Discussion Paper Series – CRC TR 224

Discussion Paper No. 791
Project C 01

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

Tiziana Assenza[1]
Anna Mogilevskaja[2]
Stefanie Huber[3]
Tobias Schmidt[4]

November 2026

[1]TSE (UT-Capitole) and IAST
[2]University of Bonn and ECONtribute Niebuhrstr. 5, 53113 Bonn, Germany. Email: amogilev@uni-bonn.de
[3]University of Bonn and ECONtribute Niebuhrstr. 5, 53113 Bonn, Germany. Email: shuber@uni-bonn.de
[4]Deutsche Bundesbank, Research Center

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

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 an experimental module in the Bundesbank Online Panel Households to show that the estimated link between inflation expectations and household consumption changes 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 questions 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 mainly question type significantly influences the estimated relationship between inflation expectations and durable consumption. While the average association 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 or insignificant link. 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: Inflation expectations; household consumption; survey design; question wording; framing effects; measurement.

JEL Classification: C83, D12, D84, E21, E31.

*We thank participants at the 2026 meetings of the European Economic Association, the Royal Economic Society, the CESifo Area Conference on Macro with Micro Data, the CEPR Workshop on Frontiers in Measurement and Survey Methods, the Verein für Socialpolitik, and the Experimental Finance Conference for valuable comments and suggestions. Huber and Mogilevskaja acknowledge funding from the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/2-390838866). Support from the German Research Foundation (DFG) through CRC TR 224 (Project C01) is gratefully acknowledged. Assenza acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements dAvenir) program (grant ANR-17-EURE-0010) and under project ANR-23-CE26-0011-01. The views expressed are those of the authors and do not necessarily reflect those of the Deutsche Bundesbank or the Eurosystem.

†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, 2015), 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 a negative association. 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 relationship between consumption and 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 common formulation in the 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). Each respondent answers several of these questions and is routed to different subsets, in part by random assignment. This yields within-person comparisons of two measures answered by the same respondent, and between-person comparisons across respondents who received different measures. By holding inflation expectations fixed within respondents while varying the consumption question, our design allows us to compare the measured relationship between

inflation expectations and consumption across elicitation framings.

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) and time horizon have no detectable impact, the question type significantly affects the sign and magnitude of the estimated relationship. Quantitative recall-based measures of past spending yield a positive but weak association with expected inflation. Forward-looking spending plans, by contrast, exhibit a negative association with expected inflation, consistent with higher expected inflation shifting spending toward the present rather than the future. Strikingly, the attitudinal question of whether it is now a good time to buy also yields a negative association with expected inflation, closely mirroring forward-looking spending plans rather than measures of realized past consumption. This is particularly noteworthy because the attitudinal question is widely used in the literature as a proxy for current consumption behavior.

We find that this divergence across question types reflects underlying construct differences: attitudinal and planned spending items appear to capture willingness to spend, while recall-based items capture realized consumption shaped additionally by constraints such as income, liquidity, or necessity. Consistent with this interpretation, our measurement comparisons suggest that attitudinal and recall-based measures capture distinct constructs. The confirmatory factor analysis (CFA) provides suggestive evidence that the attitudinal measure does not load strongly onto the same latent factor as realized spending. In addition, our aggregate-consumption benchmarking exercise indicates that quantitative recall-based measures track aggregate consumption more closely than the attitudinal measure. Overall, elicitation framing affects estimated associations, with point estimates changing sign across some formulations. This helps reconcile conflicting results in the literature and provides practical guidance for survey design and policy interpretation.

Our findings contribute to three strands of literature: survey framing, the inflation-expectations-consumption relationship, and consumer sentiment measurement.

First, we contribute to the survey design and framing literature by implementing a unified experimental design that varies several dimensions of consumption question framing within a single survey wave, allowing us to compare how the estimated relationship differs depending on the elicitation framings. We formalize a simple typology of framing dimensions and show that the estimated relationship is sensitive to some dimensions of question framing, but not others. 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 how they interpret 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. Relatedly, Haan et al. (2025) show that alternative formats for eliciting long-run stock-market expectations affect both the reported expectations and their predictive relationship with investment choices. 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. Importantly, we show that differences in question formulation can affect not only measured responses themselves, but also the estimated relationship between consumption and another economic variable—inflation expectations.

Second, we contribute to the empirical literature on inflation expectations and household consumption by showing that part of the conflicting evidence across prior studies can be traced to differences in how consumption is elicited. We develop a conceptual framework and typology that classifies 13 empirical studies by country, data source, consumption question framing, and results, revealing systematic variation in estimated relationships across different types of consumption measures. Our unified design then allows us to reproduce seemingly contradictory findings from the literature within a single dataset by varying

how consumption is elicited. This provides a coherent explanation for part of the existing empirical heterogeneity: differences in estimated relationships may reflect differences in the constructs captured by alternative consumption measures rather than differences in underlying household behavior. 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 illustrates this heterogeneity. 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). Taken together with our replications of three prior studies within the same dataset, these patterns suggest that differences in consumption elicitation can account for part of the heterogeneity in the estimated relationship between inflation expectations and consumption.

Third, we contribute to the literature on consumer sentiment and measurement validity of survey-based consumption indicators (e.g., Katona, 1975; Williams and Defris, 1981; Abeele, 1983) by investigating what alternative consumption measures capture and how closely they reflect realized consumption 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 asks whether respondents consider *now* a good time to make major purchases and has since become one of the most widely used proxies for current 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 further allows us to distinguish willingness from *ability* and *necessity* (the perceived urgency or indispensability of a purchase) as distinct drivers of

consumption. This richer typology allows us to assess whether inflation expectations are related to willingness to spend, realized spending, or both, and under what conditions. Our results show that while higher inflation expectations are often significantly associated with a lower willingness to spend in the future, this association rarely extends to realized spending, which is additionally shaped by financial constraints and replacement needs. This interpretation is reinforced by our own survey data: when asked why they make durable purchases, only a small share of respondents mention inflation expectations, while the vast majority point 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 helps explain 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 it behaves more like a forward-looking indicator of willingness to spend than a measure of realized current consumption. This distinction has important implications for economic modeling and policy design.

Central banks and researchers seeking to assess empirically the expectations channel of monetary policy rely on household surveys to measure inflation expectations and consumption behavior. Consistent with intertemporal substitution, we find that higher inflation expectations are significantly associated with lower planned future spending, a pattern closely mirrored by attitudinal consumption measures. By contrast, recall-based measures of realized spending—which our external validation shows most closely track aggregate consumption—show weak positive associations with inflation expectations that are statistically insignificant. Thus, while households willingness to spend aligns with theoretical predictions, we find little evidence that this translates into realized spending, potentially because actual consumption is also shaped by households ability and need to spend. These findings highlight the need to distinguish carefully between different consumption measures when drawing conclusions about the expectations channel from household survey data.

The remainder of the paper is organized as follows. Section 2 describes the data and the experimental design. Section 3 presents the hypothesis tests and empirical strategy. Section 4 reports the results. Section 5 discusses robustness checks, validation exercises, and replications of prior studies. Section 6 concludes. Appendix A provides a structured overview of prior studies and a comprehensive reference on this contested relationship.

2 Survey Design and Framing Variation

2.1 Data

To study how elicitation framing affects the relationship between inflation expectations and consumption, 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 module into an otherwise representative household survey in Germany. Respondents for the BOP-HH are randomly drawn from an offline-recruited 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 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.

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. The elicitation framing 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).

The resulting dataset comprises approximately 3,900 respondents, reflecting the Bundesbank's typical monthly sample size of around 4,000 participants. A sample size of about 4,000 individuals 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.

Code	Question Type	Survey Question 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 multi-person household.

Table 1: Overview of Consumption Survey Questions and Framing Dimensions

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.

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).

