0.028**
		(0.009)	(0.011)	(0.009)	(0.011)	(0.014)	(0.013)
	Pacific Islander	-0.236***	-0.027	-0.227***	-0.024	-0.222***	-0.042
		(0.015)	(0.020)	(0.016)	(0.023)	(0.023)	(0.026)
Real Income	High	0.122***	0.016**	0.111***	0.001	0.122***	0.022***
		(0.005)	(0.007)	(0.005)	(0.008)	(0.008)	(0.008)
	Low	-0.093***	-0.053***	-0.095***	-0.049***	-0.087***	-0.081***
		(0.006)	(0.009)	(0.006)	(0.010)	(0.008)	(0.011)
Employment	Employed	0.017***	0.045***			0.004	0.061***
		(0.006)	(0.008)			(0.009)	(0.010)
Occupation	Clerks/Admin/Sales			-0.022***	0.088***		
				(0.007)	(0.010)		
	Manual/Agricultural/Domestic			-0.039***	0.068***		
				(0.006)	(0.010)		
	Professional/Manager/Trained			0.051***	0.087***		
				(0.006)	(0.009)		
Dependent Children	Yes			-0.032***	-0.003	-0.027***	0.005
				(0.005)	(0.009)	(0.008)	(0.009)
Home Ownership	Living with parents					-0.101***	-0.266***
						(0.020)	(0.019)
	Mortgage					-0.042***	-0.102***
						(0.009)	(0.012)
	Other					-0.069***	-0.133***
						(0.021)	(0.022)
	Renting					-0.122***	-0.077***
						(0.010)	(0.012)
N.Obs.		75,400		71,799		35,993	
Sample		98Q2-22Q4		98Q2-21Q4		09Q1-21Q4	
McFadden R2		0.182		0.182		0.188	

Notes: Average partial effects are calculated using the delta method and averaging over the sample observations holding other variables constant at their sample values. Note that the column headers correspond to Table 2 specifications numbering, and that every two columns correspond to a separate regression. Numbers in parenthesis are robust standard errors. ***,**,* stand for 1%, 5%, 10% statistical significance, respectively.

Table A.4: *Average Inflation Expectations by Year.*

	1998	Δ	1999	Δ	2000	Δ	2001	Δ	2002	Δ
Published series (H1)	3.83		3.48		4.05		4.05		3.85	
Excl. outliers w/ hybrid IQR	3.18	-0.65	2.97	-0.51	3.54	-0.51	3.57	-0.48	3.42	-0.43
Baseline sample - unadjusted	3.17	-0.01	2.96	-0.01	3.52	-0.02	3.55	-0.02	3.40	-0.02
Baseline sample - adjusted	2.89	-0.28	2.68	-0.28	3.22	-0.30	3.26	-0.29	3.07	-0.34
	2003		2004		2005		2006		2007	
Published series (H1)	3.63		4.10		4.22		4.68		4.53	
Excl. outliers w/ hybrid IQR	3.22	-0.41	3.58	-0.52	3.86	-0.36	4.15	-0.53	3.96	-0.57
Baseline sample - unadjusted	3.23	0.01	3.56	-0.02	3.86	0.00	4.16	0.01	3.97	0.01
Baseline sample - adjusted	2.94	-0.29	3.27	-0.29	3.55	-0.31	3.85	-0.31	3.65	-0.32
	2008*		2009		2010*		2011		2012	
Published series (H1)	4.67		3.98		3.90		4.50		3.53	
Excl. outliers w/ hybrid IQR	4.14	-0.53	3.66	-0.32	3.56	-0.34	4.13	-0.37	3.17	-0.35
Baseline sample - unadjusted	4.14	0.00	3.67	0.01	3.54	-0.02	4.14	0.01	3.17	0.00
Baseline sample - adjusted	3.82	-0.32	3.33	-0.34	3.20	-0.34	3.82	-0.32	2.83	-0.34
	2013		2014		2015		2016		2017	
Published series (H1)	3.23		3.40		2.65		2.43		2.98	
Excl. outliers w/ hybrid IQR	2.95	-0.28	2.96	-0.44	2.35	-0.30	2.07	-0.36	2.50	-0.48
Baseline sample - unadjusted	2.93	-0.02	2.95	-0.01	2.33	-0.02	2.06	-0.01	2.51	0.01
Baseline sample - adjusted	2.59	-0.34	2.60	-0.35	1.99	-0.34	1.71	-0.35	2.12	-0.39
	2018		2019		2020		2021		2022	
Published series (H1)	3.18		2.75		2.85		3.65		6.98	
Excl. outliers w/ hybrid IQR	2.85	-0.33	2.45	-0.30	2.51	-0.34	3.24	-0.41	7.71	0.73
Baseline sample - unadjusted	2.83	-0.02	2.46	0.01	2.50	-0.01	3.24	0.00	7.70	-0.01
Baseline sample - adjusted	2.55	-0.28	2.26	-0.20	2.30	-0.20	3.04	-0.20	7.55	-0.15

