DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 13/2026

When wording changes what we find: The impact of inflation expectations on spending

Tiziana Assenza
(TSE (UT-Capitole) and IAST)

Stefanie Huber
(University of Bonn and ECONtribute)

Anna Mogilevskaja
(University of Bonn and ECONtribute)

Tobias Schmidt
(Deutsche Bundesbank)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank, Press and Public Relations Division, at the above address or via email: www.bundesbank.de/contact

Internet http://www.bundesbank.de

DOI https://doi.org/10.71734/DP-2026-13
ISBN 978-3-98848-070-5
ISSN 2941-7503

Bundesbank Discussion Paper No 13/2026

When Wording Changes What We Find: The Impact of Inflation Expectations on Spending *

Tiziana Assenza
TSE (UT-Capitole) and IAST

Stefanie Huber†
University of Bonn and ECONtribute

Anna Mogilevskaja‡
University of Bonn and ECONtribute

Tobias Schmidt
Deutsche Bundesbank, Research Center

Abstract

We use a randomized experiment in the Bundesbank Online Panel-Households ($n \approx 3,900$) to show that the estimated link between inflation expectations and household consumption flips sign depending on survey wording. This finding reconciles prior contradictory results and has direct implications for central bank survey design. Our experiment systematically varies elicitation framing of consumption question along three dimensions: the *reference unit* (individual vs. household), the *time horizon* (past one, 3, or 12 months), and the *question type* (attitudinal, planned, qualitative and quantitative recall-based). We find that the time horizon and question type significantly influence the estimated relationship between inflation expectations and durable consumption. While the average effect is weak, its sign and magnitude vary strongly with question design. Planned spending and attitudinal questions, such as whether it is a good time to buy, produce very similar negative associations, suggesting that respondents interpret the former as a proxy for future intentions. In contrast, quantitative recall-based questions on past spending yield a modestly positive link, especially for shorter horizons. These results highlight the critical role of survey design in shaping behavioral measurements, offering a novel explanation for mixed findings in the literature and guidance for both research and policy.

Keywords: expectations; inflation; consumption; household decision making; survey methodology; framing effects; measurement.
JEL Classification: C83, D12, D84, E31.

*A supplementary appendix is available at `https://www.bundesbank.de/dp-en/2026-13-appendix`. The data used in this study are accessible to researchers via established institutional channels. Researchers can apply for access to the Bundesbank Online Panel – Households (BOP-HH) through an application process for Scientific Use Files at the Research Data and Service Centre of the Deutsche Bundesbank (`https://www.bundesbank.de/en/bundesbank/research/rdsc/your-research-project-at-the-rdsc`). The results published and the related observations and analysis may not correspond to results or analysis of the data producers. The U.S. Survey of Consumers is available at `https://data.sca.isr.umich.edu/`. U.S. real personal durable consumption expenditures are at `https://fred.stlouisfed.org/series/DDURRA3M086SBEA`, and German real final consumption expenditures at `https://fred.stlouisfed.org/series/NCRSAXDCDEQ`. The views expressed are those of the authors and do not necessarily reflect those of the Deutsche Bundesbank or the Eurosystem. Assenza acknowledges funding from the French National Research Agency (ANR) under the *"Investissements d'Avenir"* program, grant ANR-17-EURE-0010. Huber and Mogilevskaja acknowledge funding by from the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/2-390838866).

†Corresponding Author. Niebuhrstr. 5, 53113 Bonn, Germany. Email: shuber@uni-bonn.de

‡Corresponding Author. Niebuhrstr. 5, 53113 Bonn, Germany. Email: amogilev@uni-bonn.de

1 Introduction

Macroeconomic models, particularly in the New Keynesian tradition (e.g., Galí, 2008), typically assume that household inflation expectations are a key determinant of consumption behavior. Expected inflation affects the intertemporal trade-off between consumption today and in the future, making it a central driver of aggregate demand. Central banks view inflation expectations as an important channel through which monetary policy influences household behavior and ultimately inflation itself. These theoretical and policy perspectives have motivated a large empirical literature examining whether households expecting higher inflation are more likely to increase current consumption, especially of durable goods, due to intertemporal substitution. Yet the evidence remains mixed. Some studies find a modest positive relationship, while others report no or even negative effects. This mixed empirical evidence poses a challenge for both researchers and policymakers: what explains these diverging results?

We hypothesize that these inconsistencies arise, at least in part, from how consumption is elicited in surveys. In particular, we argue that seemingly minor *elicitation framing differences*, such as how, over what horizon, and for whom consumption is reported, can lead to substantial variation in the estimated sensitivity of consumption to inflation expectations. These differences have not been systematically analyzed, yet they likely contribute to the conflicting evidence across studies.

To test this hypothesis, we implement a unified experimental framework embedded in the Bundesbank Online Panel–Households (BOP-HH), a long-standing, nationally representative monthly survey of individuals in Germany (Deutsche Bundesbank, 2025). We pose seven distinct consumption questions, each reflecting a commonly used formulation in the existing literature. These questions vary along three key elicitation framing dimensions: *reference unit* (individual vs. household spending), *time horizon* (consumption over past one, 3, or 12 months), and *question type* (attitudinal, such as whether it is a good time to buy; planned spending; quantitative recall-based spending, such as actual amounts spent; or qualitative recall-based spending, such as changes in consumption). Respondents are randomly assigned to different subsets of these questions, enabling both within- and between-subject comparisons. By keeping inflation expectations constant and systematically varying only the consumption question, our design allows us to causally identify how elicitation framing affects the measured relationship between inflation expectations and consumption.

Our findings confirm that elicitation framing meaningfully affects the estimated relationship, both in sign and magnitude, between inflation expectations and consumption. While the reference unit (individual vs. household) has no detectable impact, both the time horizon and question type significantly affect the sign and magnitude of the estimated effect. Attitudinal questions, such as whether it is now a good time to buy, yield a negative association with expected inflation, closely mirroring those from forward-looking spending plans. By contrast, quantitative recall-based measures of past consumption show a small but positive relationship, which is significant only for shorter recall periods. The divergence across question types reflects underlying construct differences: attitudinal and planned spending items measure willingness to spend, while recall-based items capture realized consumption shaped by constraints such as income, liquidity, or necessity. Consistent with this interpretation, a confirmatory factor analysis (CFA) confirms that the attitudinal measure loads on a distinct latent factor. In addition, we find that quantitative recall-based measures show the highest (and attitude the lowest) external validity. Overall, elicitation framing not only changes effect sizes but can even reverse their sign. This helps reconcile conflicting results in the literature and provides practical guidance for survey design and policy interpretation.

These findings have implications for three strands of literature: survey framing, the inflation-expectations-consumption relationship, and the empirical measurement of consumer sentiment.

First, we contribute to the survey design and framing literature by implementing a unified experimental design that identifies how different dimensions of consumption question framing shape the estimated relationship between inflation expectations and consumption. We formalize a simple typology of framing dimensions and show how each systematically influences the estimates. These results connect to a broader literature on survey methodology showing that survey design features, including wording, question order, context, and reference points, can meaningfully influence respondents' answers and their interpretation of the question (e.g., Brañas-Garza et al., 2022; Groves et al., 2011; Kalton and Schuman, 1982; Reitmann et al., 2020; Schuman and Presser, 1996; Tourangeau et al., 2000). In the context of inflation expectations, De Bruin et al. (2012) and De Bruin et al. (2017) find that minor changes in question wording can significantly alter reported inflation expectations. Pavlova (2025) finds that the question format for inflation uncertainty—such as probabilistic versus density-based approaches—substantially shifts point estimates and associated confidence intervals. Nonetheless, not all aspects of framing are consequential;

for example, Chen et al. (2025) find no framing effects when examining a specific type of linguistic framing—tense framing—which refers to whether intertemporal choice questions are phrased in the past, present, or future tense.

Our work also contributes to a growing literature on survey design and cognition in economics. Early work by Scheufele (2000), along with more recent contributions by Stantcheva (2023) and Goldin and Reck (2020), emphasizes how framing, priming, and agenda-setting affect how individuals process information and formulate survey responses. These insights, rooted in political communication and cognitive psychology, highlight that survey responses are not passive reflections of preexisting beliefs, but constructed in interaction with the survey instrument. Saris and Gallhofer (2014) further show that not only the framing, but also the structure of response categories, can systematically influence the measurement of underlying attitudes and behaviors.

We extend this literature by applying framing variation not to expectations per se, but to the elicitation of realized and intended economic behavior—specifically, household consumption. In doing so, we also contribute to the empirical literature on inflation expectations and consumption by showing that part of the conflicting evidence in prior studies arises from how consumption is elicited. Our unified design allows us to replicate the seemingly contradictory findings reported across prior studies within a single dataset, demonstrating that these differences arise systematically from how consumption is elicited. We provide a coherent explanation for prior inconsistencies, showing that framing effects reflect underlying construct differences rather than just behavioral contradictions.

Second, we contribute to the literature on inflation expectations and household consumption by developing a conceptual framework and typology that clarify how survey design choices might affect the estimated relationship between inflation expectations and consumption. We classify 13 empirical studies by country, data source, consumption question framing, and estimated effect, revealing systematic variation that maps onto differences in framing of the consumption question. This structured overview provides both a diagnostic tool for interpreting prior evidence and a reference point for future research.

A close look at the literature supports the need for such a framework. Studies using attitudinal questions tend to find mixed results, ranging from negative effects of inflation expectations on consumption (e.g., Bachmann et al., 2015, but only at the zero-lower bound; Andrade et al., 2023) to positive effects (e.g., D'Acunto et al., 2022; D'Acunto et al., 2023). Studies using quantitative or qualitative measures of past consumption often report positive effects (e.g., Burke and Ozdagli, 2023, but only for highly-educated

households; Dräger and Nghiem, 2021; Ichiue and Nishiguchi, 2015). Yet, several studies also report null results (e.g., Bachmann et al., 2015; Andrade et al., 2023; Burke and Ozdagli, 2023). This empirical heterogeneity highlights the value of our typology. It helps reconcile divergent findings by tracing them to variation in question framing rather than underlying behavioral differences. Replications of three prior studies using our unified dataset provide compelling evidence that variation in question framing alone can change the outcome of the estimated relationship.

