

Gülenay Chadwick, Meltem; Cherry, Rennae; Galimberti, Jaqueson K.

Working Paper

Nonresponse bias in household inflation expectations surveys

ADB Economics Working Paper Series, No. 706

Provided in Cooperation with:

Asian Development Bank (ADB), Manila

Suggested Citation: Gülenay Chadwick, Meltem; Cherry, Rennae; Galimberti, Jaqueson K. (2023) : Nonresponse bias in household inflation expectations surveys, ADB Economics Working Paper Series, No. 706, Asian Development Bank (ADB), Manila, <https://doi.org/10.22617/WPS230552-2>

This Version is available at:

<https://hdl.handle.net/10419/298152>

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/3.0/igo/>

NONRESPONSE BIAS IN HOUSEHOLD INFLATION EXPECTATIONS SURVEYS

Meltem Chadwick, Rennae Cherry, and Jaqueson K. Galimberti

NO. 706

December 2023

**ADB ECONOMICS
WORKING PAPER SERIES**

ADB Economics Working Paper Series

Nonresponse Bias in Household Inflation Expectations Surveys

Meltem Chadwick, Rennae Cherry,
and Jaqueson K. Galimberti

No. 706 | December 2023

The *ADB Economics Working Paper Series* presents research in progress to elicit comments and encourage debate on development issues in Asia and the Pacific. The views expressed are those of the authors and do not necessarily reflect the views and policies of ADB or its Board of Governors or the governments they represent.

Meltem Chadwick (meltem.chadwick@seacen.org) is a senior economist at the South East Asian Central Banks Research and Training Centre. Rennae Cherry (rennae.cherry@rbnz.govt.nz) is an economic analyst at the Reserve Bank of New Zealand. Jaqueson K. Galimberti (jgalimberti@adb.org) is an economist at the Economic Research and Development Impact Department, Asian Development Bank.



Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2023 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

Some rights reserved. Published in 2023.

ISSN 2313-6537 (print), 2313-6545 (electronic)
Publication Stock No. WPS230552-2
DOI: <http://dx.doi.org/10.22617/WPS230552-2>

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

By making any designation of or reference to a particular territory or geographic area, or by using the term “country” in this publication, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This publication is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Corrigenda to ADB publications may be found at <http://www.adb.org/publications/corrigenda>.

Note:
In this publication, “NZ\$” refers to New Zealand dollars.

ABSTRACT

This paper uses microdata from the Reserve Bank of New Zealand's Household Inflation Expectations survey to obtain an accurate read of households' true inflation expectations by studying how different demographic groups respond (or do not respond) to the inflation expectations question in the survey. We find nonresponses lead to substantial underrepresentation of some demographic groups in the survey: young, female, low-income, and minority ethnic groups have lower response rates. How the survey is conducted also affects item response rates. The survey response rates increase when the survey is conducted online and when inflation rates deviate from the central bank's target range. Using a sample selection model, we assess whether the survey has item nonresponse bias by comparing the demographic characteristics of responders and nonresponders. After accounting for selection, we find that observed differences in inflation expectations by gender, ethnicity, and income decrease substantially, while differences by age increase. We quantify and demonstrate how to adjust average inflation expectations for bias caused by item nonresponse. 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 nonresponse, demographic heterogeneity

JEL codes: C83, D84, E31, E71

Paper presented at the 2023 BIS Annual Workshop of the Asian Research Network, the 2023 Conference of the New Zealand Association of Economists, the 5th Bank of Indonesia-ADB Institute-Asia Pacific Applied Economics Association Workshop, and at the ADB/ERDI Seminar Series. We thank Seunghyeon Lee and Bhavesh Garg for the insightful discussions of our paper and other participants of the presentations above for many 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: Jaqueson K. Galimberti (jgalimberti@adb.org)

1 Introduction

Measures of inflation expectations are of paramount importance to monetary policy. Inflation expectations are key determinants of prices through the forward-looking behavior 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 crucial to understand the accuracy of survey data in tracking the population's beliefs about inflation. In this paper, we use microdata from the Reserve Bank of New Zealand's (RBNZ) Household Inflation Expectations survey to investigate how nonresponses to inflation expectations questions can bias measurements obtained from such surveys.

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

¹Throughout this paper, we focus on so-called item nonresponses to the specific survey question on inflation expectations instead of unit nonresponses to the whole survey. The use of survey weights corrects the incidence of unit nonresponse (see Meyer et al., 2015, for further discussion).

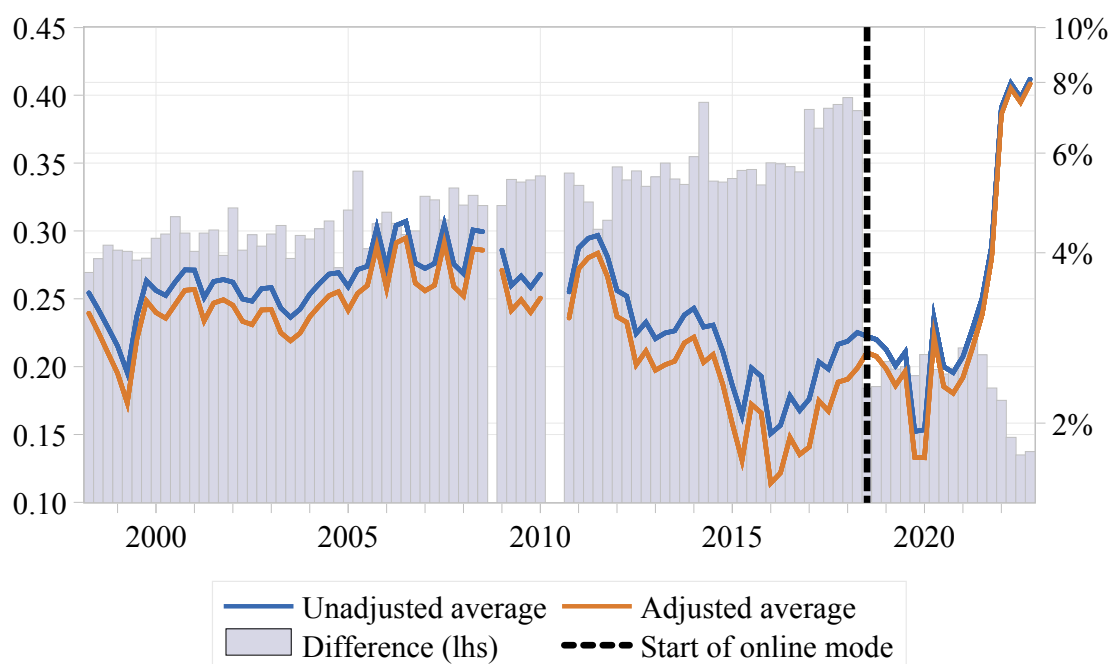
Another important finding relates to the effect of survey mode on nonresponses. Starting in Q3 2018, the survey changed from being conducted by telephone to online mode. We find that this change significantly affected the incidence of nonresponse to the inflation expectations question. The average of nonresponses decreased to about 24% since the survey moved to online mode. As evidenced in Figure 1, this change also significantly reduced the effect of nonresponse bias in estimating average inflation expectations. According to our estimates, the move to online mode generally reduced the gaps in nonresponses across the different population groups. In other words, conducting the survey online has made it more inclusive for previously underrepresented demographic groups.

Another important finding is that nonresponse bias also depends on the level of the inflation rate at the time the survey is conducted. We find that response rates tend to increase non-linearly when the previous quarter's inflation is away from the central bank's target range.² For example, an inflation rate increase from 2% to 7% increases the average response probability by 12%, while this probability barely changes over an inflation range between 0% and 4%. This effect is also apparent in Figure 1, where the adjustment for nonresponse bias decreases in magnitude since the onset of the recent increase in inflation rates.