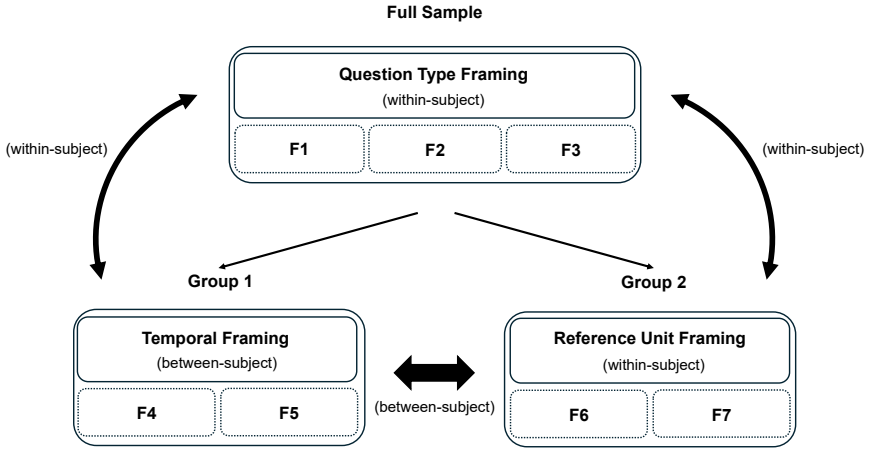
2.2 Details on Framing Variation

We now describe the structure of the survey module and the specific consumption questions administered to each group. Figure 1 provides an overview of the design.

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.

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 spending questions.



Notes: This diagram illustrates the framing experiment embedded in the BOP-HH. All respondents answer three baseline questions: planned spending (F1), 1-month quantitative recall-based spending (F2), and attitude (F3). The full sample is then randomly split into Group 1, which answers quantitative recall-based questions over 3 months (F4) or 12 months (F5), and Group 2, which answers qualitative recall-based questions at the individual (F6) and household (F7) levels. Curved arrows indicate within-subject comparisons; the horizontal arrow indicates the between-subject comparison of F5 and F7.

Figure 1: Survey Design with Framing Variation

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). Respondents living alone (single-person households) received only the individual recall-based spending question (F6), since for them, personal and household spending are equivalent.

The layered structure of the experiment yields both within-subject and between-subject comparisons across the three framing dimensions. In the following section, we discuss each elicitation framing dimension and its corresponding hypothesis test, identification strategy and empirical test.

3 Hypothesis Tests and Identification Strategy

We examine whether the estimated relationship between inflation expectations and consumption varies with the framing of the consumption question along three 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). Table 2 provides an overview of the seven hypothesis tests. For each test, it reports the framing dimension and consumption measures being compared, whether the comparison is within- or between-subject, and whether the order of the two consumption questions being compared is randomized. We discuss each framing dimension and the corresponding hypothesis tests in turn below.

Framing Dimension	Question Type	Hypothesis Test	Design	Question Order
Reference Unit	Qualitative recall-based F6/F7	$H1$: F6 vs. F7	Within-Subject	Randomized
Temporal	Quantitative recall-based F4/F5	$H2$: F4 vs. F5	Between-Subject	
Temporal	Quantitative recall-based F2/F4/F5	$H3$: F2 vs. F4/F5	Within-Subject	Not randomized
Question type	Quantitative vs. Qualitative recall-based	$H4$: F4/F5 vs. F7	Between-Subject	
Question type	Attitude vs. Quantitative recall-based	$H5$: F3 vs. F4/F5	Within-Subject	Not randomized
Question type	Attitude vs. Qualitative recall-based	$H6$: F3 vs. F6/F7	Within-Subject	Not randomized
Question type	Attitude vs. Planned spending	$H7$: F3 vs. F1	Within-Subject	Not randomized

Notes: We test elicitation framing dependencies along three dimensions: *reference unit*, *temporal horizon*, and *question type*. The *Design* column indicates whether the hypothesis is tested within- or between-subject. The *Question Order* column indicates whether the question order is randomized across respondents. Hypothesis 1 tests reference unit framing by comparing qualitative recall-based individual (F6) and household (F7) spending; questions are presented in randomized order. Hypothesis 2 and 3 examine temporal framing. Hypothesis 2 compares 3-month (F4) and 12-month (F5) quantitative recall-based spending in a between-subject comparison. Hypothesis 3 compares 1-month (F2) with pooled responses from 3-month (F4) and 12-month (F5) quantitative recall-based spending in a within-subject comparison with a fixed question order. Hypothesis 4–7 evaluate question-type framing. Hypothesis 4 uses a between-subject comparison to compare 12-month household-level quantitative (F4/F5) and qualitative (F7) recall-based spending among multi-person households. Hypothesis 5–7 contrast attitude (F3) with quantitative (F4/F5) and qualitative (F6/F7) recall-based spending, and planned spending (F1), respectively, all in a within-subject comparison with a fixed question order.

Table 2: Hypothesis Tests on Framing Variation

3.1 Reference Unit Framing

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 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 the relationship between inflation expectations and consumption.

Hypothesis Test 1 tests 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. Respondents living alone received only F6 and are therefore not part of this comparison. The resulting estimation sample is the multi-person respondents in Group 2 who answered both questions, each contributing two observations ($N = 2,584$). This makes the estimation within-subject.

Within-subject identification requires three assumptions to hold: (i) a balanced panel, (ii) temporal stability, and (iii) causal transience (List, 2025). In this study, the balanced panel assumption holds by construction for any within-subject estimation, as no participants dropped out between questions. Temporal stability is also very plausible for any within-subject estimation given the minimal time elapsed between questions within the same survey wave. Causal transience is the most demanding assumption, requiring that exposure to one treatment condition does not persist into the next—in our case, that responding to one consumption item does not affect responses to subsequent ones. In the case of Hypothesis Test 1, the order of questions F6 and F7 was randomized across respondents. This prevents question order from being systematically confounded with reference unit framing and mitigates concerns about carry-over and order effects.

We estimate the following specification:

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \qquad (3.1)$$

where C_{im} is the durable consumption response of individual i to consumption measure m, and m indexes the two measures compared in a given hypothesis test, here the individual-level (F6) and household-level (F7) qualitative recall-based questions. $E_i(\pi_{t+12})$ denotes individual i's twelve-month-ahead inflation expectation, where t is the survey month; since all responses are collected in the March 2024 wave, there is no time variation to index. The indicator $Treat_{im}$ equals one if the observation comes from the household-level question (F7) and zero if it comes from the individual-level question (F6). The coefficient of interest is β_2: it captures whether the association between inflation expectations and reported consumption differs across the two reference units. The vector X_i contains macroeconomic expectations and socio-demographic characteristics known to influence consumption: age, age squared, gender, household income, education, employment status, an East Germany dummy, household size, and city size. The error term is denoted ϵ_{im}.

In Hypothesis Test 1, Equation (3.1) is estimated using an ordered probit model, as F6 and F7 have three-category ordered response scales and require no recoding. The model includes individual random effects, $\epsilon_{im} = \alpha_i + u_{im}$, to account for the dependence between repeated responses from the same respondent; α_i is assumed to be independent of the explanatory variables. As in all within-subject comparisons below, standard errors are clustered at the individual level.

For the remaining hypothesis tests, Equation (3.1) provides the baseline specification, with β_2 capturing whether the association between inflation expectations and consumption differs between the two question formulations being compared. What differs across hypothesis tests is the pair of consumption measures compared, the estimation sample, the coding of the dependent variable, the estimator, and the error structure. In between-subject comparisons, each respondent contributes a single observation, so $Treat_{im}$ reduces to $Treat_i$.

3.2 Temporal Framing

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 relates to expected inflation.

Attitudinal consumption measures lack an explicit time frame. 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. To test whether the relation between inflation expectations and consumption depends on the temporal horizon of the consumption question, we construct two hypothesis tests.

Hypothesis Test 2 compares quantitative recall-based spending over the past 3 months (F4) with spending over the past 12 months (F5). Each respondent in Group 1 answered only one of the two questions, depending on their previously reported purchasing behavior. Respondents who reported no purchases in either horizon are not part of this comparison. The estimation sample therefore contains only individuals with strictly positive durable expenditure over the past 12 months ($N = 840$), each contributing a single observation, and the comparison is between-subject. In Equation (3.1), m indexes F4 and F5, and $Treat_i$ equals one if the response refers to the shorter horizon (F4) and zero if it refers to the longer horizon (F5).

Assignment to F4 and F5 is not random but depends on respondents reported purchasing behavior, such that the two groups differ in their purchase histories by construction. Balance tests reported in Online Appendix Table B.3 show no systematic differences in observable characteristics between the two groups, but cannot rule out selection on unobservables. We therefore interpret β_2 as indicating whether the relationship between inflation expectations and reported consumption differs between the F4 and F5 groups, rather than as identifying the causal effect of the recall horizon.

As the estimation sample includes only respondents who reported strictly positive spending, we transform F4 and F5 into log spending to account for the strong right-skewness of the intensive margin, thereby approximating normality and stabilizing variance. Equation (3.1) is estimated as a log-linear model with robust (Huber-White) standard errors. No random effects are included, since each respondent contributes a single observation.