Third, we contribute to research on consumer sentiment and survey measurement validity (e.g., Katona, 1975; Williams and Defris, 1981; Abeele, 1983) by investigating how well alternative consumption measures reflect actual behavior. We build on seminal work by Katona (1975), who emphasized that consumption depends not only on the *ability* to buy, but also on the *willingness* to buy. To operationalize this distinction, he developed the Index of Consumer Sentiment (ICS) within the Michigan Survey of Consumers. This index includes questions on households' financial outlook, macroeconomic expectations, and attitudes toward durable consumption. The latter, an attitudinal measure, has since become one of the most widely used proxies for consumption in studies on households' inflation expectations and behavior. However, already in the 1980s, several studies questioned the behavioral validity of such attitudinal proxies. For example, Williams and Defris (1981), Abeele (1983), Stuart (1984) and Van Raaij and Gianotten (1990) showed that while consumer sentiment indicators correlate with attitudinal measures, objective indicators of financial capacity, such as income or liquidity, are often stronger predictors of actual spending.

Our experimental design extends this literature by disentangling willingness, ability, and necessity (the perceived urgency or indispensability of a purchase) as distinct drivers of consumption. We assess the robustness and generalization of our findings by replicating key results from prior studies and examining both the internal and external validity of alternative consumption measures. This richer typology allows us to assess whether inflation expectations influence willingness to spend, realized spending, or both, and under what conditions. Our results show that while higher inflation expectations often decrease the willingness to spend, this effect rarely translates into realized spending, likely due to financial constraints or replacement needs. This interpretation is reinforced by our own survey data: when asked why they make durable purchases, only a small share of respondents mentioned inflation expectations, while the vast majority pointed to necessity (e.g., replacing a broken device). Attitudinal questions therefore behave more like

forward-looking indicators of willingness to spend than as reflections of realized behavior, confirming that different consumption measures capture distinct underlying constructs. Recognizing these construct differences also explains why prior studies using different question framings have produced systematically divergent results. Hence, we contribute to the literature by providing new evidence on the interpretation of a widely used consumption measure, the attitudinal question, showing that they capture forward-looking intentions rather than realized spending. This distinction has important implications for economic modeling and policy design.

Our results have important policy implications. Central banks and statistical agencies rely on household surveys to track consumption behavior and inflation expectations. If survey design systematically affects the estimated responsiveness of consumption to expected inflation, policy calibration may be distorted. Some framings may understate the expectations channel in monetary policy transmission, while others may exaggerate substitution motives. For expectations-based policy to be effective, especially near the zero lower bound, survey instruments must be designed and interpreted with careful attention to framing effects.

The remainder of the paper is organized as follows. Section 2 describes the data and the experimental design. Section 3 presents the main results. Section 4 discusses robustness, replications, and measurement validity. Section 5 concludes with policy implications. Appendix A provides a structured overview of prior studies, serving as a reference for future work on this contested relationship.

2 Survey Design and Randomized Framing Variation

2.1 Data

To study how elicitation framing affects consumption responses to inflation expectations, we use novel survey data collected through the Bundesbank Online Panel – Households (BOP-HH). This large-scale, high-frequency panel allows us to embed a randomized module into an otherwise representative household survey in Germany. Respondents for the BOP-HH are randomly drawn from an online access panel of individuals aged 16 and above, with quotas to ensure representativeness along age, gender, education, and region. For details on the methodology of the BOP-HH survey refer to Beckmann and Schmidt (2020).

In the March 2024 wave, we embedded a randomized survey experiment that varied the

elicitation framing of consumption questions along multiple dimensions. All respondents first answered a set of core questions included in every BOP-HH wave on the macroeconomic outlook, perceived inflation, one-year-ahead inflation expectations, and income expectations, establishing a common baseline. This block of questions was followed by questions on planned spending and quantitative recall-based past month's spending. After four unrelated survey items, respondents completed the attitudinal question and were then randomly assigned to one of two groups, each receiving a distinct set of follow-up consumption questions that varied along key elicitation framing dimensions. The survey concluded with questions regarding socio-demographic characteristics (for details on the full survey flow refer to Online Appendix A). The summary statistics are provided in Online Appendix Table B.1.

The sample size of approximately 3,900 respondents is based on the Bundesbank's underlying survey, which typically includes around 4,000 participants per month. While this sample size is driven by data availability, it is broadly consistent with those used in other survey-based studies examining the relationship between inflation expectations and consumption (Burke and Ozdagli, 2023; Coibion et al., 2023; Dräger and Nghiem, 2021).

Our survey module was administered without monetary or other financial incentives. However, participants in the BOP-HH surveys accumulate "points" for completing the whole questionnaire. These points can be redeemed for lottery tickets or donations but cannot be converted into cash payments.

Table 1 provides an overview of the consumption questions included in our module. This design yields a rich set of consumption-related outcomes across the different elicitation framings. All respondents answered multiple versions of consumption questions and were randomly assigned to groups subjected to different elicitation framing conditions. These conditions varied along three key dimensions: *question type* (attitudinal, planned, qualitative and quantitative recall-based), *temporal framing* (past one, 3, 12 months), and *reference unit* (individual vs. household).

All survey materials—including question wording and treatments texts—were reviewed by the Bundesbank team overseeing the BOP-HH and by the professional survey company responsible for survey implementation and panel management. The survey data are fully anonymized and the experiment did not involve sensitive content. Therefore, no additional Institutional Review Board approval was commissioned. In addition, the dataset stands out for its exceptionally low rate of missing responses (Online Appendix Table B.2).

Code	Question Type	Question Summary and Answer Categories
F1	Planned spending	Are you likely to spend more or less on major items over the *next twelve months* than in the last twelve months? (i) Plan to spend less (ii) Plan to spend roughly the same (iii) Plan to spend more
F2	Quantitative recall-based (1 month)	If you think back to *last month*, roughly how much did you spend in euro on the following items? [Input: Euro amount]
F3	Attitude	In view of the general economic situation, do you think that now is a good time to make major purchases? (i) Yes (ii) Neither good nor bad (iii) No
F4	Quantitative recall-based (3 months)	Over the *past three months*, roughly how much did your household spend on major purchases, in euro? [Input: Euro amount]
F5	Quantitative recall-based (12 months)	Over the *past twelve months*, roughly how much did your household spend on major purchases, in euro? [Input: Euro amount]
F6	Qualitative recall-based (individual purchases)	How high was *your individual spending* on major purchases over the *past twelve months* compared with an average year? (i) Significantly lower (ii) Roughly the same (iii) Significantly higher
F7	Qualitative recall-based (household purchases)	How high was *your household's spending* on major purchases over the *past twelve months* compared with an average year? (i) Significantly lower (ii) Roughly the same (iii) Significantly higher

Notes: Questions F4 and F5 were displayed conditionally. Respondents were first asked whether they had made major purchases during the past 3 months and/or the previous 4-12 months. If they answered yes for a given period, they were shown the corresponding spending question (F4 or F5). If they had made purchases in both periods, they received F5 as a follow up. F7 was only asked if the individual lives in a household with more than one member.

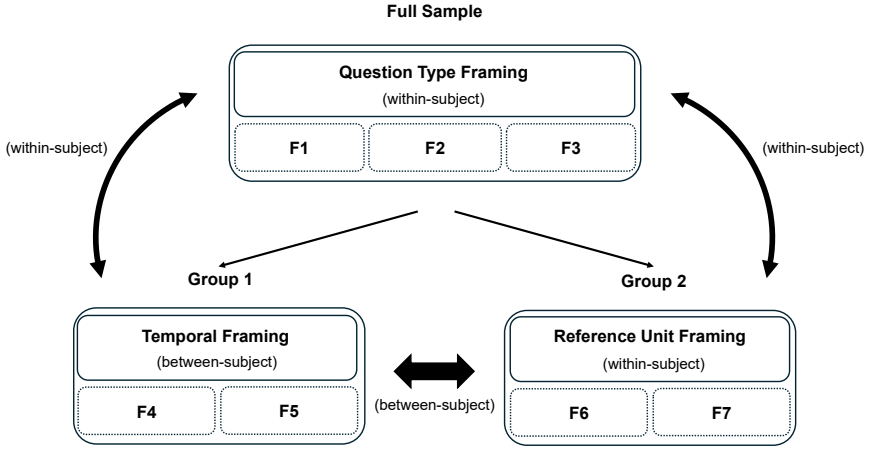
Table 1: Overview of Consumption Survey Questions and Framing Dimensions

2.2 Details on Framing Variation

The survey incorporated a randomized questionnaire experiment to examine whether the empirical relationship between inflation expectations and consumption varies with how consumption questions are framed. A key focus of the design is on *question-type framing*, as we hypothesize that differences in attitudinal versus recall-based consumption questions—often overlooked—contribute significantly to the inconsistent empirical findings in the literature. Our design allows both within-subject and between-subject comparisons across key framing dimensions.

Full sample (Question-Type Framing). All participants were asked three baseline questions on durable consumption: planned spending (F1), quantitative recall-based spending in the past month (F2), and attitude towards spending (F3). These questions establish a common baseline and allow for comparison across different formats within the same individuals. To limit respondent fatigue, the full sample was then randomly split into two groups. Each group received a distinct set of follow-up consumption questions, differing in specific elicitation framing dimensions (see Figure 1 for an overview).This setup allows for both within-subject and between-subject variation in question framing.

The experiment relies on random allocation of respondents to the groups. Online Appendix Table B.3 confirms that this criterion is met, as there are no statistically significant differences between Group 1 and Group 2 in terms of age, education, income, gender, unemployment, and region.



Notes: This diagram illustrates the structure of the randomized framing experiment implemented in the BOP-HH. All respondents answer three baseline questions—planned spending (F1), 1-month quantitative recall-based spending (F2), and attitude (F3)—enabling within-subject comparisons across question types. Group 1 answers either the 3-month (F4) or 12-month (F5) quantitative recall-based question, enabling between-subject comparisons of temporal framing (Hypothesis Test 2) and within-subject comparison with the baseline—namely, F2 vs. F4/F5 (temporal framing, Hypothesis Test 3) and F3 vs. F4/F5 (question-type framing, Hypothesis Test 5). Group 2 answers qualitative recall questions at the individual (F6) and household (F7) levels, allowing within-subject tests of reference-unit framing (F6 vs. F7, Hypothesis Test 1) and of question-type framing with attitude (F3 vs. F6/F7, Hypothesis Test 6). The horizontal arrow between Groups 1 and 2 represents a between-subject comparison of F5 and F7 (question-type framing, Hypothesis Test 4). Finally, comparing F3 and F1 within the full sample identifies differences between attitudinal and planned spending (question-type framing, Hypothesis Test 7). Curved arrows indicate within-subject comparisons across either temporal or question-type framing.