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, a 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

²We used lagged values of the inflation rate because these are the latest available information to respondents at the time the surveys are conducted each quarter.

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



RBNZ = Reserve Bank of New Zealand, lhs = left-hand side.

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 nonresponse bias. The gaps in the fourth quarter (Q4) of 2008 and Q2 and Q3 2010 are due to missing observations. The dashed line depicts when the survey switched to online mode in Q3 2018. Before Q3 2018, the survey was conducted by telephone.

Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

inflation expectation question depending on their characteristics.³ We consider several specifications for the set of explanatory variables depending on their availability across the sample period. We find that the effects of variables included in our baseline specification—gender, age, region, ethnicity, income, and employment—are robust across sample periods and to the inclusion of more information such as occupation, exposure to grocery shopping, whether there are children in the household and types of home ownership.

We then study inflation expectations bias accounting for nonresponses 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 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 Compared to estimates that do not account for selection, we find that most differences in bias across subgroups turn insignificant after accounting for nonresponse bias. Observed differences in inflation expectations by gender, ethnicity, and income turn insignificant or decrease substantially in magnitude. The only exception is age, where older individuals tend to over-predict inflation more than the young, which is stronger after accounting for selection. For robustness purposes, we also consider different estimation methodologies and find that our estimates are not sensitive to the choice of estimation method.⁵

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

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

⁴The Heckman selection model estimates the selection equation and uses the predicted probabilities from the selection equation as a correction term in the outcome equation. This correction term, known as the inverse Mills ratio, accounts for the selection bias by adjusting the coefficients in the outcome equation.

⁵The Heckman sample selection model can be estimated using either a maximum likelihood approach or the original two-step approach (see, e.g., Puhani, 2000, for more details). Here, we extend those methods to account for survey weights and derive weighted estimates.

estimates of the Heckman correction term as an additional variable in this regression, we obtain average inflation expectations adjusted for nonresponse bias—these are the adjusted average expectations reported in Figure 1. The simplicity of this approach makes it attractive for operational purposes: to obtain updated estimates every quarter, all that is required are new estimates of the inverse Mills ratio, which can be easily computed from the pre-fitted Probit model. Indeed, the fact that the Probit model estimates are relatively stable across sub-samples indicates that the adjustment is unlikely to undergo severe revisions over time. Finally, our estimates use survey weights to account for unit nonresponse bias arising from difficulty in obtaining a representative population survey sample. Although these weights can not account for determinants of nonresponses to the inflation expectations question, we also find that they are relevant for the analysis of inflation expectations bias.

1.1 Related Literature and Surveys

This paper relates to the broader literature on the heterogeneity of inflation expectations. 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. Pfajfar and Santoro (2010) document pervasive heterogeneity in forming inflation expectations in that same survey. Malmendier and Nagel (2016) show consumer inflation expectations also vary with age due to learning from experience. D'Acunto et al. (2023) document that household inflation expectations are upward-biased and systematically different across gender, income, education, and race. Our findings based on New Zealand data add to this literature by showing that some demographic differences are a product of nonresponse bias. When accounting for selection, we find that differences by gender, ethnicity, and income decrease substantially.

Our findings about the underrepresentation of some demographic groups are consistent with previous studies in the literature. Exploring United Kingdom (UK) survey microdata, an early study by Blanchflower and MacCoille (2009) also found significant nonresponse bias from young, female, and low-income respondents. Leung (2009) reported similar findings with a shorter sample from the RBNZ household survey. Our finding that online survey mode can attenuate nonresponse 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 nonresponse bias in inflation expectations surveys. One standard approach to deal with nonresponses is to replace the nonresponses or missing observations by imputation. The United States (US) 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 analyzing the survey of expectations data.

More broadly, the issue of item nonresponse 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 nonresponse bias. Comerford (2023) proposes using a verbal question to deal with nonresponse bias found in inflation expectations derived from density forecasts. Ex ante, these methods provide vital insights into survey design.

However, the required survey features are rarely available for long-running surveys. Our approach offers a potential solution to these cases.

Finally, nonresponse bias can be important for other household surveys of inflation expectations. For example, in the US MSC, a major survey in this area, nonresponses to the inflation expectations question (prior to imputation) amounted to an average of 9% of the monthly samples collected between 1978 and 2022; nonresponses also varied substantially over time, ranging from lows of about 3%, mostly observed in 1985 and 2022, to highs of about 25% observed in 1978.⁶ Another example is the Bank of England's (BoE) Inflation Attitudes Survey, where nonresponses 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.⁷ Needless to say, these are examples of surveys from advanced economies, and we would expect the relevance of item nonresponse to be even greater in less developed contexts. Our methodology can be easily applied to analyzing and adjusting these surveys using their corresponding socio-demographic information on the surveyed households.

1.2 Paper Organization

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 nonresponses to the inflation expectation question using Probit models. Section 4

⁶The MSC also distinguishes between respondents refusing to answer the inflation question (0.14% of the 1978-2022 sample, on average), those that don't know (DK) a point estimate either 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.

⁷These statistics on the BoE survey exclude the data from Q2 2020, when nonresponses declined to 0.12% due to a design issue in the switch to online mode during the coronavirus disease (COVID-19) social distancing guidance. The problem 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 nonresponses returned to usual levels. Intriguingly, the more recently launched European Central Bank (ECB) Consumer Expectations Survey does not allow the respondent to proceed with the survey without answering the inflation expectations question, which is a potential design flaw of the questionnaire.

focuses on estimates of inflation expectations bias and outlines our proposed approach to account for nonresponses using a sample selection model. Section 5 shows how the correction for nonresponses can be used to adjust indices of average inflation expectations. Finally, section 6 concludes with some remarks.

2 Data

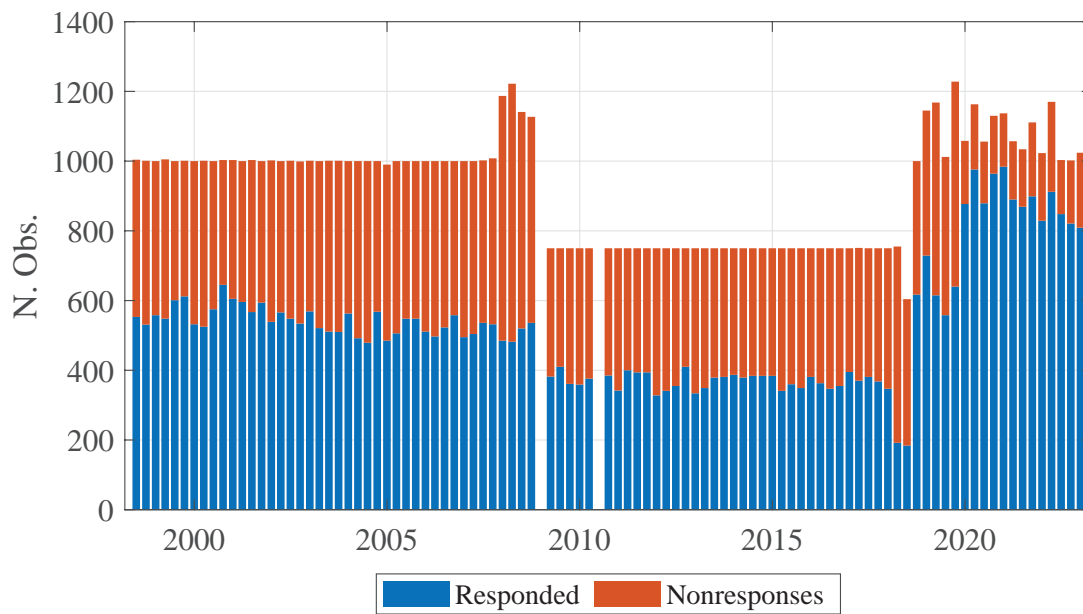
2.1 Survey Design

The RBNZ Household Inflation Expectations survey is conducted quarterly and achieves approximately 1,000 household responses per quarter. Our sample covers the period from 1998 to 2022 and contains 89,834 individual responses. The individual responses are anonymized 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 expectations question. While we attempt to understand the determinants of such nonresponses by looking at household characteristics and macroeconomic factors, it is not immediately obvious why some survey respondents do not answer the inflation expectations question. Although collected for a limited period, one question in the survey can be informative about this issue. Between Q3 2018 and Q4 2021, the survey asked a question about the respondent's understanding of inflation.⁸ Nonresponses over that sub-period amounted to 26% of the sample, but only 5% of the respondents indicated not

⁸Options to the question "What is your understanding of inflation?" included: (i) increased prices and cost of living, (ii) erosion of wages, (iii) an index/a measure, (iv) other (specify), (v) unsure, and (vi) no comment.