Hypothesis Test 3 contrasts quantitative recall-based spending over the past month (F2) with the pooled responses over the past 3 and 12 months (F4/F5). Because F4 and F5 were shown only to respondents who had indicated a purchase in the relevant period, the comparison is restricted to respondents who reported strictly positive spending both for F2 and for F4 or F5. Each of these respondents therefore contributes two observations ($N = 886$), and the comparison is within-subject. In Equation (3.1), m indexes F2 and the pooled F4/F5 responses, and $Treat_{im}$ equals one if the observation comes from the 1-month question (F2) and zero if it comes from the 3- or 12-month question.

Because Hypothesis Test 3 compares different recall horizons within the same respondent, differences in respondent composition do not confound the comparison. Although, question order was fixed, at least five thematically unrelated items separate F2 from F4/F5, creating cognitive distance between the two spending questions. This separation makes direct carry-over effects less likely and therefore supports the causal-transience assumption, although order effects cannot be ruled out entirely.

As the estimation sample includes only respondents who reported strictly positive spending, we transform all responses to F2, F4 and F5 into log spending, for the same reason as above. Equation (3.1) is estimated as a log-linear model with individual random effects and standard errors clustered at the individual level.

3.3 Question Type Framing

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 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, (ii) planned spending, (iii) qualitative recall-based questions, and (iv) quantitative recall-based questions. We investigate whether the estimated relationship between inflation expectations and consumption depends on the question type through four pairwise comparisons: qualitative versus quantitative recall-based spending (Hypothesis Test 4), attitude versus quantitative recall-based spending (Hypothesis Test 5), attitude versus qualitative recall-based spending (Hypothesis Test 6), and attitude versus planned spending (Hypothesis Test 7).

Hypothesis Test 4 compares quantitative recall-based spending (F4/F5) with household-level qualitative recall-based spending (F7). The quantitative questions were administered to Group 1 and F7 to Group 2, and assignment to the two groups is random. Because F7 was administered only to respondents living in multi-person households, we apply the same household-size restriction to Group 1, so that both arms refer to the same target

population. Conditional on this common restriction, whether a respondent receives the quantitative or qualitative recall-based measure is determined by the randomized assignment to Group 1 versus Group 2. Online Appendix Table B.4 confirms that the resulting subsamples are balanced on observable characteristics.

Because the two response formats are not directly comparable, we compare them at the extensive margin: the quantitative measure enters as an indicator of whether the respondent reported positive spending over the past 12 months. This includes also respondents who reported no purchase in either period. All respondents of Group 1 living in a multi-person household therefore contribute one observation (N = 2,626), and the comparison is between-subject. In Equation (3.1), m indexes the quantitative and the qualitative recall-based measure, and $Treat_i$ equals one if the observation comes from the qualitative question (F7).

Because the quantitative measure records only whether spending was positive and can be constructed for every Group 1 respondent from the screening questions, the conditional routing into F4 or F5 does not generate selection into the quantitative arm of Hypothesis Test 4. Respondents reporting a purchase in either period are coded as one, while respondents reporting no purchase are coded as zero, irrespective of whether the subsequent routing led them to F4, F5, or neither.

The qualitative recall-based measure also requires harmonization: F7 is a three-category ordered item, which we code as a binary indicator for increased spending. The comparison therefore contrasts a recall-based measure of spending incidence with a recall-based measure of spending change, both elicited over the past 12 months. Equation (3.1) is estimated as a logit model and standard errors are robust (Huber-White).

Hypothesis Tests 5 contrasts attitude (F3) with the pooled quantitative recall-based responses over the past 3 and 12 months (F4/F5). F3 was administered to the full sample and the quantitative questions to Group 1. As in Hypothesis Test 4, the two formats are compared at the extensive margin. Because the quantitative measure records only whether spending was positive, the conditional routing into F4 or F5 is irrelevant to this comparison and no Group 1 respondent is excluded: every one of them contributes both an attitude and a quantitative recall-based observation (N = 3,600), and the comparison is within-subject. In Equation (3.1), m indexes F3 and the quantitative recall-based measure, and $Treat_{im}$ equals one if the observation comes from the latter.

Although the question order is fixed, causal transience is plausible in this comparison. F4 and F5 ask respondents to report concrete past purchases that necessarily occurred

before the attitude was elicited, rather than to provide a second evaluative judgment. This limits the scope for the earlier attitude question to induce consistency in the subsequent recall response. While order or carry-over effects cannot be ruled out entirely, these features make it unlikely that they drive our results.

F3 has a three-category ordered response scale, which we transform into an indicator for a positive or neutral spending attitude. The comparison therefore contrasts an evaluative assessment of current buying conditions with a recall-based measure of past spending incidence. Equation (3.1) is estimated as a logit model with individual random effects and standard errors clustered at the individual level.

Hypothesis Test 6 compares attitude (F3) with the pooled qualitative recall-based responses (F6/F7). F6 was administered to all of Group 2 and F7 only to respondents in multi-person households. Where a respondent answered both, we use the household-level response (F7), so that every Group 2 respondent contributes exactly one qualitative observation alongside their attitudinal response ($N = 3,580$). The comparison is within-subject. In Equation (3.1), m indexes F3 and the pooled F6/F7 responses, and $Treat_{im}$ equals one if the observation comes from a qualitative recall-based question. As a robustness check, we restrict the qualitative measure to individual-level responses (F6), which yields similar results.

Although question order was fixed, causal transience is plausible in this comparison. F6 and F7 concern past spending behavior that necessarily occurred before the attitude was elicited and ask respondents to recall changes in spending rather than to provide a second evaluation of current buying conditions. While order or carry-over effects cannot be ruled out entirely, these features make it unlikely that they drive our results.

F3, F6 and F7 all have three-category ordered response scales, allowing the responses to enter in their original coding without harmonization. Equation (3.1) is estimated as an ordered probit model with individual random effects and standard errors clustered at the individual level.

Hypothesis Test 7 compares attitude (F3) with planned spending (F1). Both items were administered to the full sample, so every respondent contributes two observations ($N = 7,196$) and the comparison is within-subject. In Equation (3.1), m indexes F3 and F1, and $Treat_{im}$ equals one if the observation comes from the planned spending question.

Although question order was fixed, with planned spending (F1) preceding the attitudinal question (F3), four thematically unrelated survey items separate the two questions. This creates cognitive distance between the two measures, making direct carry-over effects

unlikely and supporting the causal-transience assumption.

Attitude and planned spending both have three-category ordered response scales, so no harmonization is required. Equation (3.1) is estimated as an ordered probit model with individual random effects and standard errors clustered at the individual level.

Having outlined the elicitation framing dimensions and their corresponding hypothesis tests and identification strategies, we now move on to the results.

4 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. Each dimension captures a distinct aspect of question design that may shape the estimated relationship between consumption and expected inflation. Additional details on the estimation strategies and hypothesis tests are provided in Online Appendices B.2 and B.3.

Most of our hypothesis tests are estimated with non-linear models, so coefficients cannot be interpreted directly. We therefore report how the predicted outcome changes when inflation expectations increase from 2.2 to 4.5 percent. These two values equal the first and third quartiles of inflation expectations in the full sample and approximately match those in each estimation subsample. We estimate predicted outcome changes separately for each of the two questions compared, together with the difference between the two changes, which is the quantity that tests the hypothesis. In the non-linear models, the outcome is the predicted probability of each response category and changes are in percentage points (pp). In the log-linear models of Hypothesis Tests 2 and 3, the outcome is predicted log spending, and changes can be read as approximate percentage changes in spending. Because these are computed on the log scale, differences in the level of spending between two questions, such as those arising from different recall horizons, drop out, and the difference between the two changes is equivalent to the test of the interaction term in Equation (3.1). Predicted outcomes under the control and treatment questions at the first quartile, the median and the third quartile of inflation expectations (2.2, 3 and 4.5 percent) are reported in Online Appendix B.4, Figures B.1–B.7. As an alternative summary of the estimated relationships, we also report average marginal effects in Online Appendix B.4., Tables B.5–B.6. The conclusions are unchanged.

4.1 Reference Unit Framing

Table 3, Panel A, presents the results for *Hypothesis Test 1*, comparing individual- (F6) and household-level (F7) qualitative recall-based spending. The table reports how the predicted probability of each response category changes when inflation expectations increase from 2.2 to 4.5 percent. Neither the individual-level nor the household-level qualitative recall-based measure shows a statistically significant association with expected inflation in any response category. The difference between the two changes, which tests the hypothesis directly, is statistically insignificant for every response category (Column 3). We therefore find no evidence that the association between inflation expectations and reported durable spending depends on whether spending is reported at the individual or the household level. One possible explanation is that durable goods are typically shared within households, leading to similar answers regardless of reference unit.