Figure 1: Survey Design with Randomized Framing Variation

Group 1 (Temporal Framing). This group received follow-up questions focusing on time-horizon framing of consumption. Specifically, respondents were asked about their durable expenditures over two distinct periods. Question F4 elicited quantitative recall-based spending on durable goods over the past 3 months, while question F5 addressed quantitative recall-based spending over the past 12 months (both in euro amounts). For practicality, these questions were administered conditionally: respondents first indicated whether they had made any major purchases in the last 3 months or during the 4-12 months prior. Those who had were then asked the corresponding spending question (F4 or F5). If a respondent reported purchases in both periods, they answered quantitative recall-based spending over the past 12 months (F5); if they had none, they skipped the detailed spend-

ing questions. This design introduces both within-subject and between-subject variation in temporal framing as some individuals report past spending over a short horizon, others over a longer horizon, allowing us to test whether the measured effect of inflation expectations differs between short-run and long-run spending measures.

Group 2 (Reference Unit Framing). This group received follow-up questions focusing on the reference-unit framing of consumption. Respondents answered qualitative recall-based questions about durable spending over the past year, distinguishing between their personal spending and their household's total spending. Specifically, question F6 asked respondents to assess how their individual spending on major purchases over the last 12 months compared to an average year, while question F7 referred to the same for their household's spending. All respondents from multi-person households were asked both qualitative recall-based measures (F6 and F7), providing a within-subject variation in reference unit (individual vs. household). Respondents living alone (single-person households) received only the individual recall-based spending question (F6), since for them, personal and household spending are equivalent. This design allows us to directly compare how framing the question at the individual versus household level might influence the relationship between inflation expectations and reported spending behavior.

Overall, our design introduces randomized variation in elicitation framing along three conceptual dimensions: reference unit framing (individual vs. household expenditure), temporal framing (short vs. long recall periods), and question-type framing (attitudinal, planned, qualitative and quantitative recall-based). Within each group, respondents are exposed to more than one question variant, enabling within-subject comparisons across elicitation framing dimensions. For instance, individuals in Group 1 report quantitative recall-based spending over different time horizons (F4 vs. F5), while those in Group 2 compare individual and household-level qualitative recall-based expenditures (F6 vs. F7). Additionally, all participants answer questions on planned spending, quantitative recall-based spending in the past month and attitude (e.g., F1–F3), facilitating within-subject comparisons of questions types.

Beyond these within-subject variation, the randomized assignment to Group 1 or Group 2 introduces between-subject variation. Group 1 receives quantitative recall-based questions, whereas Group 2 receives qualitative recall-based assessments, enabling comparisons of question type across randomly assigned sub-samples. A key focus of our experimental design is on question-type framing, as we hypothesize that many of the inconsistencies in the existing empirical literature—ranging from positive to null or even negative effects of

Framing Type	Question Type	Hypothesis Tests
Reference Unit	Qualitative recall-based F6/F7	$H1$: F6 vs. F7
Temporal	Quantitative recall-based F4/F5	$H2$: F4 vs. F5
Temporal	Quantitative recall-based F2/F4/F5	$H3$: F2 vs. F4/F5
Question type	Quantitative vs. Qualitative recall-based	$H4$: F5 vs. F7
Question type	Attitude vs. Quantitative recall-based	$H5$: F3 vs. F4/F5
Question type	Attitude vs. Qualitative recall-based	$H6$: F3 vs. F6/F7
Question type	Attitude vs. Planned spending	$H7$: F3 vs. F1

Notes: We test elicitation framing dependencies along three dimensions: *reference unit*, *temporal horizon*, and *question type*. For reference unit and temporal framing, we use within-question-type variation. For question-type framing, we use across-question-type variation. Hypothesis 1 tests reference unit framing by comparing qualitative recall-based individual (F6) and household (F7) spending. Hypothesis 2 and 3 examine temporal framing. Hypothesis 2 compares 3-month (F4) and 12-month (F5) quantitative recall-based spending. Hypothesis 3 compares 1-month (F2) with pooled responses from 3-month (F4) and 12-month (F5) quantitative recall-based spending. Hypothesis 4-7 evaluate question-type framing. Hypothesis 4 compares household-level quantitative (F5) and qualitative (F7) recall-based spending over 12 months. Hypothesis 5-7 contrast attitude (F3) with quantitative (F4/F5) and qualitative (F6/F7) recall-based spending, and planned spending (F1), respectively.

Table 2: Hypothesis Tests on the Effects of Framing Variation

inflation expectations on consumption—are driven by differences in how consumption is elicited.

To this end, our framing module includes multiple randomized question type variants that compare attitudinal, qualitative, and quantitative measures of spending intentions and behavior. This layered structure allows us to systematically isolate how variation in reference unit, recall horizon and question type influences the estimated relationship between inflation expectations and consumption. The corresponding hypothesis tests are summarized in Table 2.

3 Results

This section presents our main result, structured by framing dimension to assess how survey design influences the estimated relationship between inflation expectations and consumption. We present results for three types of elicitation framing: (i) reference unit, (ii) temporal horizon, and (iii) question type. The first two dimensions use *within* question-type variation, while the third relies on *across* question-type comparisons. Each dimension isolates a distinct aspect of question design that may affect the relationship between consumption and expected inflation. Additional details on the estimation strategies and hypothesis tests are provided in Appendices B.2 and B.3.

3.1 Within Question-Type Variation

Consumption questions may refer either to the respondent's own spending or to household-level spending. We refer to this distinction as *reference unit framing*. This design choice may matter because individuals may interpret the inflation's impact differently when thinking about their personal consumption versus household-wide spending, especially in multi-person households with unequal income and spending roles.

The attitudinal consumption measure (F3), widely used in the literature, typically lacks a defined reference unit—leaving it ambiguous whether respondents consider individual- or household-level behavior. By contrast, most qualitative or quantitative recall-based measures explicitly refer to household-level consumption (e.g., Ichiue and Nishiguchi, 2015; Dräger and Nghiem, 2021). Some exceptions exist, e.g., Andrade et al. (2023) elicit realized consumption over the past 12 months without specifying the unit. These differences raise the question of whether reference unit framing might contribute to the divergent empirical evidence on how inflation expectations affect consumption.

We begin by testing whether the relationship between inflation expectations and reported durable spending depends on the reference unit specified in the survey question, i.e., whether respondents report their own spending or that of their household. To this end, we compare individual- (F6) and household-level (F7) qualitative recall-based responses, both answered by all respondents in Group 2 living in multi-person households. To mitigate potential order effects, we randomized the order of questions F6 and F7 across respondents. We estimate the following specification:

$$C_{i,t}^{j} = \beta_0 + \beta_1 E_{i,t}(\pi_{t+12}) + \beta_2 E_{i,t}(\pi_{t+12}) \times Treat_i + \beta_3 Treat_i + \beta_4' X_{i,t} + \epsilon_{i,t}, \qquad (3.1)$$

where $C_{i,t}^{j}$ is qualitative recall-based spending of individual i for reference unit $j \in \{\text{individual i; household}\}$, and $E_{i,t}(\pi_{t+12})$ captures their 12-months ahead inflation expectations as measured in month t. The indicator $Treat_i$ equals one if the response refers to household spending (F7), and zero if it refers to individual spending (F6). The interaction term tests whether the effect of inflation expectations on consumption differs across reference units. The vector X_i includes controls for individual i, such as macroeconomic expectations and socio-demographic characteristics known to influence consumption: age, age squared, gender, household income, education, employment status, East Germany dummy, household size, and city size. The error term is denoted by ϵ. Equation (3.1) is estimated with an ordered probit model to account for the ordinal nature of the consumption question,

incorporating individual random effects to capture unobserved heterogeneity. Standard errors are clustered at the individual level.

Table 3, Column 1, presents the results. We find no statistically significant direct effect of expected inflation on qualitative recall-based spending. The interaction term—capturing household-level reporting—yields also no significant effect. This suggests that the effect of inflation expectations on durable spending does not depend on whether spending is reported at the individual or household level. One possible explanation is that durable goods are typically shared within households, leading to similar reports regardless of reference unit. This interpretation is consistent with the insignificant coefficient on the treatment dummy itself, indicating no systematic difference in average spending between the two reporting modes.

As a robustness check, we also estimate the effect of inflation expectations on each qualitative consumption measure separately (F6, F7). Results, reported in Online Appendix Table B.4, are consistent with our baseline findings. Furthermore, we recode qualitative recall-based spending to reflect real changes—classifying responses as real increases when nominal spending rises, and real decreases otherwise—and re-estimate equation (3.1). The results remain virtually unchanged (Online Appendix Table B.5).

Result 1 (No evidence that Reference Unit Framing matters). We find no evidence that framing the consumption question in terms of individual versus household spending systematically affects the estimated effect of inflation expectations on reported consumption.

Survey questions on consumption vary in their time horizons, a distinction we refer to as *temporal framing*. Some questions use short recall periods (e.g., one or 3 months), while others span longer intervals (e.g., 12 months). These differences may affect how reported consumption responds to expected inflation.

Attitudinal consumption measures lack an explicit time frame, making it ambiguous whether responses refer to past or future consumption. By contrast, recall-based or planned spending questions typically specify a clear time horizon. For example, Burke and Ozdagli (2023) use a 3-month window for durable spending and a 1-month horizon for non-durable spending, while Andrade et al. (2023), Dräger and Nghiem (2021) and Ichiue and Nishiguchi (2015) employ a 12-month recall period.