Figure 2: Responses and Nonresponses to Inflation Expectation Question



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

Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

understanding inflation (answering “unsure” or “no comment”). Hence, for the majority of nonresponses to the inflation expectations question, 81% to be precise, the reason for not providing an estimate of future inflation is not a lack of understanding about inflation.

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

The dataset contains various 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.⁹ While the survey has been redeveloped over time to improve data quality and better align with international best practices 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 Q1 1998 to Q3 2008, Pacific Island respondents could identify as one of six Pacific Island ethnic groups: Cook Island, Niuean, Fijian, Tongan, Samoan, or other Pacific Islands. In Q4 2008, the six Pacific Island ethnic groups were aggregated into a single ethnic umbrella group, leaving Pacific Island respondents with one response option: Pacific Islanders. To maintain consistency, respondents who identified with one

⁹See Table A.1 in Appendix 6 for tabulations of these data and further discussion in the following subsection.

of the six Pacific Island ethnic groups from Q1 1998 to Q3 2008 were grouped and categorized 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 three categories: under NZ\$50,000 (low), NZ\$50,000-NZ\$100,000 (median), and over NZ\$100,000 (high). In the absence of nominal household income, real personal income is used to 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, or high, real household income would have been low, median, or high, respectively.¹⁰ Figure 3 presents the evolution of our sample compositions according to different demographic variables.

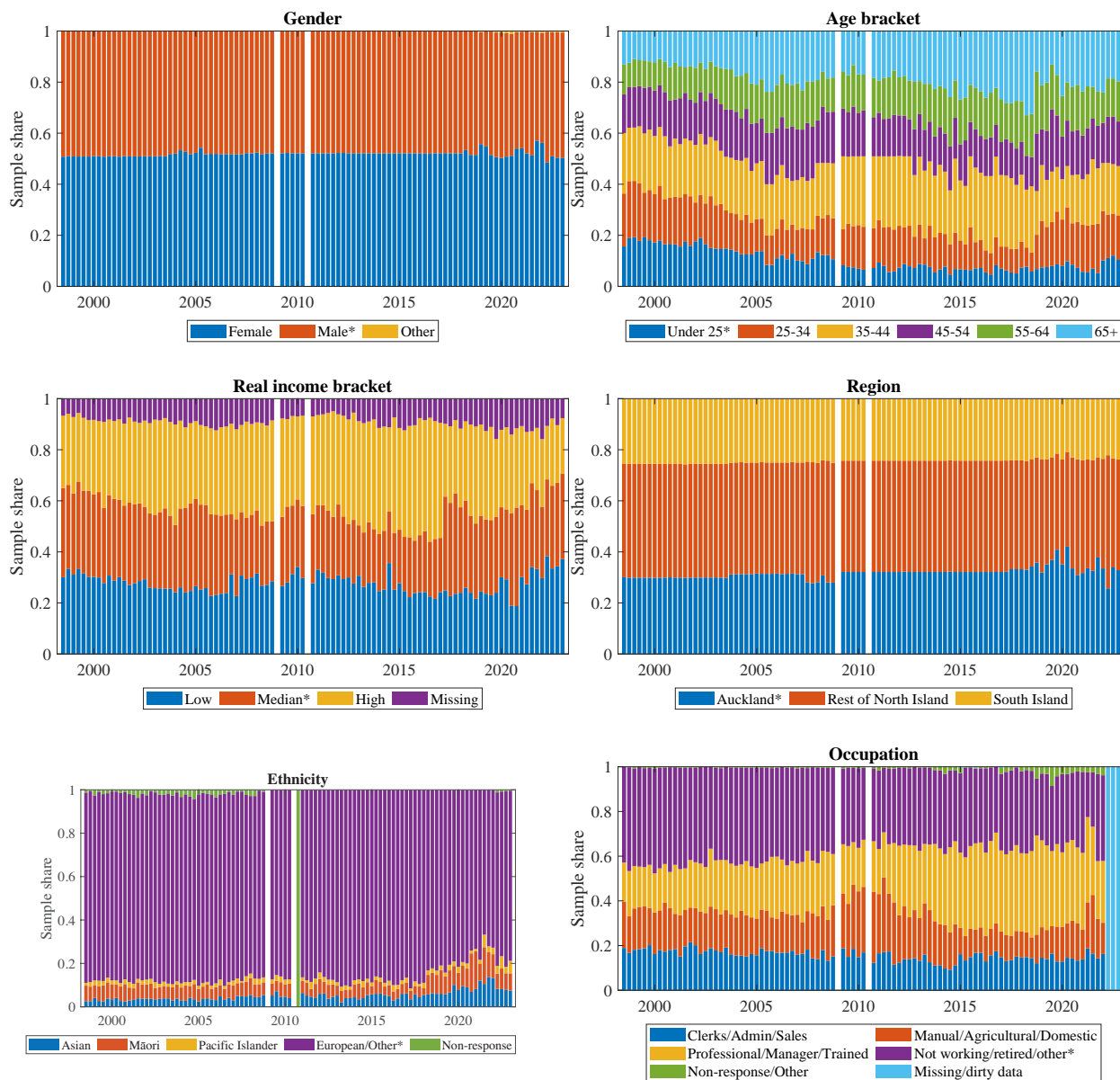
2.2 Sample Statistics

Over the period that the survey was conducted by telephone, Q2 1998 to Q2 2018, 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

¹⁰Upon 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 0.702, and the hypothesis of independence is strongly rejected (p-value=0).

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.

Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

respondents who identified with other ethnicity groups decreased by 10.73%. Survey respondents aged 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 reduced by 18.69%. Furthermore, female respondents increased marginally by 0.89%.