As a robustness check, 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 unchanged (Online Appendix Table B.7).

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 relationship between inflation expectations and reported consumption.

4.2 Temporal Horizon Framing

Table 3, first block of Panel B, presents the results for *Hypothesis Test 2*, contrasting 3-month (F4) and 12-month (F5) quantitative recall-based spending. The table reports the approximate percentage change in reported spending when inflation expectations increase from 2.2 to 4.5 percent. For F5 we find no significant association. For F4, an increase in expected inflation from 2.2 to 4.5 percent is associated with approximately 14% higher reported spending ($p = 0.022$). The difference between the two estimates, about 18pp, is significant ($p = 0.017$). Because respondents were routed to F4 or F5 based on their reported purchase incidence, we interpret this as suggestive evidence that shorter recall horizons potentially strengthen the association.

The second block of Panel B presents the results for *Hypothesis Test 3*, which compares 1-month quantitative recall-based spending (F2) with the pooled reports of 3-month and

12-month recall-based spending (F4/F5). Neither association is statistically significant, nor is the difference between them. We therefore find no evidence that the estimated association between reported consumption and inflation expectations differs across recall horizons when the comparison holds the respondent fixed.

Outcome	Control question (1)	Treatment question (2)	Difference (3)
Panel A: Reference unit framing (percentage points)			
$H1$: F6 (control) vs. F7 (treatment), RE ordered probit, $N = 2{,}584$			
Pr(Decreased spending)	0.07	−0.87	−0.94
	(0.63)	(0.56)	(0.55)
Pr(Same spending)	0.01	−0.23	−0.24
	(0.10)	(0.15)	(0.13)
Pr(Increased spending)	−0.08	1.10	1.18
	(0.72)	(0.71)	(0.66)
Panel B: Temporal framing (percent)			
$H2$: F5 (control) vs. F4 (treatment), OLS, $N = 840$			
Log spending	−3.81	14.05*	17.86*
	(4.31)	(6.12)	(7.46)
$H3$: F4/F5 (control) vs. F2 (treatment), RE GLS, $N = 886$			
Log spending	8.23	5.90	−2.32
	(4.55)	(5.56)	(5.46)

Notes: This table reports how the average predicted outcome changes when inflation expectations increase from 2.2 to 4.5 percent under the control and the treatment question for Hypothesis Tests 1–3, based on regression specification (3.1). Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the framing used. Panel A examines whether the relationship depends on the reference unit (individual vs. household), based on responses to individual- (F6) and household-level (F7) qualitative recall-based spending using within-subject variation, estimated with a random effects ordered probit model. Entries in Panel A are changes in the predicted probability of each response category, in pp. Panel B assesses whether the time horizon of the consumption question affects the relationship. Hypothesis Test 2 compares the log of quantitative recall-based spending over the past 3 months (F4) and the past 12 months (F5) using a between-subjects design, estimated by OLS. Hypothesis Test 3 contrasts the log of quantitative recall-based spending over the past month (F2) with the pooled responses from F4 and F5 using within-subject variation, estimated with a random effects GLS model. Entries in Panel B are changes in predicted log spending multiplied by 100 and can be read as the approximate percentage change in spending. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in random effects models. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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

Re-estimating the specification separately for F2 vs. F4 and F2 vs. F5 yields similar results (Online Appendix Tables B.8), indicating that pooling F4 and F5 does not drive the findings.

Taken together, the two tests provide no robust evidence that temporal framing affects the estimated relationship between inflation expectations and reported consumption.

While Hypothesis Test 2 points to a stronger association at the shorter recall horizon, this comparison is between respondents routed to different questions based on their reported purchase incidence. The within-subject comparison in Hypothesis Test 3, which holds respondent composition fixed, shows no significant difference across recall horizons.

Result 2 (No robust evidence that temporal framing matters). When the same respondents report spending over different recall horizons, the association between inflation expectations and recall-based spending does not differ across horizons. Only the between-subject comparison shows a stronger association for the 3-month than for the 12-month horizon, which may, however, reflect differences between respondents routed to each horizon rather than the horizon itself.

4.3 Question Type Framing

Panel 1 of Table 4 presents the results for *Hypothesis Test 4*, which compares quantitative (F4/F5) and qualitative (F7) recall-based spending at the extensive margin, reporting changes in predicted probabilities in percentage points. For the quantitative measure, the association between expected inflation and the probability of reporting positive spending is insignificant. For the qualitative measure, an increase in expected inflation from 2.2 to 4.5 percent is associated with a 1.97pp higher probability of reporting increased spending ($p = 0.006$). The difference between the two changes is not statistically significant ($p = 0.093$), providing no evidence that the two recall-based question types differ in their relation to expected inflation.

As a robustness check, we include single-person households, using their individual-level response (F6) instead of F7. The quantitative measure (F4/F5) now includes all Group 1 respondents. The extended sample is balanced, and the difference between the two measures remains insignificant (Online Appendix Tables B.9–B.10).

Panel 2 presents the results for *Hypothesis Test 5*, which compares attitudinal responses (F3) with quantitative recall-based spending (F4/F5) at the extensive margin. An increase in inflation expectations from 2.2 to 4.5 percent is associated with a 2.88pp lower probability of reporting that it is a good time, or neither a good nor a bad time, to buy durable goods ($p < 0.001$). For the quantitative measure, the association with the probability of reporting positive spending is insignificant. The difference between the two changes is around 3.4pp and significant ($p < 0.001$), indicating that the attitudinal consumption measure is more strongly associated with inflation expectations. A plausible explanation is

that attitudinal questions capture willingness to spend more directly, whereas recall-based reports reflect realized consumption, which is also shaped by ability, liquidity constraints, and necessity.

Outcome	Control question (1)	Treatment question (2)	Difference (3)
$H4$: F4/F5 (control) vs. F7 (treatment), logit, $N = 2{,}626$			
Pr(Positive spending/Increased spending)	0.10	1.97**	1.87
	(0.87)	(0.72)	(1.12)
$H5$: F3 (control) vs. F4/F5 (treatment), RE logit, $N = 3{,}600$			
Pr(Neutral or good time/Positive spending)	−2.88***	0.50	3.37***
	(0.63)	(0.72)	(0.93)
$H6$: F3 (control) vs. F6/F7 (treatment), RE ordered probit, $N = 3{,}580$			
Pr(Bad time/Decreased spending)	2.12***	−0.45	−2.57***
	(0.55)	(0.44)	(0.68)
Pr(Neutral time/Same spending)	−0.73***	−0.09	0.64**
	(0.20)	(0.09)	(0.23)
Pr(Good time/Increased spending)	−1.39***	0.55	1.93**
	(0.36)	(0.53)	(0.62)
$H7$: F3 (control) vs. F1 (treatment), RE ordered probit, $N = 7{,}196$			
Pr(Bad time/Plan to decrease)	2.00***	1.78***	−0.22
	(0.39)	(0.46)	(0.51)
Pr(Neutral time/Plan the same)	−0.66***	−0.58***	0.08
	(0.14)	(0.16)	(0.18)
Pr(Good time/Plan to increase)	−1.34***	−1.19***	0.14
	(0.26)	(0.31)	(0.34)

Notes: This table reports how the average predicted probability of each response category changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, for Hypothesis Tests 4–7 on question-type framing, based on regression specification (3.1). Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the question type. Hypothesis Test 4 compares quantitative (F4/F5) and qualitative (F7) recall-based spending at the extensive margin, in a between-subject comparison restricted to multi-person households, estimated by logit. Hypothesis Test 5 compares attitude (F3) with quantitative recall-based spending (F4/F5) at the extensive margin, in a within-subject comparison of all Group 1 respondents, estimated by random-effects logit. In both tests, quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item; qualitative recall-based spending (F7) is coded as a binary indicator for increased spending, and attitude (F3) as a binary indicator for a positive or neutral spending attitude. Hypothesis Test 6 compares attitude (F3) with qualitative recall-based spending (F6/F7), taking the household-level response (F7) where both were answered, and Hypothesis Test 7 compares attitude (F3) with planned spending (F1); both are within-subject comparisons of three-category ordered items entering without transformation, estimated by random-effects ordered probit. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method; they are robust (Huber-White) in Hypothesis Test 4 and clustered at the individual level in the random-effects models. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

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

The findings are robust to estimating F3–F4 and F3–F5 separately, indicating that pooling F4 and F5 does not drive the result. An alternative comparison of attitude (F3) with 1-month quantitative recall-based spending (F2) yields a closely aligned pattern (Online Appendix Table B.11).