Framing	Reference Unit	Temporal	
	$H1$	$H2$	$H3$
Measures	F6 vs. F7	F4 vs. F5	F2 vs. F4/F5
Treatment	F7	F4	F2
	(1)	(2)	(3)
Inflation Expectations $E_{i,t}(\pi_{t+1})$	-0.0023	-0.0166	0.0084
	(0.0212)	(0.0187)	(0.0165)
$E_{i,t}(\pi_{t+1}) \times Treatment$	0.0334	0.0776*	0.0056
	(0.0192)	(0.0324)	(0.0230)
$Treatment$	0.0214	−0.6923***	−1.7226***
	(0.0828)	(0.1526)	(0.1069)
Individual Random Effects	+		+
Constant	+	+	+
Controls	+	+	+
N	2,584	840	1,279
(Pseudo) R^2	0.0137	0.0917	0.2490

Note: This table reports estimation results corresponding to regression specifications (3.1) (Column 1) and (3.2) (Columns 2 and 3) which test the hypothesis 1-3. Column 1 examines whether the relationship between inflation expectations and reported durable spending depends on the reference unit (individual vs. household), based on responses to individual- (F6) and household-level (F7) qualitative recall-based spending. Columns 2 and 3 assess whether the time horizon of the consumption question affects the estimated relationship. Column 2 compares quantitative recall-based spending over the past 3 months (F4) and past 12 months (F5) using a between-subjects design (Hypothesis Test 2). Column 3 contrasts quantitative recall-based spending over the past month (F2) with the pooled responses from F4 and F5 using within-subject variation (Hypothesis Test 3). Inflation expectations are entered as a continuous variable truncated at ±30. All regressions include controls for expected economic growth and a standard set of sociodemographic characteristics: age, age squared, gender, household income, education, employment status, an East Germany dummy, household size, and city size. Individual random effects are included in Columns 1 and 3. In Column 1, the estimation method is ordered probit and in Column 2 and 3 the estimation method is log-linear. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: Reference Unit and Temporal Framing Effects: Hypothesis Tests 1-3

To test for temporal framing effects, we conduct two comparisons. First, we use a between-subject design, restricted to Group 1, comparing quantitative recall-based spending over the past 3 months (F4) and 12 months (F5). Second, using a within-subject design, we compare responses to quantitative recall-based spending over the past month (F2) with the pooled responses from the 3- and 12-month questions (F4/F5). This dual approach allows us to assess whether the responsiveness of reported spending to inflation expectations varies by the time horizon over which consumption is elicited. We estimate the following model for each comparison:

$$C^h_{i,t} = \beta_0 + \beta_1 E_{i,t}(\pi_{t+12}) + \beta_2 E_{i,t}(\pi_{t+12}) \times Treat_i + \beta_3 Treat_i + \beta_4' X_{i,t} + \epsilon_{i,t}, \qquad (3.2)$$

where $C^h_{i,t}$ denotes quantitative recall-based spending over horizon $h \in \{1; 3; 12\}$ measured in months. $E_{i,t}(\pi_{t+12})$ denotes individual i's 12 months-ahead inflation expectations collected at month t. $Treat_i$ is a dummy equal to one if the response refers to the shorter

time horizon (F4 for the between-subject design, F2 for the within-subject comparison). The key coefficient of interest is β_2, which captures whether the effect of expected inflation on reported consumption differs by time horizon—that is, whether temporal framing moderates the inflation-consumption relationship. Controls $X_{i,t}$ include macroeconomic expectations and standard socio-demographics. The error term is denoted by ϵ.

To ensure consistent estimation, we focus on the intensive margin of spending as respondents report amounts only if they previously indicated having made a purchase. Because quantitative recall-based spending is highly right-skewed, we apply a log transformation to approximate normality and stabilize variance. The comparison between F4 and F5 relies on between-subject variation in recall horizons that arises from respondents purchasing behavior, rather than random assignment. The estimation uses robust (Huber-White) standard errors. By contrast, Hypothesis 3—identified through within-person comparisons in Group 1—includes individual random effects and standard errors clustered at the individual level.

Table 3, Column 2, reports the results from the between-subject comparison of the 3-month (F4) versus the 12-month (F5) quantitative recall-based spending. The coefficient on the treatment dummy is negative and highly significant ($p < 0.001$), confirming that reported spending is mechanically lower when the question refers to a shorter time horizon. The interaction term is significant at the 5% level ($p = 0.017$), indicating that the effect of inflation expectations on spending is stronger when the recall period is shorter.

As a robustness check, we estimate the relationship between inflation expectation and consumption separately for 3-month (F4) and 12-month (F5) quantitative recall-based spending (Online Appendix Table B.6). The relationship is positive and significant ($p = 0.025$) for spending over the past 3 months, and insignificant for the 12-months horizon. These results confirm our baseline finding that shorter recall horizons lead to a stronger link between inflation expectations and consumption. One possible explanation is that recent purchases are more salient and therefore recalled more accurately, reducing measurement error in reported spending.

Column 3 presents results from the within-subject comparison between quantitative recall-based spending over the past month (F2) and the pooled reports over the past 3 and 12 months (F4/F5). As in Column 2, the treatment dummy is negative and significant, consistent with the fact that people naturally report higher total spending when asked about a longer period of time. However, the interaction term is small and statistically insignificant, suggesting that the estimated sensitivity of reported consumption to inflation

expectations does not differ across horizons in this within-subject setting.

Hypothesis 3 is identified through within-person comparisons in Group 1. Past-month recall-based spending (F2) is part of the core questionnaire and always preceded the follow-up question (F4 or F5), which depended on reported purchases. Because F2's position was fixed, randomizing its order was not feasible. However, the follow-up questions were separated from F2 by numerous unrelated survey items, making systematic priming or anchoring effects unlikely. To verify this, we re-estimated the specification separately for F2 vs. F4 (3 months) and F2 vs. F5 (12 months) and obtained nearly identical results (Online Appendix Tables B.7 and B.8). These disaggregated estimations confirm that pooling F4 and F5 does not affect the findings, providing strong evidence that order effects do not drive the observed pattern.

Taken together, the results from Columns 2 and 3 suggest that temporal framing can affect the estimated responsiveness of recall-based consumption to inflation expectations. However, this effect appears to be specification-dependent. While the between-subject comparison (Column 2) shows a stronger response at shorter horizons, the within-subject comparison (Column 3) yields no significant difference.

Result 2 (Temporal Framing matters). Reported consumption increases as the recall period lengthens, simply because longer periods capture more purchases. We find some evidence that the responsiveness of recall-based spending to inflation expectations is stronger when the question refers to a shorter horizon. This pattern is not robust across all specifications.

3.2 Across-Question-Type Variation

As noted above, prior studies use different types of consumption questions. Some questions refer to realized consumption, others measure intentions, and some assess attitudes towards spending. We refer to this distinction as *question-type framing*. Attitudinal questions primarily capture willingness to spend, while questions on realized spending reflect a combination of willingness, ability, and necessity (Abeele, 1983; Katona, 1975; Van Raaij and Gianotten, 1990). These conceptual differences suggest that the estimated relationship between inflation expectations and consumption may vary substantially by question type. While most studies rely on attitudinal measures (e.g., Bachmann et al., 2015; D'Acunto et al., 2023, 2022; Marenčák, 2023), others employ quantitative recall-based reports (e.g., Burke and Ozdagli, 2023) or qualitative recall-based assessments (e.g., Ichiue and Nishiguchi, 2015).

Our survey includes four distinct formats: (i) attitudinal questions (e.g., "Is now a good time to buy a durable good?"), (ii) planned spending (e.g., "Do you plan to increase or decrease?"), (iii) qualitative recall-based questions (e.g., "Did your spending increase or decrease?"), and (iii) quantitative recall-based questions (e.g., "How much did you spend?"). We investigate whether the estimated relationship between inflation expectations and consumption depends on the type of question by running a set of pair-wise comparisons. Specifically, we compare: (i) qualitative vs. quantitative recall-based, (ii) attitudinal vs. quantitative recall-based, (iii) attitudinal vs. qualitative recall-based, and (iv) attitudinal vs. planned spending measures. We estimate the following regression specification:

$$C_{i,t}^{k} = \beta_0 + \beta_1 E_{i,t}(\pi_{t+12}) + \beta_2 E_{i,t}(\pi_{t+12}) \times Treat_i + \beta_3 Treat_i + \beta_4' X_{i,t} + \epsilon_{i,t}, \qquad (3.3)$$

where $C_{i,t}^{k}$ is a binary or ordered multinomial measure of consumption from question type $k \in \{\text{attitudinal}, \text{qualitative recall-based}, \text{quantitative recall-based}, \text{planned spending}\}$. To enable comparison across question types, we convert consumption responses into a harmonized binary format as needed. $E_{i,t}(\pi_{t+12})$ denotes the respondent's i 12-months-ahead inflation expectation. $Treat_i$ equals one if the consumption question refers to the alternative format in a given comparison (e.g., in Column 1, $Treat_i = 1$ for F7; Table 4 specifies the treatment mapping for each column). The interaction term tests whether the expectation-consumption link differs by question type k. Controls $X_{i,t}$ follow the previous specifications, and the error term is denoted by ϵ.

The estimation approach varies across hypotheses to reflect differences in design and dependent variable type. Hypothesis 4 is estimated using a logit model on a between-subject design with robust (Huber-White) standard errors. Hypotheses 5–7 rely on within-subject designs that include individual random effects and cluster standard errors at the respondent level. Hypothesis 5 uses a logit specification, while Hypotheses 6 and 7 apply ordered probit models to account for the ordinal nature of the dependent variables.

Column 1 of Table 4 contrasts two measures of recall-based durable spending using between-subject variation: 12-month quantitative (F5) and qualitative recall-based spending over the same period (F7), both at the household level. The treatment indicator equals one for the qualitative framing (F7). The negative and highly significant coefficient ($p < 0.001$) on the treatment dummy indicates that respondents are less likely to report having increased their spending (F7) than to report any spending at all (F5). This may reflect that "increased spending" is interpreted more narrowly, or that it is cognitively easier

to recall whether one spent anything than to assess changes relative to the previous year. The interaction term is insignificant, suggesting no meaningful difference in how inflation expectations relate to spending across these two question types. As a robustness check, however, we estimate the effect of inflation expectations separately for quantitative (F5) and qualitative (F7) recall-based spending. These separate regressions suggest a stronger relationship for the qualitative measure (F7). Expected inflation significantly increases the likelihood of reporting increased spending (F7), but not the likelihood of reporting any spending at all (F5) (Online Appendix Table B.9).