Over the whole sample, from 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 NZ\$100,000 was the most frequently reported real household income range, followed by NZ\$50,000 to NZ\$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 NZ\$40,000, 31.99% reported incomes between NZ\$40,000 to NZ\$80,000, and 19.75% reported incomes over NZ\$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 average one-year-ahead inflation expectation of males has marginally surpassed that of females. Similarly, respondents aged 35 and under have traditionally had higher average one-year-ahead inflation expectations than those aged 35 and over. However, this trend has inverted in recent years, resulting in lower inflation expectations among respondents under 35 than those over 35. Likewise, the 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 nonresponses 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 who responded versus those who did not respond to the inflation question. This non-random incidence of item nonresponses

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 collected over time, with the green shaded areas reflecting available observations by variable. The most notable omission is the data for the Q4 2008 and the Q2 2010, which is 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 data availability will determine the number of observations available for each specification. For example, grocery 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 significantly impact estimates derived based on this type of data. The method for outlier detection used in constructing aggregate indices from the RBNZ household survey of inflation expectations also changed over time. Starting from 2022, the methodology shifted from a fixed interval rule, which discarded inflation expectations lower than -2% and higher than 15%, to an interquartile range (IQR) method based on a whisker equal to 1.5.¹¹

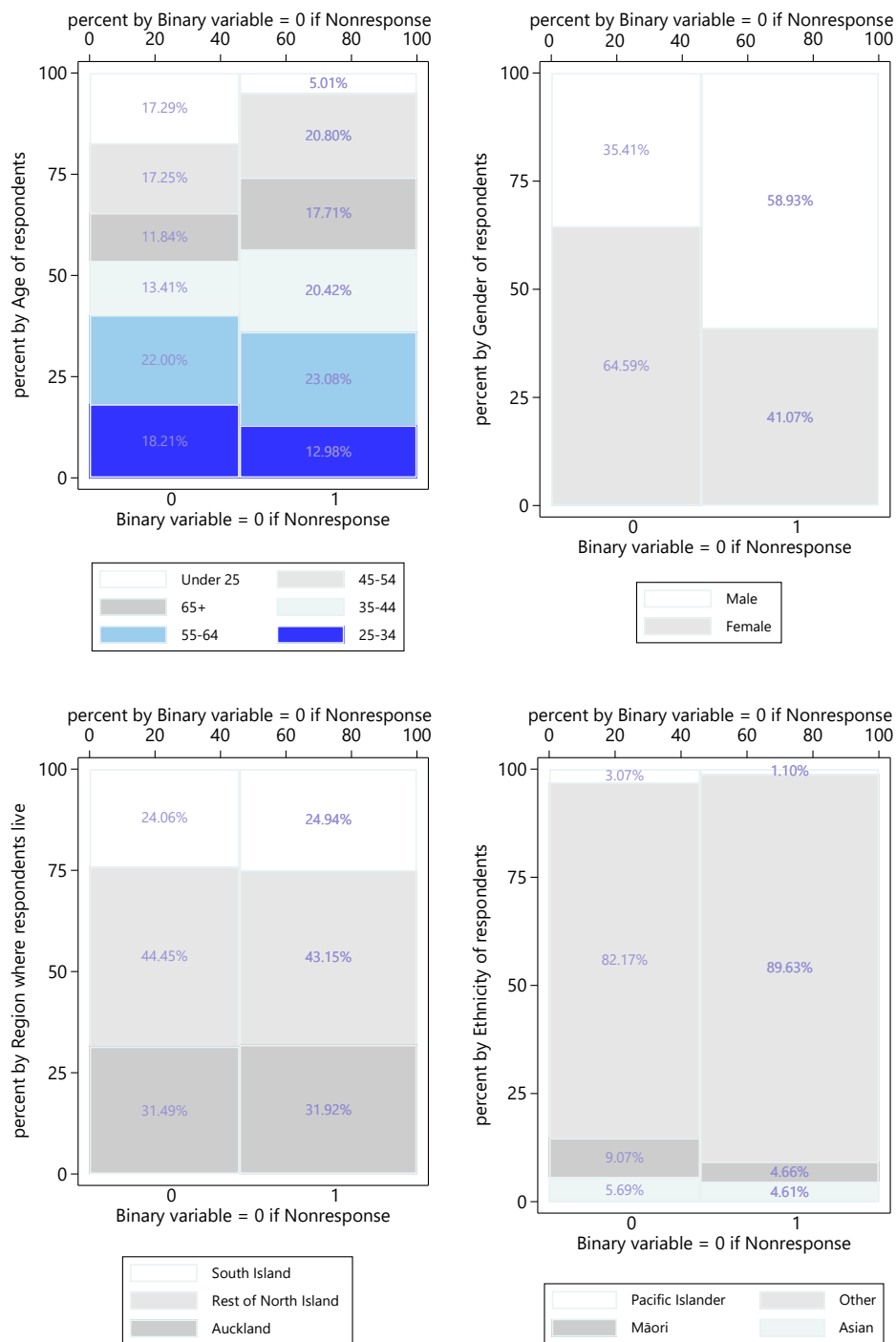
Table 1 shows how different outlier detection rules compare for our sample. There is a clear issue with the consistency of the previous rules between the two sub-periods divided by the 2022 change. While the fixed interval rule turned too restrictive from 2022, the new IQR rule implied a higher number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

¹¹The IQR method excludes observations falling outside the following limits:

$$\begin{aligned} \text{Lower limit} &= Q1 - \text{whisker} \times IQR, \\ \text{Upper limit} &= Q3 + \text{whisker} \times IQR, \end{aligned}$$

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 Nonresponses to Inflation Expectation Question by Group



Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

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.

Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

Table 1: *Frequency of Outliers by Detection Rule*

Sample	Rule			
	(-2,15)	$\pm 1.5 \times IQR$	$\pm 3 \times IQR$	Hybrid IQR
Q2 1998 to Q4 2021	1.84%	4.80%	2.48%	4.80%
Q1 2022 to Q4 2022	18.55%	16.19%	10.19%	10.19%

Notes: The statistics are the 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 Q2 1998 to Q4 2021 and a *whisker* = 3 for the sample from Q1 2022 to Q4 2022 to attenuate the difference in outlier frequency across these subperiods.

Source: Authors' estimate.

considering that 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. The hybrid IQR method achieves greater consistency in the number of outliers excluded across the sample. In total, 4,535 observations were excluded from the analysis.

3 Determinants of Responses to the Inflation Question

3.1 Probit Model

As a first step in finding the nonresponse bias and correcting for it, we study the determinants of each household's nonresponse behavior. To do so, we estimate different Probit models that relate the choice of response to the household's demographic characteristics. 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 general structure of probit models we estimate.

$$Pr(\text{Response}_i = 1 | \mathbf{X}_i, \mathbf{Z}_i, Y_i, \mathbf{Q}_i, O_i) = \Phi(\beta_0 + \beta_1 \mathbf{X}_i + \beta_2 \mathbf{Z}_i + \beta_3 Y_i + \beta_4 \mathbf{Q}_i + \beta_5 O_i). \quad (1)$$

Demographic variables are represented by \mathbf{X}_i , and macroeconomic variables are represented by \mathbf{Z}_i . Note the subscript i indexes for the survey/response observation. Macroeconomic variables do not vary across responses within a survey wave but across survey waves. Y_i is the yearly trend. \mathbf{Q}_i are quarter seasonal dummy variables. O_i is also a dummy variable that indicates whether the response i occurred after Q2 2018 when the survey switched to online mode. $Response_i$ 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.

3.2 Model Estimates

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.¹² Females seem to be less likely to respond to the inflation expectations question, with an average probability of responding about 20% lower than men. Māori and Pacific Islanders are also more likely to be in the nonresponse category. In contrast, older individuals who are employed and 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 the effects of conducting the survey in online mode in this section.

We also find a significant downward trend in the responses to the inflation question. Despite 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 nonresponses of about 1% per year. This is consistent with more broad evidence of

¹²Estimates of the coefficients of equation 1 are presented in Table A.2 in Appendix 6.

increased rates of both unit and item nonresponses in household surveys (Meyer et al., 2015). These findings are robust across the different specifications.

Figures 6 and 7 summarize some of the results illustrated in the first column of Table 2. Accordingly, gender is an essential 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 significantly lower response rate is attached to being a Māori or a Pacific Islander.