Panel 3 presents the results for *Hypothesis Test 6*, which compares attitudinal responses (F3) with qualitative recall-based spending (F6/F7). Both items are three-category ordered, so changes in predicted probabilities are reported for each response category. For the attitudinal question, an increase in inflation expectations from 2.2 to 4.5 percent is associated with a significant shift in the probability of every response category; for qualitative recall-based spending, none of the three changes is significant, and the difference between the two measures is significant in each category.

The findings are robust to comparing F3 separately with individual-level (F6) and household-level (F7) qualitative recall-based spending, indicating that the result does not depend on the reference unit of the recall measure (Online Appendix Table B.12). Recoding qualitative recall-based spending to reflect real changes, as in Hypothesis Test 1, further strengthens the result: the probability of reporting a real increase in spending rises significantly with inflation expectations, while the probability of a positive or neutral attitude falls, so that the two measures move in opposite directions (Online Appendix Table B.13).

Taken together with the results of Panel 2, these findings show that attitudinal measures are more strongly associated with inflation expectations than either qualitative or quantitative recall-based measures. This pattern is consistent with inflation expectations being more closely related to willingness to spend than to realized spending.

Panel 4 of Table 4 presents the results of *Hypothesis Test 7*, contrasting attitudinal responses (F3) with planned spending (F1). For both consumption measures, expected inflation is significantly associated with all response categories ($p < 0.001$). The changes go in the same direction for both consumption measures: an increase in expected inflation from 2.2 to 4.5 percent is associated with a lower probability of the most favorable category and a higher probability of the least favorable category. The differences between the two measures are statistically insignificant across all response categories, providing no evidence that their associations with inflation expectations differ. Attitudes and forward-looking spending plans thus appear to capture similar sentiment.

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 reflect differences in the underlying constructs captured by these measures. Specifically, both attitude and planned spending reflect willingness to spend, whereas recall-based measures capture realized consumption. Consistent with this interpretation, we find no question-type fram-

ing effects when comparing quantitative and qualitative recall-based measures, both of which refer to realized past consumption. Attitudinal and planned-spending measures, in contrast, exhibit a stronger association with inflation expectations than recall-based measures, consistent with realized consumption additionally reflecting ability and necessity. A further distinction is that recall-based measures require respondents to retrieve past consumption and may therefore be more susceptible to recall error, whereas attitudes and planned spending do not require retrospective recall. Both differences may contribute to the stronger association of attitudinal and planned-spending measures with inflation expectations.

Result 3 (Question-type framing operates across constructs, not within them). The estimated association between inflation expectations and durable consumption differs systematically across question types that capture different constructs. Attitudinal and planned-spending measures, which both primarily elicit willingness to spend, are more strongly associated with expected inflation than recall-based measures of realized spending, which additionally reflect ability and necessity. We find no evidence of a difference between the two recall-based measures, nor between attitudinal responses and planned spending, indicating that the observed question-type framing effects arise across, rather than within, these underlying constructs.

These findings demonstrate that the estimated relationship between inflation expectations and durable consumption is highly dependent on how consumption is elicited in surveys. The most consistent framing effects emerge from question-type differences, particularly the stronger association from attitudinal questions. Effects along the reference unit and time horizon dimension, by contrast, are limited.

Our findings help reconcile divergent results in the empirical literature by showing that they need not reflect differences across samples, countries, time periods, or economic environments: different consumption measures can yield different estimates of the same underlying relationship even within the same sample and at the same point in time. Our unified framework thus shows that the various survey questions used to measure consumption in the related literature are not interchangeable. In particular, attitudinal measures are more strongly associated with inflation expectations than recall-based reports of realized spending, consistent with these measures capturing different constructs—willingness to spend versus realized consumption—while planned spending closely aligns with attitudes. 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.

5 Robustness Checks, Validation Exercises, and Replication

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. This section provides additional evidence on the robustness and interpretation of these findings.

First, we assess the sensitivity of our results to alternative estimation strategies, including linear, censored and two-part models. We then conduct three complementary validation exercises. The first validation exercise benchmarks selected survey measures against aggregate consumption data. This comparison assesses whether measures that differ in their relationship with inflation expectations also differ in how closely they track realized aggregate consumption.

The second validation exercise examines the relationships among the alternative survey measures. We use pairwise correlations and confirmatory factor analysis to assess whether attitudinal, planned, and recall-based questions appear to capture closely related or distinct dimensions of spending. We then use respondents stated motivations for durable purchases to further interpret why attitudes and realized spending may diverge.

The third validation exercise applies the empirical specifications of prior studies within our unified survey framework. This allows us to assess whether divergent patterns documented in the literature can be reproduced when country, institutional environment, survey fielding period, sample, and survey platform are held constant, while the consumption measure varies.

5.1 Alternative Estimation Strategies

To assess the robustness of our results to estimation choices, we re-estimate the baseline specification (3.1) using alternative methods. Estimating the relationship between inflation expectations and each consumption measure in separate regressions, rather than jointly, yields changes in predicted outcomes of the same sign and similar magnitude (Online Appendix Tables B.14–B.20).

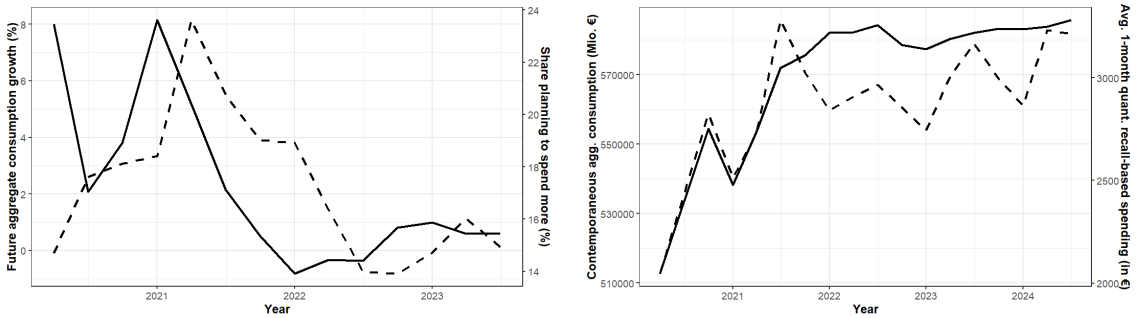
Replacing the ordered probit and logit models with linear estimations yields interaction effects that align with the differences in the change of predicted probabilities (Online

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 the baseline specification for Hypothesis Tests 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).

5.2 Validation Exercise I: Aggregate Consumption Benchmarking

We compare selected survey-based consumption measures with aggregate consumption data. In the BOP-HH survey, planned spending (F1) and past-month quantitative recall-based spending (F2) are the only consumption measures available across repeated waves. We therefore compare these measures with German national-account data. Assessing the aggregate relevance of the attitudinal measure (F3) requires a separate dataset: the Michigan Survey of Consumers, which has been widely used in prior research. Figure 2 presents these time-series comparisons. Details on the methodology are provided in Online Appendix C.1.



Notes: Panel A plots 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: Survey-Based Consumption Measures and Aggregate Consumption in Germany

Panel A of Figure 2 plots average planned spending (dashed line)—defined as the share of respondents reporting plans to increase durable spending—alongside subsequent

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 pattern is consistent with the interpretation that the quantitative recall-based measure more closely reflects realized consumption than the planned-spending measure.

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 provide suggestive evidence that past-month quantitative recall-based spending is more closely aligned with realized aggregate consumption than the planned-spending or attitudinal measures.

5.3 Validation Exercise II: Comparing Survey Measures

This second validation exercise examines whether the alternative survey-based consumption measures are interchangeable or instead capture distinct constructs and behavioral dimensions. The results in Section 4 show that these measures differ in their relationship with inflation expectations, while Section 5.2 shows that they also differ in how closely they align with aggregate consumption data. We therefore compare the survey measures to each other by examining pairwise correlations across measures and using confirmatory factor analysis to test whether attitudinal and recall-based measures load onto a common latent construct. Details on the confirmatory factor analysis are provided in Online Appendix C.2.