Framing	Question Type			
	$H4$	$H5$	$H6$	$H7$
Measures	F5 vs. F7	F3 vs. F4/F5	F3 vs. F6/F7	F3 vs. F1
Treatment	F7	F4/F5	F6/F7	F1
	(1)	(2)	(3)	(4)
Inflation Expectations $E_{i,t}(\pi_{t+1})$	0.0102	-0.0791^{***}	-0.0320^{***}	-0.0298^{***}
	(0.0133)	(0.0176)	(0.0083)	(0.0058)
$E_{i,t}(\pi_{t+1}) \times Treatment$	0.0377	0.0888^{***}	0.0411^{***}	0.0033
	(0.0213)	(0.0218)	(0.0118)	(0.0076)
$Treatment$	-0.5589^{***}	-1.7862^{***}	0.2197^{***}	-0.0084
	(0.1175)	(0.1221)	(0.0559)	(0.0378)
Individual Random Effects		+	+	+
Constant	+	+	+	+
Controls	+	+	+	+
N	3,096	7,136	3,580	7,196
Pseudo R^2	0.0286	0.1006	0.0257	0.0163

Note: This table reports estimation results corresponding to regression specification (3.3) which test the hypotheses 4–7 on question-type framing. Column 1 (Hypothesis Test 4) compares quantitative (F5) and qualitative (F7) recall-based spending. Quantitative recall-based spending (F5) is coded as a binary indicator for positive expenses, while qualitative recall-based spending (F7) is coded as a binary indicator for increased spending. The treatment is F7. Column 2 (Hypothesis Test 5) compares attitude (F3) with the pooled responses of quantitative recall-based spending (F4/F5). Attitude (F3) is transformed into a binary indicator capturing a positive or neutral spending attitude; quantitative recall-based spending (F4/F5) is coded as binary indicators for positive expenses. The treatment is F4/F5. Column 3 (Hypothesis Test 6) compares attitude (F3) with the pooled responses of qualitative recall-based spending (F6/F7). The treatment is F6/F7. Column 4 (Hypothesis Test 7) compares F3 with planned spending (F1). The treatment is F1. Inflation expectations are entered as a continuous variable truncated at ± 30. All regressions include controls for expected economic growth and a standard set of sociodemographic characteristics: age, age squared, gender, household income, education, employment status, an East Germany dummy, household size, and city size. Columns 1 and 2 use a logit model, while Columns 3 and 4 use an ordered probit model. All specifications include individual random effects, except for the estimation in Column 1. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Question-Type Framing Effects: Hypothesis Tests 4-7

Hypothesis 4 leverages the random assignment of respondents to Groups 1 and 2. While assignment to the groups is random, answering the 12-month quantitative recall-based measure (F5) is conditional on respondents reported purchasing behavior and answering the household-level qualitative recall-based measure (F7) is conditional on respondents'

household size. This introduces behavioral and demographic selection into the respective subsamples, but since this behavior is orthogonal to treatment assignment, the comparison between F5 (Group 1) and F7 (Group 2) remains valid. To support this claim, we report balancing tests comparing F5 and F7 respondents in Online Appendix Table B.10. The results show that observable characteristics—such as age, gender, employment status, and region of residence—are well balanced across the two groups. Differences in income and education reflect the exclusion of one-person households in Group 2 rather than any failure of randomization. When pooling household- and individual-level recall-based responses (F6 and F7), the sample becomes fully balanced, and our results remain robust (Online Appendix Tables B.11–B.12).

Column 2 uses within-subject variation to compare attitudes (F3) and 3- and 12-month quantitative recall-based spending (F4/F5). The treatment indicator equals one for the quantitative recall-based question-type framing (F4/F5). The negative and highly significant treatment coefficient ($p < 0.001$) shows that respondents are more likely to express a positive or neutral attitude than to report realized spending, highlighting a gap between sentiment and behavior. Inflation expectations are negative and significant ($p < 0.001$), indicating that higher expected inflation reduces the willingness to buy. The positive and highly significant interaction term ($p < 0.001$) shows that this relationship is weaker when consumption is elicited via quantitative recall-based questions rather than attitudinal questions. Thus, attitudinal questions yield stronger associations with inflation expectations than recall-based reports, possibly because inflation expectations primarily affect the willingness to spend—captured by attitude—whereas the effect on quantitative recall-based reports on realized consumption is attenuated by constraints on ability (e.g., income or liquidity) and the necessity to purchase durable goods.

Hypothesis 5 relies on within-person comparisons in Group 1. Given the large number of consumption-related survey questions and the surveys module structure, the relative order of attitude (F3) and the quantitative recall-based measures (F4/F5) could not be randomized. Robustness checks comparing F3–F4 and F3–F5 separately yield highly consistent results (Online Appendix Tables B.13–B.14). Moreover, since past-month recall-based measure (F2) is conceptually similar to the 3-month (F4) and 12-month (F5) quantitative recall-based measures and precedes F3 in the questionnaire, we use the F2–F3 comparison as an additional test for potential ordering effects (Online Appendix Table B.15). The results closely align with our main findings, indicating that pooling responses or order effects is unlikely to drive our results.

Column 3 compares attitudinal responses (F3) to qualitative recall-based spending (F6/F7), again exploiting within-subject variation. The treatment indicator equals one for the qualitative questions. The positive and highly significant treatment effect ($p < 0.001$) shows that respondents are more likely to report increased past spending (F6/F7) than to say it is a good time to buy (F3). While the inflation expectations coefficient remains negative and significant ($p < 0.001$), the interaction effect is again positive and significant ($p < 0.001$), implying that this relationship is weaker when consumption is elicited qualitatively.

As a robustness check, we recode qualitative recall-based spending to reflect real changes, classifying responses as real increases when nominal spending rises and as real decreases otherwise. The results remain robust to this alternative specification (Online Appendix Table B.5).

Hypothesis 6 relies on within-person comparisons. F3, F6 and F7 are part of the module we introduced into the survey. Due to the large number of consumption-related survey questions, we were only able to randomize the order of F6 and F7. Robustness checks comparing F3–F6 and F3–F7 separately yield highly consistent results across specifications, indicating that pooling responses or order effects is unlikely to drive our findings (Online Appendix Tables B.16–B.17).

Taken together with the results from Column 2, these findings reinforce the idea that attitudinal measures are more sensitive to inflation expectations than either qualitative or quantitative recall-based measures, and that inflation expectations are primarily related to the willingness to spend.

Column 4 compares attitudinal responses (F3) to planned spending (F1), exploiting within-person variation. The treatment indicator equals one for the planned spending question. Because F1 is part of the Bundesbank Online Panels core questionnaire, the relative order of the two questions could not be randomized. Importantly, they are separated by several unrelated survey items, which reduces the risk of immediate order effects. Neither the treatment effect nor the interaction term is statistically significant. This implies no meaningful difference in how inflation expectations relate to attitudes versus forward-looking plans—both appear to capture similar sentiment. As a robustness check, we estimate the effect of expected inflation separately for attitude (F3) and planned spending (F1) and find very similar effect sizes (Online Appendix Table B.18).

Taken together, these results indicate that attitudinal consumption responses align more closely with forward-looking spending plans than with recall-based reports of past

consumption. This suggests that the observed question-type framing effects arise because these measures capture different constructs of consumption. Specifically, both attitude and planned spending reflect willingness to spend, whereas recall-based measures capture realized consumption. Consistent with this, we find no question-type framing effects when comparing quantitative and qualitative recall-based measures—i.e., when the construct is the same. Because attitudinal and planned spending are less susceptible to recall bias and past consumption constraints, they exhibit a stronger association with inflation expectations than recall-based measures.

Result 3 (Attitudes Resemble Plans, not Behavior). Attitudinal responses align more closely with forward-looking consumption plans than with recall-based reports of past spending, consistent with both capturing the same underlying construct of willingness to spend. By contrast, recall-based questions—whether qualitative or quantitative—reflect realized consumption that is potentially constrained by other factors such as income or liquidity constraints.

Result 4 (Question-Type Framing Reflects Construct Differences). The estimated relationship between inflation expectations and durable consumption varies systematically with the type of survey question used to elicit consumption. Attitudinal and planned-spending measures—both capturing willingness to spend—yield stronger (more negative) associations with expected inflation than recall-based measures, which capture realized spending. No framing effects emerge when comparing different recall-based measures or when comparing attitudes with planned spending, indicating that question-type framing arises only across different constructs, not within them.

These findings demonstrate that the estimated relationship between inflation expectations and durable consumption is highly dependent on how consumption is elicited in the survey. Across the framing dimensions—time horizon, and question-type—we find that both the sign and magnitude of the effect vary systematically. Some framing effects, such as the stronger association from attitudinal questions, are consistent across specifications; others—like the effect of recall period—are more specification-sensitive. These findings help reconcile the empirical inconsistencies in the literature by showing that divergent results can arise from differences in survey design alone. Our unified framework shows that the various survey questions used to measure consumption in the related literature are

not interchangeable. For example, the attitudinal consumption question yields stronger correlations with inflation expectations than recall-based spending reports, reflecting that they capture distinct constructs (e.g. willingness to spend vs. realized consumption). For researchers and policymakers relying on survey data, this highlights the importance of accounting for framing effects when interpreting consumption responses or when comparing results across studies.

4 Robustness, Replication, and Measurement Validity

The presented results suggest that the estimated relationship between inflation expectations and durable consumption is highly sensitive to how consumption is elicited in survey questions. Hence, this paper highlights the importance of elicitation framing to understand how inflation expectations relate to consumption. In this section, we perform additional checks to test the reliability of our findings.

First, we test robustness to sample variation by re-estimating key hypothesis tests without pooling across related consumption questions. Second, we assess the sensitivity of our results to alternative estimation strategies, including linear, censored and two-part models. Third, we replicate the core empirical designs of prior studies using our unified dataset, illustrating that divergent results in the literature can be reproduced by varying only the framing of the consumption question (i.e., elicitation framing effects). Lastly, we assess the internal and external validity of our consumption measures by examining their pairwise correlations, their loading on common factors and their correlations with aggregate consumption, highlighting how question types capture distinct constructs.

4.1 Sample Variation

In our baseline tests of *temporal* (Hypothesis Test 3) and *question-type framing* effects (Hypothesis Tests 5–6), we pooled responses across related consumption questions (e.g., combining different recall periods or reference units) to increase statistical power (see Table 2 for a summary of the hypothesis tests). While this pooling is reasonable when the effects are homogeneous, it raises the concern that the estimated effects might be driven by particular question variants. To address this, we re-estimate each hypothesis test without pooling, i.e., using each question variant individually. The disaggregated results are reported in Online Appendix Tables B.7–8 (Hypotheses 3a–3b), B.13–B.15 (Hypotheses 5a–5c), and B.16–B.17 (Hypotheses 6a–6b); Online Appendix Tables B.19 and B.20 provide an overview of all hypothesis test results.

The disaggregated estimates closely mirror the pooled ones: the sign and significance of the inflation-consumption relationship remain stable across specifications, confirming that elicitation-framing effects are robust and that aggregation does not drive our findings. Hence, these robustness checks support the pooling approach used in the main analysis and reinforce the validity of our conclusions reported in Section 3.

4.2 Alternative Estimation Strategies

To assess the robustness of our results to estimation strategy choices, we re-estimate baseline specifications (3.1)–(3.3) using alternative methods. Replacing the ordered probit and logit models with OLS yields estimates nearly identical in sign and magnitude (Online Appendix Tables B.21–B.22).

In the baseline analysis of temporal framing effects, we use the *3-month* (F4) and *12-month* (F5) quantitative recall-based questions, excluding respondents with zero spending. These estimates therefore capture only the intensive margin of consumption (spending amount conditional on spending), ignoring the extensive margin of consumption (probability of spending at all). To account for both margins, we re-estimate specification (3.2) for Hypotheses 2 and 3 using Tobit and two-part models. The Tobit model jointly captures the probability and amount of spending, while the two-part model treats these as separate decisions. Both approaches include zero-spending households and test whether this affects our main results. The findings are consistent with the baseline estimates, confirming the robustness of our results to the choice of estimator (Online Appendix Tables B.23–B.26).