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

Variables	Categories	(1)	(2)	(3)	(4)
Gender (base=Male)	Female	-0.211*** (0.004)	-0.223*** (0.004)	-0.235*** (0.007)	-0.199*** (0.005)
Age (base=under 25)	25-34	0.131*** (0.008)	0.131*** (0.008)	0.168*** (0.011)	0.058*** (0.013)
	35-44	0.246*** (0.007)	0.255*** (0.008)	0.321*** (0.011)	0.134*** (0.013)
	45-54	0.324*** (0.008)	0.327*** (0.008)	0.393*** (0.011)	0.194*** (0.013)
	55-64	0.343*** (0.008)	0.337*** (0.008)	0.405*** (0.011)	0.186*** (0.013)
	65+	0.372*** (0.007)	0.353*** (0.008)	0.371*** (0.012)	0.207*** (0.013)
Region (base=Auckland)	Rest of North Island	-0.012*** (0.004)	-0.012*** (0.005)	-0.020*** (0.006)	-0.010* (0.006)
	South Island	-0.001 (0.005)	-0.000 (0.005)	-0.001 (0.007)	-0.008 (0.007)
Ethnicity (base=European/Other)	Asian	-0.053*** (0.009)	-0.055*** (0.010)	-0.035** (0.016)	-0.060*** (0.012)
	Māori	-0.152*** (0.007)	-0.149*** (0.008)	-0.185*** (0.012)	-0.118*** (0.010)
	Pacific Islander	-0.192*** (0.013)	-0.192*** (0.014)	-0.222*** (0.022)	-0.166*** (0.018)
Real income (base=Median)	High	0.104*** (0.004)	0.096*** (0.005)	0.108*** (0.007)	0.096*** (0.006)
	Low	-0.089*** (0.005)	-0.092*** (0.005)	-0.096*** (0.008)	-0.089*** (0.007)
Employment (base=Unemployed/Other)	Employed	0.026*** (0.005)			0.017** (0.007)
Occupation (base=Unemployed/Other)	Clerks/Admin/Sales		-0.004 (0.006)	-0.041*** (0.008)	
	Manual/Agricultural/Domestic		-0.017*** (0.006)	-0.034*** (0.008)	

Continued on the next page

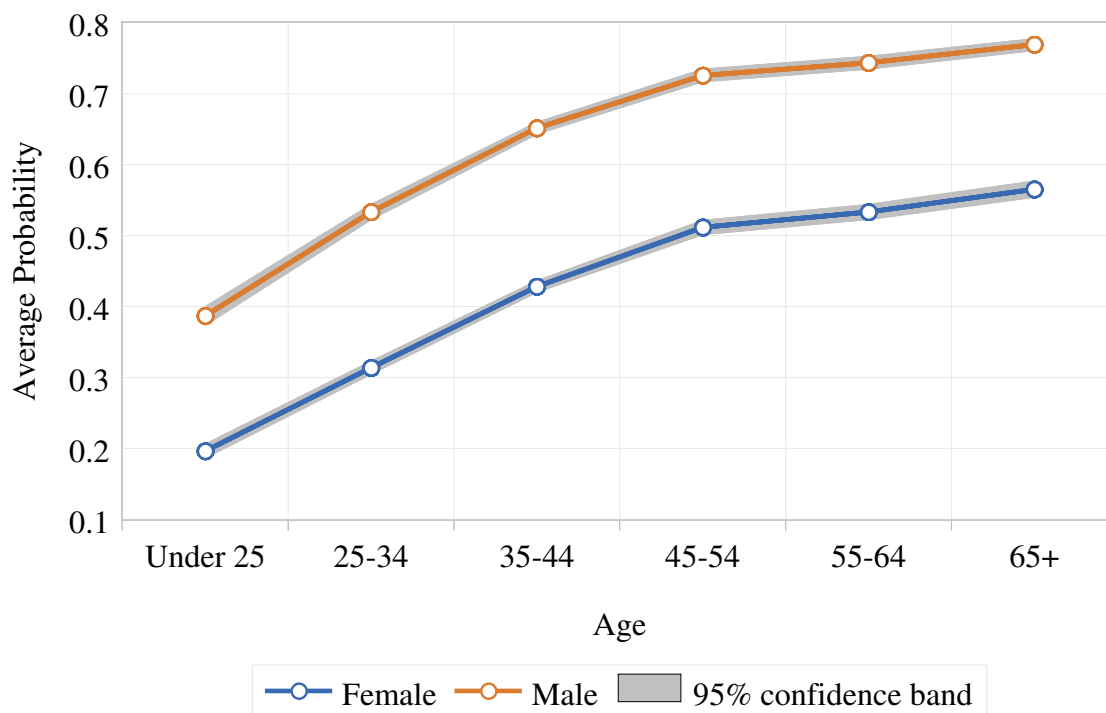
Variables	Categories	(1)	(2)	(3)	(4)
	Professional/Manager/Trained		0.055*** (0.005)	0.050*** (0.008)	
Dependent children (base=No)	Yes		-0.031*** (0.004)	-0.032*** (0.006)	-0.019*** (0.007)
Groceries shopping (base=No)	Yes - jointly/shared			0.013 (0.010)	
	Yes - main			-0.025*** (0.008)	
Home ownership (base=Owner)	Living with parents				-0.224*** (0.013)
	Mortgage				-0.077*** (0.007)
	Other				-0.114*** (0.016)
	Renting				-0.129*** (0.008)
Online (base=No)	Yes	0.335*** (0.001)	0.338*** (0.001)		0.336*** (0.001)
Year trend		-0.009*** (0.000)	-0.010*** (0.000)	-0.014*** (0.000)	-0.004*** (0.001)
Lagged inflation		0.002 (0.001)	0.001 (0.002)	0.008*** (0.003)	0.006* (0.003)
No. of observations		75,400	71,799	36,011	35,993
Sample period		Q2 1998 to Q4 2022	Q2 1998 to Q4 2021	Q2 1998 to Q3 2008	Q1 2009 to Q4 2021
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.

Source: Authors' estimate.

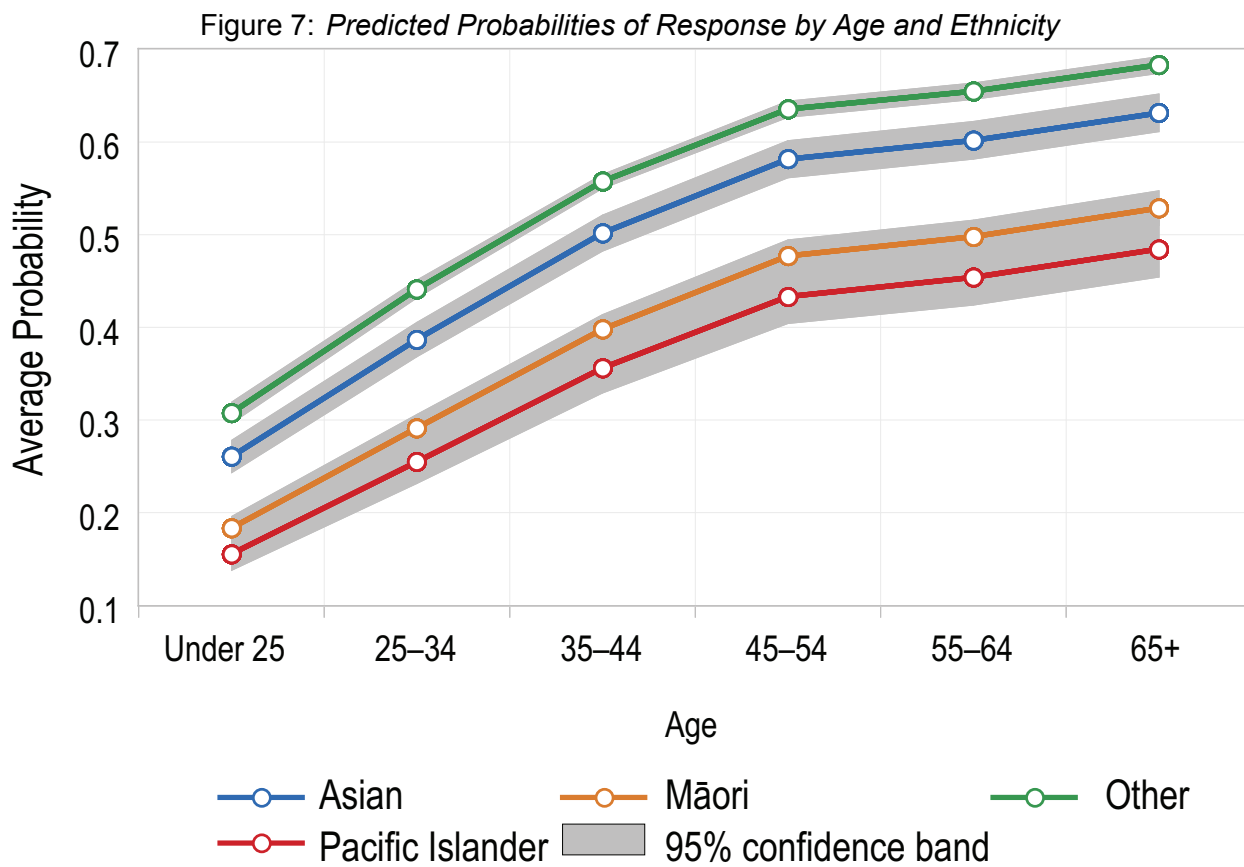
Figure 8 illustrates the link between lagged inflation rates and the predicted response probabilities. In our Probit estimations, we have found evidence of a quadratic relationship between responses and lagged inflation rates. Intriguingly, the averaged nonlinear effects of lagged inflation over our sample are mostly insignificant, as reported in Table 2. However, this is a misleading artefact 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% to 7%, the slope of the response probability turns steeply positive. This may imply that when inflation moves out of the “rational inattention” zone, where economic agents hardly notice