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 corre-

	F1	F2	F3	F4	F5	F6	F7
F1	1						
F2	-0.2184	1					
F3	0.1748	0.0777	1				
F4	-0.1747	0.6981	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 categorical 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

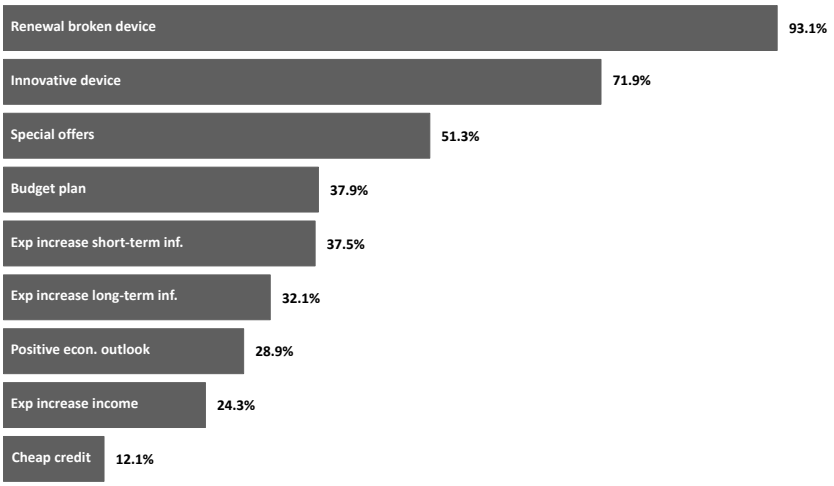
lations at similar recall horizons. Similarly, the qualitative recall-based questions (F6 and F7) are highly correlated despite differences in reference unit. By contrast, correlations across question-types are weak. The attitudinal measure (F3) is only weakly correlated with recall-based spending measures, and its modest association with planned spending (F1) suggests that both capture forward-looking sentiment, although they are not entirely equivalent: planned spending additionally invites consideration of financial feasibility, consistent with the within-subject heterogeneity documented in the results section.

This observation is further supported by three one-factor confirmatory factor analyses (CFAs), which test whether attitude and recall-based measures load onto a common latent factor. Across all specifications, attitude and realized spending measures never load strongly onto a common factor: in Panel A, realized spending loads moderately to strongly while attitude loads very weakly; in Panel B, attitude loads moderately while realized spending loads weakly; and in Panel C, the model does not converge (Online Appendix Table C.1). We therefore interpret the CFA results as suggestive evidence, consistent with the view that attitude captures a different dimension of consumption than realized spending. Different consumption questions reflect distinct behavioral dimensions rather than interchangeable measures.

5.4 Why Attitudes and Realized Spending Diverge

These findings reinforce our earlier conclusion in Section 4: 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).



Notes: 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 question used.

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

To further interpret the weak relationship between inflation expectations and 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%). By comparison, 37.5% cited expected short-term price increases, and only 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 alone. They are therefore consistent with the interpretation that attitudes toward purchasing primarily capture willingness or sentiment, whereas realized spending also reflects constraints and circumstances surrounding the purchase decision.

5.5 Validation Exercise III: Replicating Prior Studies

We conduct a *conceptual 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 iden-

tifying 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 association 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 likely a driver of these inconsistencies and thereby offer a potential reconciliation of the literature. Online Appendix D 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 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), as shown in Online Appendix Table D.1. However, we do not replicate the positive relationship they report for qualitative recall-based spending. In our data, expected inflation has a negligible or even slightly negative association with qualitative recall-based spending (Online Appendix Table D.2). 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 prices (as opposed to stable prices) is associated with a higher likelihood of reporting positive recall-based spending and more favorable spending attitudes, consistent with their extensive-margin result (Online Appendix Tables D.3–D.4). Similarly, we find a negative association at the intensive margin, i.e. higher average inflation expectations are associated with lower consumption. While the association is larger at the extensive margin than at 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, the intensive margin is more strongly associated with spending attitudes 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 (Online Appendix Table D.5). Ichiue and Nishiguchi (2015) also also examine how inflation perceptions relate to qualitative recall-based spending, but we cannot replicate this part of their study due to data-limitations.

In summary, this conceptual replication exercise shows that several contradicting findings from the literature can be reproduced within our unified survey framework. By holding the country, institutional environment, fielding period, sample, and survey platform constant while varying the consumption concept elicited, they provide further evidence that differences in survey design contribute to the divergent relationships between inflation expectations and consumption reported across studies. This also lends some support to the external validity of our results: the fact that our dataset can reproduce patterns documented in studies from other cultural, institutional, and macroeconomic contexts suggests that the underlying measurement issue is not unique to the German context.

Taken together, the three validation exercises support the interpretation that the alternative survey questions capture related but distinct dimensions of consumption. Attitudinal and planned-spending measures are more closely aligned with willingness to spend, whereas recall-based measures more directly reflect realized expenditure. Realized durable spending, however, depends not only on willingness but also on whether households have the financial capacity to purchase and whether a purchase is necessary. Consistent with this interpretation, quantitative recall-based spending aligns more closely with aggregate consumption, while the prior-study replications reproduce several of the divergent patterns in the literature when the consumption concept elicited differs. Thus, the mixed evidence on inflation expectations and consumption is likely to reflect, at least in part, differences in whether survey questions capture sentiment, intended spending, or realized spending under practical constraints.

6 Conclusion

Understanding how inflation expectations shape household consumption is central to both economic theory and policy. Despite extensive empirical research, no consensus has

emerged on how inflation expectations relate to consumption decisions. We show that 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 an experiment embedded in a representative household survey, we show that the estimated relationship between consumption and inflation expectations varies systematically with the elicitation framing of consumption questions.

The main result is straightforward: elicitation framing matters, but not always. The overall relationship between expected inflation and consumption is weak, while 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. For temporal framing (past 1, 3, or 12 months), we find suggestive evidence that expected inflation relates stronger to consumption measured over shorter time horizons. Question-type framing (attitudes, planned spending, quantitative and qualitative recall-based spending), however, produce significant differences in the observed relationship. For the attitudinal measure, which asks about the general attractiveness of buying durables at present, respondents with higher inflation expectations are less willing to spend. Most notably, this measure closely mirrors planned future consumption: both are negatively associated with expected inflation. By contrast, the quantitative and qualitative recall-based measures show at most a weak positive relationship with inflation expectations. This contrast suggests that the different question types capture distinct constructs. Attitudinal and planned-spending measures capture a common forward-looking construct—households willingness to spend. Recall-based measures, by contrast, more closely reflect realized spending, which is also shaped by households ability and necessity to spend.

This interpretation is supported by three validation exercises. First, aggregate consumption benchmarking shows that quantitative recall-based spending tracks realized aggregate consumption substantially more closely than planned-spending or attitudinal measures. To shed new light on what drives realized spending, we also directly ask respondents about their main reasons for purchasing durable goods. Respondents report that their purchases are primarily driven by factors such as the need to replace broken items, product innovation, and special offers, while expected price increases play a considerably smaller role. Second, correlations across the survey measures show greater alignment within than across question types. Factor analysis further supports the distinction between attitudinal and recall-based measures as capturing different constructs. Third, our conceptual

replication of prior studies reproduces several of the divergent patterns documented in the literature within our unified survey framework, suggesting that differences in the consumption concept elicited can help explain conflicting findings across studies.

The external validity of our findings should nevertheless be interpreted with caution, as our evidence comes from a German household survey. The relationship between inflation expectations and consumption may differ across cultural, institutional, and macroeconomic contexts. At the same time, our replication of patterns documented in other settings suggests that the underlying measurement issue is not unique to the German context.

Our findings have important implications for researchers and central banks seeking to assess the expectations channel of monetary policy and the predictions of macroeconomic theory. Consistent with intertemporal substitution, we find that higher inflation expectations are associated with lower planned future spending, a pattern closely mirrored by the attitudinal consumption measure. By contrast, recall-based measures of realized spending show at most a weak positive association with inflation expectations. Thus, the relationship predicted by theory is apparent in households willingness and plans to spend, but much less evident in realized consumption.

At the same time, caution is warranted when using recall-based measures to assess this mechanism. While the empirical literature commonly uses retrospective spending questions as measures of current consumption, these measures relate spending over a preceding period to inflation expectations elicited at the time of the survey and referring to future inflation. This timing mismatch means that such estimates do not provide a direct test of whether previously held inflation expectations caused realized consumption to change.