Notes: The table presents annual averages of the quarterly cross-section averages of one-year-ahead inflation expectations. The published series is the historical data published at the RBNZ website and using different outlier detection rules across time – see the main text for details. The second series is based on our calculations using the hybrid IQR outlier detection rule. The third series uses the same outlier detection rule as the previous, but restricts the sample to observations with available demographic information to estimate our baseline Probit specification. The fourth series is adjusted for sample selection bias according to our methodology. (*) Due to missing micro observations, the averages for 2008 are based on quarters Q1 to Q3, and the averages for 2010 are based on quarter Q1 and Q4.

References

Blanchflower, D. G. and R. Kelly (2008). Macroeconomic literacy, numeracy and the implications for monetary policy. Technical report, Working Paper Bank of England.

Blanchflower, D. G. and C. MacCoille (2009). The formation of inflation expectations: an empirical analysis for the uk. Technical report, National Bureau of Economic Research.

Borio, C., M. Lombardi, J. Yetman, and E. Zakrajšek (2023). The two-regime view of inflation. *BIS Papers*.

Bruine de Bruin, W., W. van der Klaauw, M. van Rooij, F. Teppa, and K. de Vos (2017). Measuring expectations of inflation: Effects of survey mode, wording, and opportunities to revise. *Journal of Economic Psychology 59*, 45–58.

Bruine de Bruin, W., W. Vanderklaauw, J. S. Downs, B. Fischhoff, G. Topa, and O. Armantier (2010). Expectations of inflation: The role of demographic variables, expectation formation, and financial literacy. *Journal of Consumer Affairs 44*(2), 381–402.

Bürgi, C. R. S. (2023). How to deal with missing observations in surveys of professional forecasters. *Journal of Applied Economics 26*(1), 2185975.

Comerford, D. A. (2023). Response bias in survey measures of expectations: Evidence from the survey of consumer expectations' inflation module. *Journal of Money, Credit and Banking n/a*(n/a).

Curtin, R. T. (1996). Procedure to estimate price expectations. Technical report, Survey of Consumers, University of Michigan.

D'Acunto, F., U. Malmendier, and M. Weber (2023). What do the data tell us about inflation expectations? In *Handbook of economic expectations*, pp. 133–161. Elsevier.

D'Acunto, F., U. Malmendier, J. Ospina, and M. Weber (2021). Exposure to grocery prices and inflation expectations. *Journal of Political Economy 129*(5), 1615–1639.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: journal of the econometric society 42*(4), 679–694.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society 47*(1), 153–161.

Leung, C. (2009, June). The demographics of household inflation perceptions and expectations. *Reserve Bank of New Zealand Bulletin 72*, 34–42.

McGovern, M. E., D. Canning, and T. Bärnighausen (2018). Accounting for non-response bias using participation incentives and survey design: An application using gift vouchers. *Economics letters 171*, 239–244.

Meyer, B. D., W. K. C. Mok, and J. X. Sullivan (2015, November). Household surveys in crisis. *Journal of Economic Perspectives 29*(4), 199–226.

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys 14*(1), 53–68.