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



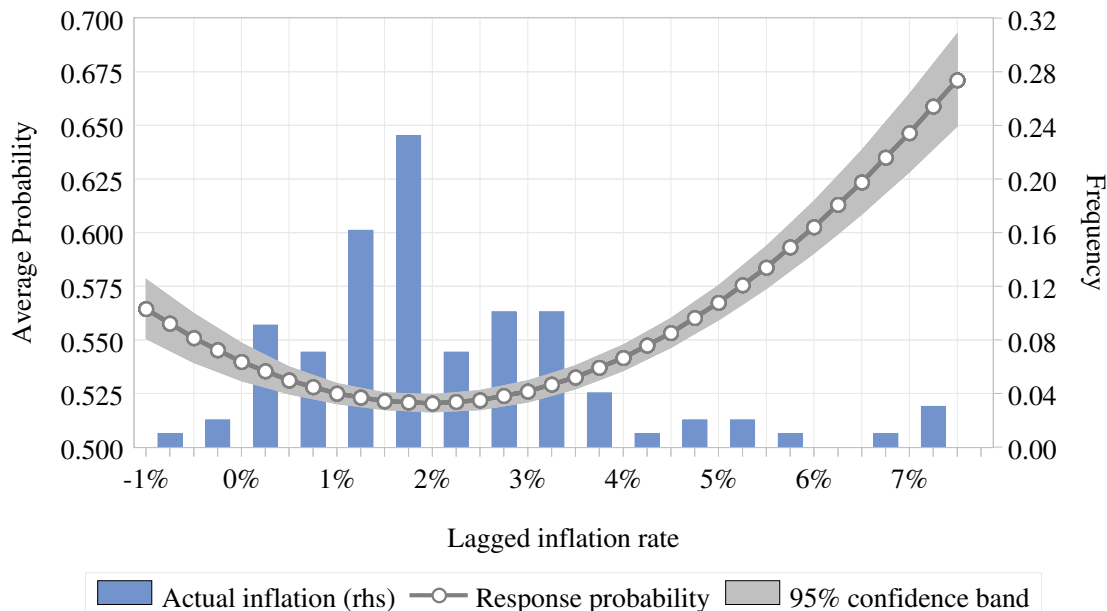
Notes: Response probabilities are based on the baseline Probit model estimates, column (1) from Table A.2. They are obtained as the average of predicted probabilities holding other variables constant at their sample values.

Source: Authors' estimate.



Notes: 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.

Source: Authors' estimate.

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.

Source: Authors' estimate.

inflation, price changes snap into sharp focus and increase agents' ability or willingness to respond to the inflation expectations question.¹³

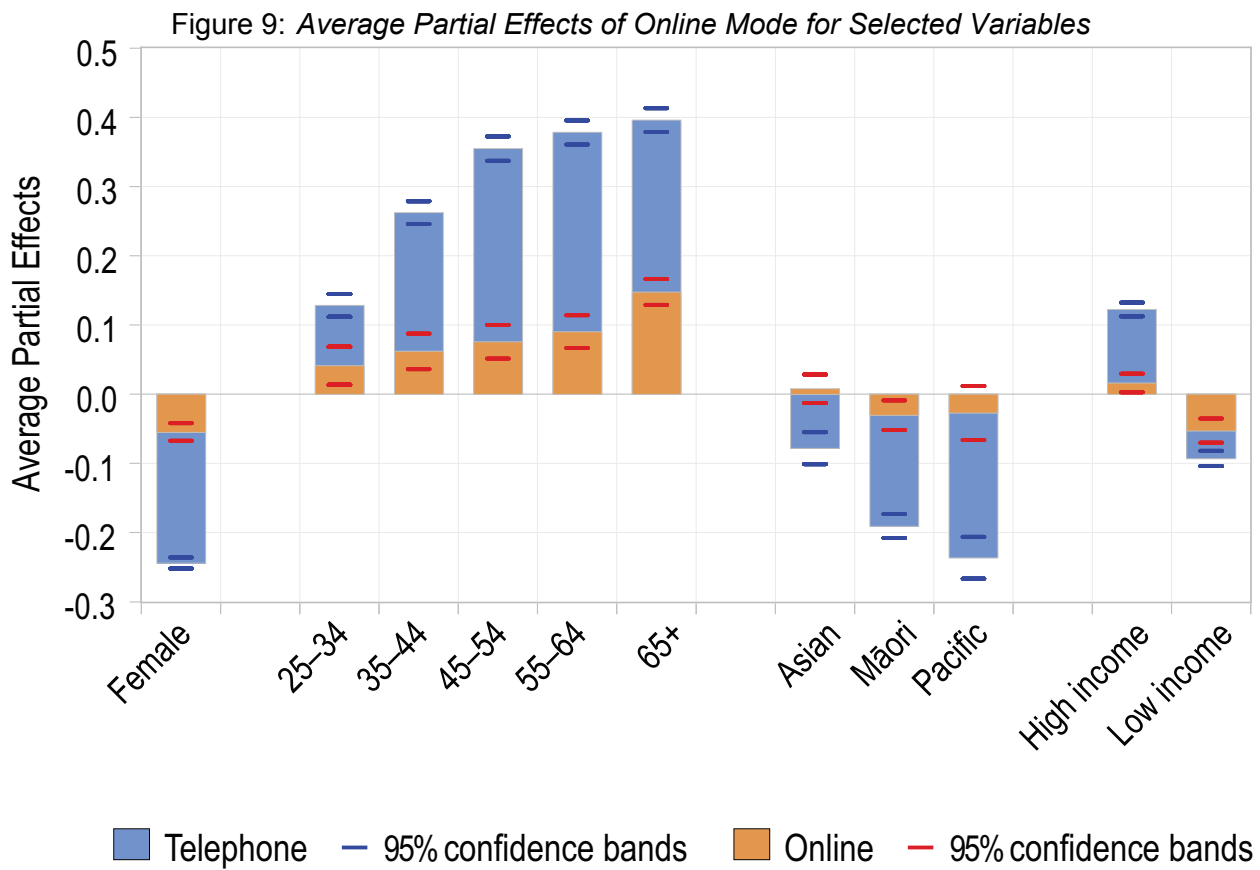
The estimates using the extended specifications (columns 2 to 4 in Table 2) provide additional insights into the determinants of (non-)responses to the inflation expectations question. First, only the more skilled occupations, covering professionals, managers and trained service workers, show a probability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