For researchers and policymakers, the broader implication is that the choice of consumption measure should depend on the object of interest. Attitudinal and planned-spending questions are informative about households willingness and intentions to spend, whereas recall-based measures are better suited to capturing realized spending. These measures should therefore be viewed as complementary rather than interchangeable. More generally, estimates of the relationship between inflation expectations and consumption should be interpreted in light of the specific consumption concept elicited, particularly when comparing findings across studies or using survey evidence to assess the expectations channel of monetary policy.

Appendix A

This section summarizes all empirical papers on the relationship between expected inflation and consumption. This is a 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 et al. (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-based 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 significant and positive. 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	Significant, positive 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). Significant, positive 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	Insignificant estimate, unless one considers heterogeneity across education, income and mortgage balances. The more educated, the ones with lower income and a mortgage have a significant, positive estimate.

Note: All qualitative and quantitative consumption measures in this table refer to recall-based 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	Insignificant 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 significant, positive 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, significant, positive effect across most phases, but not when inflation decreases.

Note: All qualitative and quantitative consumption measures in this table refer to recall-based 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 Paper 01/2020.

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. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

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.

Haan, P., Sun, C., Weinhardt, F., and Weizsäcker, G. (2025). Measuring long-run expectations that correlate with investment decisions. Technical report, DIW Discussion Papers.

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. Journal. Series A: General*, 145(1):42–57.

Katona, G. (1975). *Psychological Economics*. Elsevier Scientific Publishing Company.

Kikuchi, J., Nakazono, Y., and Tango, K. (2020). Inflation Expectations and Consumer Spending: Micro-Data Evidence. *Available at SSRN 5018789*.

List, J. A. (2025). The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs. Working Paper 33456, National Bureau of Economic Research.

Marenčák, M. (2023). State-dependent inflation expectations and consumption attitudes. *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*, 231:106899.

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. Temi di Discussione (Working Paper) 1276, Bank of Italy.

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 Publications.

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*. Cambridge University Press.

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.

Online Appendix

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

The online appendix contains the following materials: (i) an overview of the survey structure (Online Appendix A); (iv) figures of the predicted outcomes and the average marginal effects for each hypothesis test (Online Appendix B.4) (iii) all hypothesis tests, additional results and robustness checks (Online Appendix B); (iv) details on the aggregate consumption benchmarking and the confirmatory factor analysis (Online Appendix C); and (v) replications of prior studies examining the relationship between expected inflation and consumption (Online Appendix D).

A Survey Structure

In addition to the questions used in our paper, the BOP-HH survey includes a wide range of items. As our key questions are not presented consecutively in the survey, we briefly outline the overall structure:

The survey begins by eliciting macroeconomic expectations, including perceived and expected inflation. This is followed by the core consumption questions: F1 (planned spending) and F2 (1-month quantitative recall-based spending). Next are questions on home ownership and house price expectations.

Our experimental module follows, starting with questions on household type and attitude towards durable consumption. Respondents are then randomly assigned to one of two groups: one answers quantitative recall-based questions, the other qualitative recall-based questions. After this, we ask about the reasons for purchasing durable goods. The survey concludes with socio-demographic questions.

The complete questionnaire for the March wave is available online.

German: https://www.bundesbank.de/resource/blob/918014/108b89f71062014688c2e3694a394cfd/472B63F073F071307366337C94F8C870/questionnaire-51-data.pdf

English: https://www.bundesbank.de/resource/blob/918016/709ec3dea159eb06b9eb1d12cb2eb1cb/472B63F073F071307366337C94F8C870/questionnaire-51-data.pdf

The identification numbers for the consumption questions and the experimental module are CM006 and P5100 to P5106.

4 Questions

CP003A | Core | Perceived problems | percprob_[a,d,j,k]

Respondent group: all

The order of the four categories is generated randomly for each respondent.

Question: To what extent do you think the following developments/matters are a serious problem at present?

Note: Please select one answer for each row.

1 No problem at all
2 -> 9 [no label]
10 An extremely serious problem

a Climate change
d Economic situation
j Refugee crisis
k Geopolitical situation

CM001 | Core-M | Expectations qualitative | expmacroquali_[a-i]

Respondent group: all

After the term “inflation rate”, an info box (i) with the following definition is shown: “Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index.”

The order of the nine categories is generated randomly for each respondent.

Now we would like to ask you about your assessment of general economic developments in Germany over the next twelve months.

Question: What developments do you expect in the following metrics over the next twelve months? Will …

1 Decrease significantly
2 Decrease slightly
3 Remain roughly the same
4 Increase slightly
5 Increase significantly

Note: For more information, please click the (i) button.

a The unemployment rate in Germany
b Rents in your area
c Lending rates
d Interest rates on savings accounts
e The inflation rate (i)
f Property prices in your area
g Economic growth in Germany
h Fuel prices
i {If respondent group = refresher: The German stock index (DAX)}

--- | Core-M | THE INFLATION RATE – INTRO

Respondent group: all

Now we would like you to think more carefully about the development of the inflation rate.

The inflation rate

Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as deflation.

CQ002 | Core-Q | Inflation development | devinfpoint

Respondent group: refresher + drandom0 = 1

Range of valid values: -100.0 to +100.0

Question: What do you think the rate of inflation or deflation was in Germany over the past twelve months?

Note: If you assume there was deflation, please enter a negative value. Values may have a maximum of one decimal place.

Please enter a value here:

[Input field] percent

CM002 | Core-M | Expectation inflation or deflation | infdef

Respondent group: all

Question: Do you think inflation or deflation is more likely over the next twelve months?

Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as deflation.

Please select one answer.

1 Inflation more likely
2 Deflation more likely

CM003 | Core-M | Inflation expectations quantitative | inflexppoint

Respondent group: all

Both the inflation and deflation rates are entered and stored as positive values. The value entered under CM002 indicates whether the respondent is expecting inflation or deflation.

Range of valid values: 0.0 to 100.0

If CM002 = 1 | -9997 | -9998

Question: What do you think the rate of inflation will roughly be over the next twelve months?

If CM002 = 2

Question: What do you think the rate of deflation will roughly be over the next twelve months?

Note: Inflation is the percentage increase in the general price level. It is mostly measured using the consumer price index. A decrease in the price level is generally described as deflation.

Please enter a value in the input field (values may have one decimal place).

[Input field] percent

CM004 | Core-M | Inflation expectations probabilistic | infexprob_[a-j]

Respondent group: all

The programming of the question requires the sum of the ten variables to be 100. The current sum of all entered points is shown to the respondent when answering the question. Respondents are asked to correct their responses if the sum does not equal 100.

Question: In your opinion, how likely is it that the rate of inflation will change as follows over the next twelve months?

Note: The aim of this question is to determine how likely you think it is that something specific will happen in the future. You can rate the likelihood on a scale from 0 to 100, with 0 meaning that an event is completely unlikely and 100 meaning that you are absolutely certain it will happen. Use values between the two extremes to moderate the strength of your opinion. Please note that your answers to the categories have to add up to 100.

a The rate of deflation (opposite of inflation) will be 12% or higher.
b The rate of deflation (opposite of inflation) will be between 8% and less than 12%.
c The rate of deflation (opposite of inflation) will be between 4% and less than 8%.
d The rate of deflation (opposite of inflation) will be between 2% and less than 4%.
e The rate of deflation (opposite of inflation) will be between 0% and less than 2%.
f The rate of inflation will be between 0% and less than 2%.
g The rate of inflation will be between 2% and less than 4%.
h The rate of inflation will be between 4% and less than 8%.
i The rate of inflation will be between 8% and less than 12%.
j The rate of inflation will be 12% or higher.

--- | --- | START SPLIT SAMPLE 1

--- | Split sample 1 | RANDOMISATION 1

Respondent group: all

The sample is split randomly into two groups. Randomisation dummy (drandom1) is used for question CM008A/C.

One factor variable: drandom1
group A1 – one-half of the sample
group B1 – one-half of the sample

CM008C | Core-M | Quantitative long-term inflation | Inflexppoint_long_c

Respondent group: all

Range of valid values: -100.0 to 100.0

Input filter: drandom1 = 1

Question: And what value do you think the rate of inflation or deflation will take over the twelve months between March 2026 and March 2027?

Note: Please enter a value in the input field (values may have one decimal place). If you assume that prices will fall (deflation), please enter a negative value.

[Input field] percent

CM008A | Core-M | Quantitative long-term inflation | Inflexppoint_long_a

Respondent group: all

Range of valid values: -100.0 to 100.0

Input filter: drandom1 = 2

Question: What value do you think the rate of inflation or deflation will take on average over the next five years?