4.3 External Validity: Replication Exercise

We conduct a *conceptional replication* of key empirical models from Dräger and Nghiem (2021), Andrade et al. (2023), and Ichiue and Nishiguchi (2015) by applying their original model specification to our data. Leveraging our unified survey design—with seven distinct consumption measures that vary by question type (attitudinal, planned, qualitative and quantitative recall-based), time horizon, and reference unit—we closely replicate the identifying variation used in each study. This approach allows us to investigate whether the conflicting empirical results in the literature can be reproduced within a single dataset. Our main objective is to assess, for each study, whether we can replicate the direction and statistical significance of the key effect of household's inflation expectations on their consumption decisions within our dataset.

By replicating the divergence in prior findings within our harmonized framework, we

demonstrate that the framing of the consumption question is a likely driver of these inconsistencies and thereby offer a potential reconciliation of the literature. Online Appendix C provides full implementation details and presents the replication results. For clarity, we display each study's original key coefficients side-by-side with our estimates based on analogous measures.

Dräger and Nghiem (2021) find that consumers who expect higher inflation tend to report higher current overall spending (a significantly positive effect of expected inflation on consumption) but a lower willingness to buy durable goods. We are able to replicate the direction of their attitude result (higher inflation expectations are associated with more cautious spending attitudes). However, we do not replicate Dräger and Nghiem (2021)'s positive effect on qualitative recall-based spending. In our data, expected inflation has a negligible or even slightly negative association with qualitative recall-based spending. The likely reason is measurement: Dräger and Nghiem (2021) use a broad total-spending measure, whereas our survey focuses only on durable goods.

Andrade et al. (2023) report that households expecting stable prices (the extensive margin) are significantly less likely to consume durable goods than those expecting inflation, and identify the extensive margin as the main channel linking inflation expectations to spending. In our replication, we also find that a higher share of respondents predicting rising price (as opposite to stable prices) is associated with greater quantitative recall-based spending and more favorable spending attitudes, consistent with their extensive-margin result. Similarly, we find a negative effect for the intensive margin, i.e. higher average inflation expectations dampen consumption. While the extensive margin has a stronger effect than the intensive margin in our data as well, the magnitude of this difference is smaller: the gap is about nine times larger in our case, compared to 18 times in Andrade et al. (2023). As a result, our estimate for the extensive margin does not reach statistical significance. A notable difference is that, for attitudes, the intensive margin appears more influential in our data than the extensive margin, whereas Andrade et al. (2023) report the opposite. Nonetheless, we share the finding—both in terms of direction and significance—that inflation expectations matter for households' consumption decisions.

Ichiue and Nishiguchi (2015), using Japanese data, show that higher inflation expectations are associated with a decrease in planned future real spending. We replicate and confirm this finding. Ichiue and Nishiguchi (2015) also investigate the impact of inflation perceptions on qualitative recall-based spending, but we cannot replicate this part of their study due to data-limitations.

Overall, these replication exercises illustrate that question framing is a first-order driver of divergent results. By using a unified dataset with seven distinct consumption questions, we see that simply changing how consumption is measured, can flip the apparent relationship with inflation expectations. This confirms our main conclusion: differences in question framing largely explain the conflicting findings in the literature about inflation expectations and consumption.

4.4 Internal & External Validity: Correlation Exercise of Measures

In Sections 3 and 4.1–4.3, we demonstrated that different framings of consumption questions yield markedly different estimates for the relationship with inflation expectations. This raises a fundamental question: are these measures interchangeable, or do they capture distinct constructs and behavioral dimensions? To address this, we assess both the internal consistency (i.e., correlation across measures, confirmatory factor analysis) and external validity (i.e., alignment with aggregate national consumption data) of the seven survey-based consumption measures used in our experiment. Details on the confirmatory factor analysis and the time series of aggregate durable consumption are provided in Online Appendix D. This includes variable definitions, coding procedures, and national accounts data sources used for benchmarking.

Validation Result 1 (Internal Validity: Survey Measures Are Not Interchangeable). Table 5 displays pairwise correlations among the consumption survey measures. We find strong internal consistency within question-types. For instance, the quantitative recall-based measures (F2, F4, F5) show moderate-to-high correlation, with stronger correlations at similar recall horizons. Similarly, the qualitative recall-based questions (F6 and F7) are highly correlated despite differences in reference unit. By contrast, correlations across question-types are weak. The attitudinal measure (F3) is only weakly correlated with both recall-based and planned spending measures. Its strongest association is observed with planned spending (F1), supporting the view that attitudes reflect forward-looking sentiment—such as willingness to spend—rather than realized consumption.

To validate this distinction more formally, we estimate three one-factor confirmatory factor analyses (CFAs) combining attitude with quantitative and qualitative recall-based measures, respectively. In one specification (Panel B), the attitude loads moderately while recall measures load weakly, but in all other cases attitude either loads weakly or cause non-convergence (Online Appendix Table D.1). This pattern confirms that attitude captures a distinct latent construct, separate from realized spending behavior, and that different

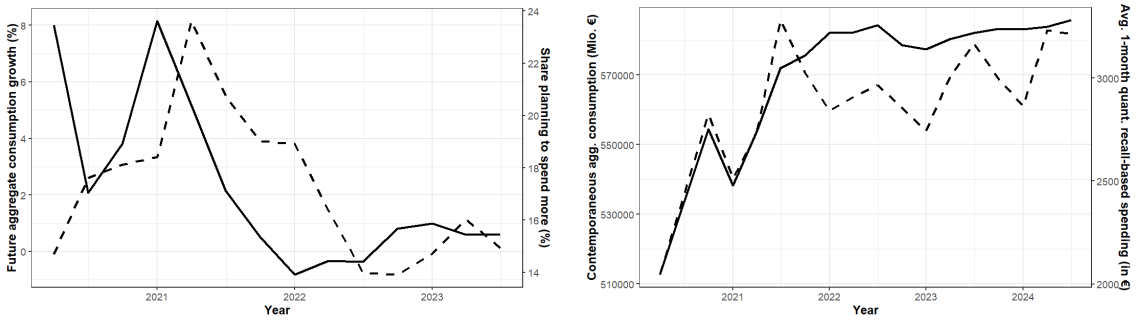
	F1	F2	F3	F4	F5	F6	F7
F1	1						
F2	-0.2184	1					
F3	0.1748	0.0777	1				
F4	-0.1747	0.6421	0.0929	1			
F5	-0.0684	0.3171	0.1177		1		
F6	-0.0848	0.2567	0.1347			1	
F7	-0.1327	0.2866	0.0859			0.6835	1

Notes: This table reports the correlation coefficients between the consumption measures. We estimate the Spearman's Rank Correlation Coefficient if one of the consumption measures is a multinomial ordered variable, otherwise we estimate the Pearson correlation coefficient. F1: planned spending; F2/F4/F5: quantitative recall-based (1, 3, 12 months); F3: attitude; F6/F7: qualitative recall-based (individual, household).

Table 5: Correlation between Consumption Measures

consumption questions reflect distinct behavioral dimensions rather than interchangeable measures.

Validation Result 2 (External Validity: F2 tracks Realized Consumption Best). We next assess external validity by comparing which survey-based consumption measures matches with national account data. For this purpose, we use the time series of planned spending (F1) and past-month quantitative recall-based spending (F2), the only measures available in repeated waves in the BOP-HH survey. To assess external validity of the attitudinal measure (F3), we have to rely on a different dataset, the Michigan Survey of Consumers, which has been widely used in prior work (University of Michigan, Survey Research Center, 2025). Figure 2 plots the corresponding time series comparisons.



Note: Panel A plot the time series of average planned spending (dashed line) and future aggregate consumption growth in Germany (solid line; International Monetary Fund, 2025; Real Final Consumption Expenditure for Germany [NCRSAXDCDEQ], retrieved from FRED)). Panel B plots the time series of average past-month spending (dashed line) and contemporaneous aggregate consumption in Germany (solid line).

Figure 2: When and why do people make major purchases?

Panel A of Figure 2 plots the time series of average planned spending (dashed line)—defined as the share of respondents reporting plans to increase durable spending—alongside future aggregate consumption growth in Germany (International Monetary Fund, 2025;

Real Final Consumption Expenditure for Germany [NCRSAXDCDEQ], retrieved from FRED). Although the two series exhibit broadly similar dynamics, they do not move in lockstep. This is reflected in a moderate correlation of $r = 0.29$. In contrast, Panel B of Figure 2 shows that past-month quantitative recall-based spending (dashed line), averaged across respondents, closely tracks contemporaneous aggregate consumption (solid line), with a strong correlation of $r = 0.84$. This tight co-movement suggests that the survey measure has high external validity.

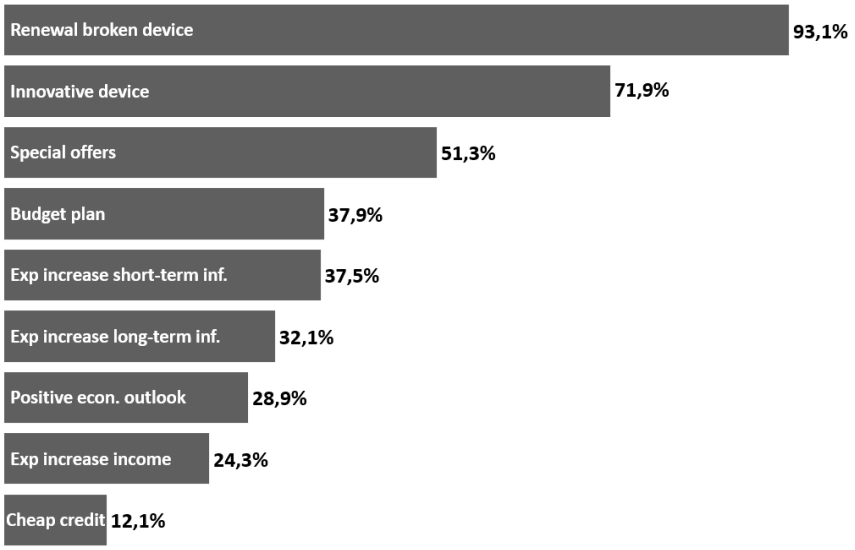
Turning to the U.S. evidence, we compute the attitudinal measure (F3) as the share of respondents who report that it is a good time to purchase durable goods and correlate this indicator with the year-over-year percentage change in real personal consumption expenditures on durable goods (U.S. Bureau of Economic Analysis, 2025; Real personal consumption expenditures: Durable goods (chain-type quantity index) [DDURRA3M086SBEA], retrieved from FRED). The Michigan attitudinal index shows a moderate contemporaneous correlation with U.S. aggregate durable consumption growth ($r = 0.37$), but its forward-looking correlation is near zero ($r = 0.07$). Taken together, these comparisons suggest that the quantitative recall-based measure (F2) is the most externally valid measure for realized consumption.