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

Households with dependent children and those not owning a freehold house tend to respond less to the inflation question. At the same time, the effect of grocery shopping is only significant and negative when the respondent was the main person responsible for that chore. The latter effect can be a critical determinant of the relevance of the beliefs elicited by the survey. Previous research suggests exposure to grocery prices as an essential determinant of consumers' belief formation (D'Acunto et al., 2021; D'Acunto et al., 2023). Although being the main responsible for grocery shopping in the household is found to decrease the probability of responding to the inflation expectations question, the magnitude of that effect is small relative to other determinants.

Finally, we also look at how conducting the survey in online mode can change the effects of demographic characteristics on response probabilities. To do that, we re-estimate our Probit model specifications with additional terms interacting the online dummy variable with the demographic variables. These estimates are summarized in Figure 9, focusing on the case extending our baseline specification.

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



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 6 for the corresponding numerical estimates and other specifications.

Source: Authors' estimate.

4 Inflation Expectations Bias

4.1 Econometric Framework

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,

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

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.

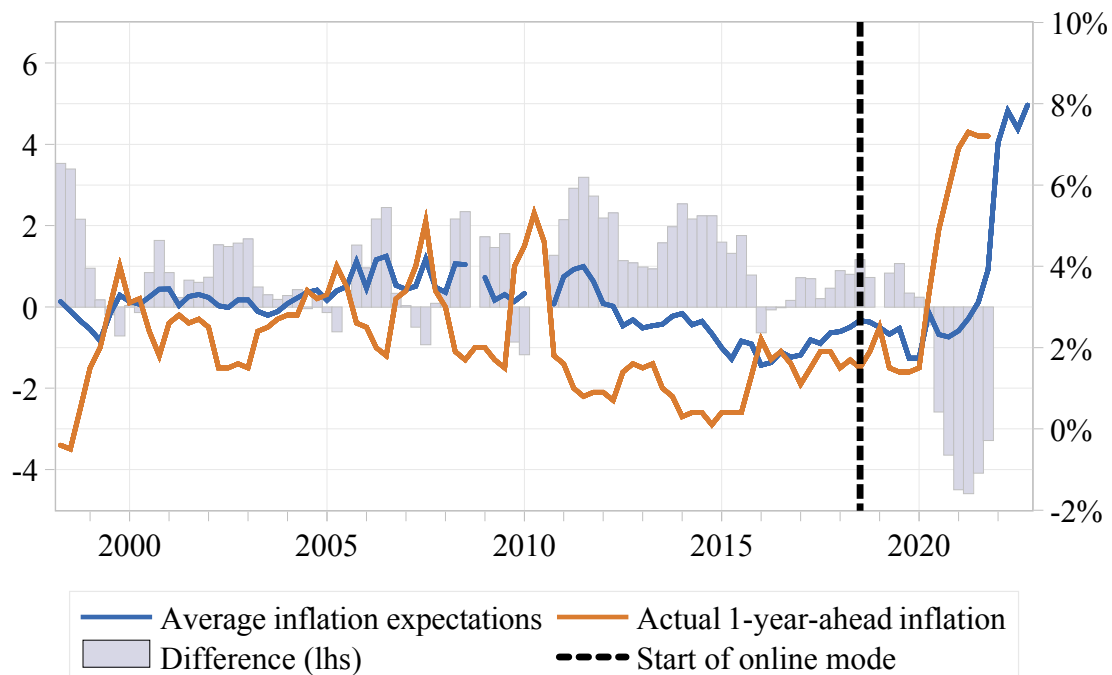
The richness of information available in the microdata from the survey of households' expectations allows us to go one step further and attempt to understand how inflation expectations bias depends on household characteristics. For that purpose, we run regressions of the following form:

$$\pi_i^e - \pi_{t+1} = \alpha_0 + \alpha_1 \mathbf{X}_i' + \alpha_2 \mathbf{Z}_i' + \alpha_3 \mathbf{Q}_i + \alpha_4 O_i + u_i, \quad (3)$$

where, as in our previous notation, \mathbf{X}_i' contains household characteristics, \mathbf{Z}_i' contains macroeconomic variables at the time the respondent answered the survey, \mathbf{Q}_i are quarter dummy variables, and O_i is a dummy variable for the survey mode. Note the household characteristics and the macroeconomic variables included in Equation 3 need not be the same as those included in the Probit model of Equation 1. As discussed in this section, identification in the sample selection model posits the exclusion of variables that are not relevant to the outcome equation.

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,

Figure 10: Average Inflation Expectations and Actual Inflation Rates



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

Source: Authors' estimate.

the missing responses to the inflation expectation question are not random. 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 chose this specification as a selection equation because most of the other variables we used for robustness usually restrict our sample size. In contrast, the inclusion of the online interaction terms provides 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).

4.2 Estimation Results

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; 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 pp higher than respondents from the younger (<25 years) base age bracket. Pacific Islanders have an average bias of 0.34 pp higher than the base ethnic group (Others/NZ Europeans), while Māori people and Asians have an average bias of 0.16 and 0.13 pp higher than the base group, respectively. Of course, these estimates reflect only the sample of respondents who responded to the inflation expectations question.

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

Variables (base category)	Categories	(1)	(2)	(3)	(4)
		No Selection OLS	No Selection WLS	Heckman Selection Two Step	Heckman Selection Max. Likelihood
Gender (base=Male)	Female	0.240*** (0.023)	0.276*** (0.023)	-0.005 (0.040)	0.105* (0.063)
Age (base=Under 25)	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 (base=European/Other)	Asian	-0.093 (0.065)	0.130* (0.066)	0.045 (0.067)	0.082 (0.069)
	Māori	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 (base=Median)	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 (base=No)	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)
No. of observations		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.

Source: Authors' estimate.

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 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.

These are interesting findings in relation to previous literature. Systematic differences in inflation expectations by gender have long been reported (e.g., Jonung, 1981) and corroborated by many studies across countries. For example, D'Acunto et al. (2021) find that women have persistently higher inflation expectations than men and associate this difference with gender roles in household grocery shores. There is also evidence about differences by race (admittedly, not exactly the same as ethnicity, but a reasonable proxy) and income. Using cross-sectional information from the New York Fed Survey of Consumer Expectations, D'Acunto et al. (2023) document that average inflation expectations of Blacks tend to be above those of Whites and Asian Americans, while lower-income households also tend to report higher expectations. Although their findings are based on different surveys and populations, our results indicate that part of those differences by gender, race, and income may be explained by nonresponse bias. In other words, the sources associated with observed heterogeneity in surveys of inflation expectations across these characteristics are, in fact, determining participation and

responses to the inflation expectations question rather than the subgroup's average expectations.

Third, the effects of age increase in magnitude after accounting for selection, particularly steepening the profile of higher over-predictions for older respondents. According to the two-step estimates (column 3), inflation expectations of respondents older than 35 are more than two times higher than what is estimated without accounting for selection (column 2). This finding can have important implications for the literature on learning from experience. Malmendier and Nagel (2016) showed that the dispersion of household inflation expectations by age can be associated with their lifetime experiences of inflation. Our results suggest that these effects could be even stronger when accounting for nonresponse bias. Besides, an upward-sloping profile of inflation expectations by age is more clearly identified when accounting for selection. Hence, the age effects may also be associated with exposure to prices of age-specific baskets of products that would not be apparent from purely observed data.