Validation Result 3 (Attitudinal Measures Reflect Sentiment). These findings reinforce our earlier conclusion in Section 3: attitudinal questions (such as "good time to buy") primarily capture consumer sentiment—i.e., willingness to spend—but do not account for structural constraints such as liquidity or necessity, and therefore only weakly predict actual consumption. Our findings are consistent with early evidence from the 1980s, showing that attitudinal measures are incomplete proxies for realized behavior (Abeele, 1983; Stuart, 1984; Van Raaij and Gianotten, 1990; Williams and Defris, 1981).

To better understand why inflation expectations may fail to consistently predict realized consumption, we asked respondents directly about their motivations for durable purchases (Figure 3). The most frequently cited reason was the need to replace a broken device (93.1%), followed by a desire for innovative products (71.9%) and attractive offers (51.3%). In contrast, only 37.5% cited expected short-term price increases, and just 24.3% mentioned anticipated income changes. These responses highlight that durable spending is primarily driven by structural factors such as necessity and affordability, rather than inflation expectations or general willingness to spend.

In summary, this validation exercise yields three key takeaways. First, the seven consumption survey measures used in the literature are not interchangeable and should be

interpreted in light of their framing and the underlying construct they capture. Second, quantitative recall-based measures (F2), which ask about recent realized spending exhibit the strongest external validity. Third, while attitudinal questions remain very valuable—especially for measuring consumer sentiment—they are weak proxies for realized consumption. Hence, survey framing and measurement validity should be central considerations in both research and policy work of consumption and inflation expectations.



Note: The figure shows the share of consumers who at least somewhat agree with various statements about when and why they buy durable goods, including necessity, expected price increases (short- and long-term), special offers, optimism about the economy, access to cheap credit, replacement needs, adherence to a budget plan, and expected pay raises. See Online Appendix A for the survey questions used.

Figure 3: When and why do people make major purchases?

5 Conclusion

Understanding how inflation expectations shape household consumption is central to both theory and policy. Despite extensive empirical research, no consensus has emerged on how inflation expectations affect consumption decisions. A key reason for this ambiguity lies in how consumption is measured in surveys. We provide a structured overview of the empirical literature on inflation expectations and consumption, synthesizing divergent findings and methodological differences. Using a randomized experiment embedded in a representative household panel survey, we show that the estimated sensitivity of consumption to inflation expectations varies systematically with the elicitation framing of consumption questions. Finally, we evaluate the internal and external validity of our findings and discuss their implications for survey-based consumption research.

The main result is straightforward: elicitation framing matters. Although the overall

relationship between expected inflation and consumption is weak, both its sign and magnitude depend on how consumption is elicited. We find no evidence that elicitation framing by reference unit (individual vs. household spending) affects responses. However, both temporal framing (past one, 3, 12 months) and question-type framing (attitudes, planned spending, quantitative and qualitative recall-based spending) significantly influence the observed relationship. When asked about the general attractiveness of buying durables at present, respondents with higher inflation expectations are less willing to spend. Most notably, the attitudinal question produces results that closely mirror those from planned future consumption, with both showing a negative association with expected inflation. This suggests that attitude and planned spending capture a forward-looking construct (e.g. willingness to spend). Conversely, when asked about recall-based spending, the relationship turns modestly positive or insignificant.These results indicate that question-type framing effects largely reflect construct differences: expected inflation primarily influences willingness to spend, while realized spending is constrained by income, liquidity, and necessity. This interpretation is supported by the internal and external validity tests, which show that recall-based consumption measures align more closely with realized behavior than attitudinal questions. These patterns are robust across specifications and replicate findings from prior studies when their question formats are emulated. Furthermore, we find that respondents tend to buy goods for reasons that are unrelated to inflation expectations, such as replacing a broken device, upgrading an outdated one, or taking advantage of special offers.

The implications are both methodological and behavioral. Question framing is not a minor detail—it determines which which construct of consumption is being measured. Attitudinal and planned spending questions reflect willingness to spend, while recall-based quantitative and qualitative questions reflect realized behavior, often shaped by ability or necessity. Ignoring these distinctions risks conflating sentiment with action or understating the heterogeneity of consumer responses to inflation expectations.

These findings point to several practical recommendations for the design and interpretation of consumption questions. First, our results suggest that consumption should not be treated as a fixed construct. Different question framings can yield systematically different responses. While it may not be feasible for every survey to include multiple versions of the consumption question, it is important to be aware of these framing effects when designing the survey and interpreting results. Second, given the influence of framing on the estimated relationship between inflation expectations and consumption, it is

important for studies to interpret results in light of the specific framing used. This would also facilitate comparison across studies. Third, our results provide practical guidance on the use of different consumption measures. Attitudinal and planned-spending questions reflect willingness to spend and are particularly informative for assessing sentiment and intentions. Recall-based questions more reliably capture realized consumption and are suited when the objective is to approximate actual spending behavior. These measures should be treated as complementary tools rather than substitutes.

Our findings may extend beyond the German population, albeit with some limitations. While the link between inflation expectations and consumption may vary across cultural and institutional contexts—depending on whether intertemporal substitution or precautionary saving dominates—our evidence suggests that framing effects arise from differences in constructs rather than from country-specific behavior. Because these effects reflect survey design rather than local behavioral patterns, they should generalize more broadly, providing tentative support for the external validity of our results.

For policymakers, these findings raise important concerns. If framing alters the estimated strength—or even direction—of the expectations-consumption link, models calibrated on survey data may misstate household responsiveness to monetary policy. As central banks increasingly rely on survey-based indicators to assess household expectations, measurement strategies must account for framing effects.

Appendix A

This section summarizes all empirical papers on the relationship between expected inflation and consumption. This is non-exhaustive list. We list the data, the inflation expectations and consumption measures employed by the paper. We categorize the consumption questions (quantitative and qualitative recall-based, attitudinal as well as reference unit and recall period), and summarize briefly the central findings.

Paper	Data	Measure of Expected Inflation	Question Type	Survey Question	Survey Answers	Question Type	Estimate of Expected Inflation
Bachmann et al. (2015)	Michigan Survey of Consumers; USA, 1984:1 to 2012:12	By about what percent do you expect future prices to go (up/down) on the average, during the next 12 months?	Quantitative; Prices	About the big things people buy for their homes—such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?	(i) good, (ii) bad, (iii) neutral	Attitude	Generally, no significant estimate. At the ZLB, negative and significant estimate.
Ichiue and Nishiguchi (2015)	Opinion Survey; Japan, 1993:1 to 2013:2	What is your outlook for prices 1 year from now?	Qualitative; Prices	How has your household changed its spending compared with 1 year ago?	(i) has increased, (ii) has remained the same, (iii) has decreased	Qualitative; Household level; 12 months	Significant, positive estimate.
Kikuchi and Nakazono (2020)	Online Survey; SCI-personal (Intage); 2015:Q4	Japan, What will the levels of CPI be over the next one-year periods given that the current level of CPI is 10,000?	Quantitative; Inflation			Scanner data; Non-durable goods; Annual and semi-annual	Significant, positive estimate.
Rondinelli and Zizza (2020)	Survey on Household Income and Wealth; Italy, 1991 to 2016	Below you find some intervals for inflation. We would like to know your opinion about inflation in Italy one year head. Distribute 100 points among the following alternatives.	Distributional; Inflation	What is the total value of the objects bought?	Combined spending on furniture, furnishings, household appliances, sundry equipment, PC and other items	Quantitative; Household level; Annual	Significant, positive estimate for total expenditure during high inflation but not when differentiating between durable and non-durable goods.
				How likely is it (from 0 to 100) that your household will buy a new car to replace the present one (the car used most often)?		Probabilistic	Significant, positive estimate.
Dräger and Nghiem (2021)	Consumer Survey on Expectations, Consumption and Saving; Germany, 2015:10 to 2015:12	By what percentage do you think prices in general will increase or decrease on average over the next 12 months?	Quantitative; Prices	How would you say your total expenditures in the past 12 months compare to an average year in the past?	They were (i) considerably higher, (ii) about the same, (iii) considerably lower	Qualitative; Household level; 12 months	Significant, positive estimate.
				When looking at the current economic situation, do you think now is a good or a bad time for people to make large purchases such as furniture or electronic devices and so on?	(i) Now is a good time, (ii) Neither a good nor a bad time, (iii) Now is a bad time	Attitude	Significant, negative estimate.
Duca-Radu et al. (2021)	EU Consumer Survey; EA countries, 2003:5 to 2016:12	By how many per cent do you expect consumer prices to go up/down in the next 12 months?	Quantitative; Prices	In the view of the general economic situation, do you think that now it is the right moment for people to make major purchases such as furniture, electrical/electronic devices, etc.?	(i) yes, it is the right moment now, ii) it is neither the right moment nor the wrong moment, iii) no, it is not the right moment now, iv) don't know	Attitude	Significant, positive estimate.

Note: All qualitative and quantitative consumption measures in this table refer to recall-base spending.