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 pp, after controlling for the effects of the other variables.¹⁴

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 gender, income and ethnicity, turn statistically insignificant after accounting for selection bias.

¹⁴The 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 Q3 2020 onwards due to the sluggish adjustment of expectations to the increase of actual inflation rates 4 quarters ahead starting from Q3 2021 (see Figure 10).

5 Adjusted Indices

In this section, we show how the correction for nonresponses 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.,

$$\pi_i^e = \hat{\delta}_1 I(t = 1998q2) + \dots + \hat{\delta}_{99} I(t = 2022q4) + \hat{\epsilon}_i, \quad (4)$$

where $I(\bullet) = 1$ when the condition between brackets is true, and $\{\hat{\delta}_t\}$ 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 nonresponse 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,

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

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 (Q1 2018) to a maximum of -0.135 (Q3 2022).¹⁵

¹⁵ Annual averages of these estimates are also presented in Table A.4 in Appendix 6, 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.

6 Concluding Remarks

In this paper, we provide an attempt to understand the demographic determinants of nonresponses 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 nonresponse 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 low-income households. A switch in the conduct of the survey to online mode is found to substantially decrease nonresponses to the inflation question, as well as decrease 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 primarily due to sample selection bias. 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 nonresponse bias, it does not correct for item nonresponse bias. Because both inflation perceptions and nonresponses 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 is, on average, about 0.30 pp higher than a measure that accounts for item nonresponse bias.

Our findings also allow us to draw 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 responding 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.

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
Māori	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	

Source: Reserve Bank of New Zealand Household Inflation Expectations Survey.

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)
	Māori	-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)

Continued on the next page

Variables	Categories	(1)	(2)	(3)	(4)
	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.o. of observations		75,400	71,799	36,011	35,993
Sample period		Q2 1998 to Q4 2022	Q2 1998 to -Q4 2021	Q2 1998Q2 to Q3 2008	Q1 2009 to Q4 2021
McFadden R^2		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.

Source: Authors' estimate.

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

Variables	Categories	(1)		(2)		(4)	
		Online=0	Online=1	Online=0	Online=1	Online=0	Online=1
Gender	Female	-0.244*** (0.004)	-0.055*** (0.006)	-0.252*** (0.004)	-0.043*** (0.007)	-0.250*** (0.006)	-0.048*** (0.008)
Age	25-34	0.128*** (0.008)	0.041*** (0.014)	0.127*** (0.009)	0.042*** (0.016)	0.083*** (0.017)	0.059*** (0.017)
	35-44	0.262*** (0.008)	0.062*** (0.013)	0.268*** (0.008)	0.059*** (0.015)	0.197*** (0.019)	0.081*** (0.017)
	45-54	0.355*** (0.009)	0.076*** (0.012)	0.352*** (0.009)	0.061*** (0.015)	0.283*** (0.020)	0.078*** (0.017)
	55-64	0.378*** (0.009)	0.090*** (0.012)	0.363*** (0.009)	0.080*** (0.014)	0.277*** (0.020)	0.093*** (0.016)
	65+	0.396*** (0.009)	0.148*** (0.009)	0.367*** (0.009)	0.137*** (0.011)	0.302*** (0.020)	0.133*** (0.015)
Region	Rest of North Island	-0.017*** (0.005)	0.010 (0.007)	-0.015*** (0.005)	0.009 (0.008)	-0.016** (0.007)	0.011 (0.009)
	South Island	-0.007 (0.006)	0.020** (0.008)	-0.003 (0.005)	0.017** (0.009)	-0.015* (0.008)	0.016 (0.010)
Ethnicity	Asian	-0.078*** (0.012)	0.008 (0.010)	-0.080*** (0.012)	0.007 (0.012)	-0.104*** (0.016)	0.004 (0.014)
	Māori	-0.190*** (0.009)	-0.030*** (0.011)	-0.184*** (0.009)	-0.020* (0.011)	-0.180*** (0.014)	-0.028** (0.013)
	Pacific Islander	-0.236*** (0.015)	-0.027 (0.020)	-0.227*** (0.016)	-0.024 (0.023)	-0.222*** (0.023)	-0.042 (0.026)
Real income	High	0.122*** (0.005)	0.016** (0.007)	0.111*** (0.005)	0.001 (0.008)	0.122*** (0.008)	0.022*** (0.008)
	Low	-0.093*** (0.006)	-0.053*** (0.009)	-0.095*** (0.006)	-0.049*** (0.010)	-0.087*** (0.008)	-0.081*** (0.011)
Employment	Employed	0.017*** (0.006)	0.045*** (0.008)			0.004 (0.009)	0.061*** (0.010)
Occupation	Clerks/Admin/Sales			-0.022*** (0.007)	0.088*** (0.010)		
	Manual/Agricultural/Domestic			-0.039*** (0.006)	0.068*** (0.010)		
	Professional/Manager/Trained			0.051*** (0.006)	0.087*** (0.009)		
Dependent children	Yes			-0.032*** (0.005)	-0.003 (0.009)	-0.027*** (0.008)	0.005 (0.009)
Home ownership	Living with parents					-0.101*** (0.020)	-0.266*** (0.019)
	Mortgage					-0.042*** (0.009)	-0.102*** (0.012)
	Other					-0.069*** (0.021)	-0.133*** (0.022)
	Renting					-0.122*** (0.010)	-0.077*** (0.012)
No. of observations		75,400		71,799		35,993	
Sample period		Q2 1998–Q4 2022		Q2 1998–Q4 2021		Q1 2009–Q4 2021	
McFadden R^2		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.

Source: Authors' estimate.

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

IQR = interquartile range.

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 quarters Q1 and Q4.

Source: Authors' estimate.

References

- 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.
- D'Acunto, F., U. Malmendier, and M. Weber (2021, May). Gender roles produce divergent economic expectations. *Proceedings of the National Academy of Sciences* 118(21), 2008534118–.
- 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.
- Jonung, L. (1981). Perceived and expected rates of inflation in Sweden. *The American Economic Review* 71(5), 961–968.
- Leung, C. (2009, June). The demographics of household inflation perceptions and expectations. *Reserve Bank of New Zealand Bulletin* 72, 34–42.

- Malmendier, U. and S. Nagel (2016). Learning from inflation experiences. *The Quarterly Journal of Economics* 131(1), 53–87.
- 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.
- Pfajfar, D. and E. Santoro (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization* 75(3), 426–444.
- Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14(1), 53–68.
- Weber, M., S. Frache, D. Georgarakos, B. Candia, B. H. Meyer, O. Coibion, T. Ropele, S. Kumar, G. Kenny, R. Lluberas, Y. Gorodnichenko, and J. Ponce (2023). Tell Me Something I Don't Already Know: Learning in Low and High-Inflation Settings. Technical report, Chicago Booth Research Paper No. 23-14, Fama-Miller Working Paper, University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2023-102.

Nonresponse Bias in Household Inflation Expectations Surveys

This work shows that some demographic groups end up underrepresented in household inflation expectations surveys due to item nonresponse. Nonresponses can lead to misleading inferences about inflation bias, and this paper shows how to correct for such nonresponse bias in average inflation expectations. These findings have important implications for how central banks use household inflation expectations measures, and how policymakers communicate with the population. One main implication is that policy outreach can be improved with more targeted communications.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68 members—49 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.



ASIAN DEVELOPMENT BANK

6 ADB Avenue, Mandaluyong City

1550 Metro Manila, Philippines

www.adb.org