Table Appendix A.1: The Inflation ExpectationsConsumption Puzzle: A Structured Review of the Literature

Paper	Data	Measure of Expected Inflation	Question Type	Survey Question	Survey Answers	Question Type	Estimate of Expected Inflation
Binder and Brunet (2022)	Survey of Consumer Finance; USA, 1951	What do you think will happen to the prices of things you buy during 1951—in general, do you think they will go up, or down, or stay about where they are now?	Qualitative; prices	Is this a good or bad time to buy autos and large household items?	Probably the same as in Bachmann et al. (2015). Answers are not given in the paper.	Attitude	Insignificant estimate. However, consumers who provided a price related explanation for their readiness to spend, the estimate is positive and significant. Significant, negative estimate for the shift of durables and car consumption from 1951 to 1950.
D'Acunto et al. (2022)	GfK; Germany, 2000:1 to 2013:12	How will consumer prices evolve during the next twelve months compared to the previous twelve months?	Qualitative; Prices	Given the current economic situation, do you think its a good time to buy larger items such as furniture, electronic items, etc.?	(i) it's neither a good nor a bad time, (ii) no, it's a bad time, (iii) yes, it's a good time	Attitude	Significant, positive estimate.
Andrade et al. (2023)	INSEE monthly consumer confidence survey; France, 2004:1 to 2018:12	In comparison with the past 12 months, how do you expect consumer prices will develop in the next 12 months?	Quantitative; Prices	Have you made any major purchase over the last 12 months? (such as furniture, household appliances, electronic or computer equipment, etc.) In view of the current general economic situation, do you think now is the right time for people to make major purchases (such as furniture, household appliances, electronic or computer equipment, etc.)?	(i) Yes, (ii) No, (iii) Don't know (i) Yes, now is the right time, (ii) It is neither the right time nor the wrong time, (iii) No, it is the wrong time, (iv) Don't know	Quantitative (but binary); Individual level; 12 months Attitude	Positive, significant effect for the extensive margin (stable prices vs. increasing prices). Insignificant estimate for the intensive margin (quantitative expected inflation given the positive inflation is expected). Positive, significant effect for the extensive margin (stable prices vs. increasing prices). Insignificant estimate for the intensive margin (quantitative expected inflation given the positive inflation is expected).
Burke and Ozdagli (2023)	RAND-American Life Panel; USA, 5:2009 to 2013:1 New York Fed mini module on household expectations; 5:2008 to 11:2012	During the next 12 months, what do you think is the percent chance that prices in general will. The numbers need to add up to 100 percent.	Distributional; Prices	The next questions are about categories of spending that households tend to have less frequently. We would like to know what your household paid - if anything - for any of these items over the last 3 calendar months.	Combined spending on refrigerators, stoves, ovens, washers, dryers, dishwashers, televisions, computers, furniture, floor coverings, and miscellaneous household furnishings and equipment.	Quantitative; Household level; 3 months	No significant estimate, unless one considers heterogeneity across education, income and mortgage balances. The more educated, the ones with lower income and a mortgage have a positive, significant estimate.

Note: All qualitative and quantitative consumption measures in this table refer to recall-base spending.

Table Appendix A.2: The Inflation ExpectationsConsumption Puzzle: A Structured Review of the Literature

Paper	Data	Measure of Expected Inflation	Question Type	Survey Question	Survey Answers	Question Type	Estimate of Expected Inflation
Burke and Ozdagli (2023)	RAND-American Life Panel; USA, 5:2009 to 2013:1 New York Fed mini module on household expectations; 5:2008 to 11:2012	During the next 12 months, what do you think is the percent chance that prices in general will. The numbers need to add up to 100 percent.	Distributional; Prices	Please, provide your best estimate of how much in total your household spent in the following categories. Please include spending by all members of your household, that is, by you or anyone living with you. Even if the amount your household spent last calendar month was unusual, please report that amount.	Clothing, food (home and away), utilities (phone/cable/internet, electricity, water, heating), gasoline, personal care, hobbies and leisure equipment, house cleaning, gardening, health care and medical expenditures (not including drugs), other child spending, entertainment.	Quantitative; Household level; 1 month	No significant estimate.
Coibion et al. (2023)	CentEr Internet Panel; Netherlands, 2018:4 to 2018:7	How much do you think consumer prices in general will change in the next twelve months in the Netherlands?	Quantitative; Prices	How much did your household spend in total on purchases of durable goods in the last three months? What was your households average monthly spending on nondurable goods and services in the last three months?	(i) [...] euros, (ii) I did not buy any durables, (iii) I do not know (i) [...] euros, (ii) I did not buy any durables, (iii) I do not know	Quantitative; Household level; 3 months Quantitative; Household level; 3 months	Significant, negative estimate across all months. Significant, positive estimate in the pooled regression, but out of three monthly only one regression has a positive, significant estimate. The rest is insignificant.
D'Acunto et al. (2023)	Consumer Survey of Statistics Finland; 1996:1 to 2015:3	By what percentage do you think consumer prices will change over the next 12 months?	Quantitative; Prices	In view of the general economic situation in Finland, do you think that now it is the right moment for people to buy durable goods such as furniture, home appliances, cars, etc.?	(i) It is neither the right moment nor the wrong moment, (ii) No, it is not the right moment now, (iii) Yes, it is the right moment now	Attitude	Significant, positive estimated for high IQ individuals. Insignificant estimate for low IQ individuals.
Marenčák (2023)	Harmonized EC consumer survey program; Slovakia, $2009_01to2023:08$	Which development of consumer prices do you expect over the next 12 months? By what percentage do you think consumer prices will change over the next 12 months?	Qualitative; Prices Quantitative; Prices	In view of the general economic situation in Slovakia, do you think that now it is or it is not the right moment for major purchases (furniture, household appliances, electronics, etc.)?	(i) Yes, it's a good time, It's neither a good nor a bad time, No, it's a bad time	Attitude	Positive, significant effect of expecting higher inflation during surge of inflation and normal times. Insignificant effect during deflationary times and drop in inflation. Very small, positive, significant effect across most phases, but not when inflation decreases.

Note: All qualitative and quantitative consumption measures in this table refer to recall-base spending.

Table Appendix A.3: The Inflation ExpectationsConsumption Puzzle: A Structured Review of the Literature

Bibliography

Abeele, P. V. (1983). The index of consumer sentiment: Predictability and predictive power in the eec. *Journal of Economic Psychology*, 3(1):1–17.

Andrade, P., Gautier, E., and Mengus, E. (2023). What matters in households inflation expectations? *Journal of Monetary Economics*, 138:50–68.

Bachmann, R., Berg, T. O., and Sims, E. R. (2015). Inflation expectations and readiness to spend: Cross-sectional evidence. *American Economic Journal: Economic Policy*, 7(1):1–35.

Beckmann, E. and Schmidt, T. (2020). Bundesbank online pilot survey on consumer expectations. Technical report, Technical Paper.

Binder, C. C. and Brunet, G. (2022). Inflation expectations and consumption: Evidence from 1951. *Economic Inquiry*, 60(2):954–974.

Brañas-Garza, P., Ciacci, R., and Ramírez, E. G. R. (2022). Anchors matter: Eliciting maternal expectations on educational outcomes. *Journal of Economic Psychology*, 90:102510.

Burke, M. A. and Ozdagli, A. (2023). Household inflation expectations and consumer spending: evidence from panel data. *Review of Economics and Statistics*, 105(4):948–961.

Chen, J. I., Han, Y., and He, T.-S. (2025). Does future time reference framing increase temporal discounting? evidence from english monolinguals. *Journal of Economic Psychology*, 107:102796.

Coibion, O., Georgarakos, D., Gorodnichenko, Y., and Van Rooij, M. (2023). How does consumption respond to news about inflation? field evidence from a randomized control trial. *American Economic Journal: Macroeconomics*, 15(3):109–152.

D'Acunto, F., Hoang, D., Paloviita, M., and Weber, M. (2023). Iq, expectations, and choice. *The Review of Economic Studies*, 90(5):2292–2325.

D'Acunto, F., Hoang, D., and Weber, M. (2022). Managing households expectations with unconventional policies. *The Review of Financial Studies*, 35(4):1597–1642.

De Bruin, W. B., Van der Klaauw, W., Topa, G., Downs, J. S., Fischhoff, B., and Armantier, O. (2012). The effect of question wording on consumers reported inflation expectations. *Journal of Economic Psychology*, 33(4):749–757.

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

Deutsche Bundesbank (2025). Bundesbank online panel – households (bop-hh). Research Data and Service Centre. Retrieved from `https://www.bundesbank.de/en/bundesbank/research/rdsc/your-research-project-at-the-rdsc`. Last accessed on May 30, 2025.

Dräger, L. and Nghiem, G. (2021). Are consumers' spending decisions in line with a euler equation? *Review of Economics and Statistics*, 103(3):580–596.

Duca-Radu, I., Kenny, G., and Reuter, A. (2021). Inflation expectations, consumption and the lower bound: Micro evidence from a large multi-country survey. *Journal of Monetary Economics*, 118:120–134.

Galí, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press, Princeton, NJ.

Goldin, J. and Reck, D. (2020). Revealed-preference analysis with framing effects. *Journal of Political Economy*, 128(7):2759–2795.

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Ichiue, H. and Nishiguchi, S. (2015). Inflation expectations and consumer spending at the zero bound: Micro evidence. *Economic Inquiry*, 53(2):1086–1107.

International Monetary Fund (2025). Real final consumption expenditure for germany [ncrsaxdcdeq]. Federal Reserve Bank of St. Louis (FRED). Retrieved from `https://fred.stlouisfed.org/series/NCRSAXDCDEQ`. Last accessed on May 30, 2025.

Kalton, G. and Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 145(1):42–57.

Katona, G. (1975). *Psychological economics.* Elsevier.

Kikuchi, J. and Nakazono, Y. (2020). *Inflation expectations and consumer spending: micro-data evidence*. Economic and Social Research Institute, Cabinet Office.

Marenčák, M. (2023). State-dependent inflation expectations and consumption choices. *Available at SSRN 4660603*.

of Education, F. M. and Research (2023). The german qualifications framework for lifelong learning.

Pavlova, L. (2025). Framing effects in consumer expectations surveys. *Journal of Economic Behavior & Organization*, 231.

Reitmann, A.-K., Goedhuys, M., Grimm, M., and Nillesen, E. E. (2020). Gender attitudes in the arab region–the role of framing and priming effects. *Journal of economic psychology*, 80:102288.

Rondinelli, C. and Zizza, R. (2020). Spend today or spend tomorrow? the role of inflation expectations in consumer behaviour. *The Role of Inflation Expectations in Consumer Behaviour (April 27, 2020). Bank of Italy Temi di Discussione (Working Paper) No*, 1276.

Saris, W. E. and Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass communication & society*, 3(2-3):297–316.

Schuman, H. and Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15:205–234.

Stuart, O. (1984). A note on the measurement, quantification, and use of consumer confidence. *Journal of Economic Psychology*, 5(2):125–138.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). The psychology of survey response.

University of Michigan, Survey Research Center (2025). Survey of consumers. Retrieved from https://data.sca.isr.umich.edu/. Last accessed on May 30, 2025.

U.S. Bureau of Economic Analysis (2025). Real personal consumption expenditures: Durable goods (chain-type quantity index) [ddurra3m086sbea]. Federal Reserve Bank of St. Louis (FRED). Retrieved from https://fred.stlouisfed.org/series/DDURRA3M086SBEA. Last accessed on May 30, 2025.

Van Raaij, W. F. and Gianotten, H. J. (1990). Consumer confidence, expenditure, saving, and credit. *Journal of Economic Psychology*, 11(2):269–290.

Williams, R. A. and Defris, L. (1981). The roles of inflation and consumer sentiment in explaining australian consumption and savings patterns. *Journal of Economic Psychology*, 1(2):105–120.