Note: Please enter a value in the input field (values may have one decimal place). If you assume that prices will fall (deflation), please enter a negative value.

[Input field] percent

--- | --- | END SPLIT SAMPLE 1

CQ005 | Core-Q | Interest rate expectations | expint_sav

Respondent group: refresher + drandom0 = 2

Range of valid values: -100.00 to 100.00

{If refresher: We would now like to ask you a little more about your expectations regarding interest rate developments and your income.}

{If drandom0 = 2: We would now like to ask you a little more about your expectations regarding interest rate developments.}

Question: What do you expect interest rates on savings accounts to be on average over the next twelve months?

Note: Please enter a value in the input field (values may have two decimal places). If you assume that interest rates will be negative, please enter a negative value.

[Input field] percent

CQ003 | Core-Q | Income expectations probabilistic | incexp_[a-l]

Respondent group: refresher + drandom0 = 3

Show info box "Household" as in CQ008

The programming of this question requires the sum of the variables to be 100. The current sum of all entered points is shown to the respondent when answering the question. Respondents are asked to correct their responses if the sum does not equal 100.

{If drandom0 = 3: We would now like to ask you a little more about your expectations regarding your income.}

Question: In your opinion, how likely is it that the average monthly net income of your household (i) will change as follows over the next twelve months?

Note: The aim of this question is to determine how likely you think it is that something specific will happen in the future. You can rate the likelihood on a scale from 0 to 100, with 0 meaning that an event is completely unlikely and 100 meaning that you are absolutely certain it will happen. Use values between the two extremes to moderate the strength of your opinion. Please note that your answers to the categories have to add up to 100.

a Decrease by €2,000 or more
b Decrease by between €1,500 and less than €2,000
c Decrease by between €1,000 and less than €1,500
d Decrease by between €500 and less than €1,000
e Decrease by between €250 and less than €500
f Decrease by between €0 and less than €250
g Increase by between €0 and less than €250
h Increase by between €250 and less than €500
i Increase by between €500 and less than €1,000
j Increase by between €1,000 and less than €1,500
k Increase by between €1,500 and less than €2,000
l Increase by €2,000 or more

CQ004 | Core-Q | Past expenditure | spentlastmon_[a-i]

Respondent group: all

Range of valid values for every item: 0 to 100,000

You will now be shown some everyday items that you can or need to buy.

Question: If you think back to last month, roughly how much did you spend in euro on the following items?

Note: Please enter an amount in every field (with no decimal places). If you are not quite sure, please give a rough estimate.

a Major purchases (e.g. car, furniture, electrical appliances, etc.): [Input field] euro
b Essential goods (e.g. food and beverages, non-food items such as cleaning products or similar): [Input field] euro
c Clothing and footwear: [Input field] euro
d Entertainment/recreation (e.g. restaurant visits, cultural events, gym): [Input field] euro
e Mobility (e.g. fuel, car loans and running costs, bus and train tickets): [Input field] euro
f Services (e.g. hairdresser, childcare, medical costs): [Input field] euro
g Travel, holidays: [Input field] euro
h Housing costs (e.g. rent, mortgage, ancillary costs): [Input field] euro
i Savings (e.g. savings account, shares, bonds): [Input field] euro

CM006 | Core-M | Planned expenditure | spendintent_[a-i]

Respondent group: all

Question: Are you likely to spend more or less on the following items over the next twelve months than in the last twelve months? What about ...

1 Plan to spend more
2 Plan to spend roughly the same
3 Plan to spend less

Please select one answer for each row.

a Major purchases (e.g. car, furniture, electrical appliances, etc.)
b Essential goods (e.g. food and beverages, non-food items such as cleaning products or similar)
c Clothing and footwear
d Entertainment/recreation (e.g. restaurant visits, cultural events, gym)
e Mobility (e.g. fuel, car loans and running costs, bus and train tickets)
f Services (e.g. hairdresser, childcare, medical costs)
g Travel, holidays
h Housing costs (e.g. rent, mortgage, ancillary costs)
i Savings (e.g. savings account, shares, bonds)

CZ004 | Z | Typical credit transfer | credit_transfer_[a-e]

Respondent group: all

The programming of the question requires the sum of the five variables to be ten. Respondents are asked to correct their responses if the sum does not equal ten.

Question: Think for a moment about how you transfer money. How did you make your last ten credit transfers? Please allocate these to the following types of transfer:

Notes:
- Please note that your answers to the categories have to add up to ten.
- Do not include payments using payment services such as Apple Pay, Google Pay or PayPal.
- Instant payments allow you to transfer money from one bank account to another in just a few seconds.

a Regular online bank transfer
b Instant online bank transfer or giropay transfer
c Transfer terminal
d Paper-based transfer order
e Other

X I do not make credit transfers.

CQ008 | Core-Q | Home ownership | homeown

Respondent group: all

Show info box in CQ008, CQ003, CS006, CS007, CS008, CS009

A household is defined as persons who live together and "maintain together", i.e. share the cost of daily life, not calculating their purchases separately.

Households in their own right are
(1) persons or groups of persons who live in a shared residence without a family or partner, or
(2) domestic staff residing at that address.

Persons who are temporarily absent, e.g. commuters, students, trainees, children under 18 in joint custody, and persons in hospital or on vacation, are also deemed to be part of a household.
What matters is that the person is only temporarily absent and normally lives in the household, or the address of the household is registered as their main residence.

In the following section, we would like to ask you some questions on real estate.

Question: Does your household (i) live in a rented property or an owner-occupied apartment or house?

Please select one answer.
1 Rent and do not own any other home(s)
2 Rent but own other home(s)
3 Live in own apartment
4 Live in own house

CM005 | Core-M | House price expectations quantitative | exphp_point

Respondent group: all

Range of valid values: -100.0 to +100.0

Question: By what percentage do you think property prices in your area will change over the next twelve months?

Note: Please enter a value in the input field (values may have one decimal place). If you assume that property prices will fall, please enter a negative value.

[Input field] percent

CQ001 | Core-Q | House price expectations probabilistic | exphp_prob_[a-j]

Respondent group: refreshers + drandom0 = 4

The programming of the question requires the sum of the variables to be 100. The current sum of all entered points is shown to the respondent when answering the question. Respondents are asked to correct their responses if the sum does not equal 100.

Question: In your opinion, how likely is it that property prices in your area will change as follows over the next twelve months?

Note: The aim of this question is to determine how likely you think it is that something specific will happen in the future. You can rate the likelihood on a scale from 0 to 100, with 0 meaning that an event is completely unlikely and 100 meaning that you are absolutely certain it will happen. Use values between the two extremes to moderate the strength of your opinion. Please note that your answers to the categories have to add up to 100.

a Decrease by 12% or more
b Decrease by between 8% and less than 12%
c Decrease by between 4% and less than 8%
d Decrease by between 2% and less than 4%
e Decrease by between 0% and less than 2%
f Increase by between 0% and less than 2%
g Increase by between 2% and less than 4%
h Increase by between 4% and less than 8%
i Increase by between 8% and less than 12%
j Increase by 12% or more

P5100 | 2024_04 | Household type | hhtype

Respondent group: all

Show info box “Household” as in CQ008

Question: Do you live in a single-person or multi-person household (i)?

1 Single-person household
2 Multi-person household

P5101 | 2024_04 | Attitude towards spending | attitude

Respondent group: all

The following questions are about major purchases that have been made or are planned by you or your household.

Question: In view of the general economic situation, do you think that now is a good time to make major purchases?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 Yes, now is a good time
2 It is neither a good nor a bad time
3 No, now is a bad time

--- | --- | START SPLIT SAMPLE 2

--- | Split sample 2 | RANDOMISATION 2

Respondent group: all

The sample is split randomly into two groups. Randomisation dummy (drandom2) is used for questions P5102A-P5105B.

One factor variable: drandom2
group A2 – one-half of the sample
group B2 – one-half of the sample

P5102A | 2024_04 | Purchases over the past twelve months | purchase_[a,b]

Respondent group: all

Input filter: drandom2 = 1

Question: Did your household make one or more major purchases during the following periods?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 Yes
2 No

a Over the past three months (December 2023 to February 2024)
b Four to twelve months ago (March to November 2023)

P5103A | 2024_04 | Purchases over the past three months | purchase_3m

Respondent group: all

Input filter: drandom2 = 1 & (purchase_a = 1 & (purchase_b != 1))

Range of valid values: 1 to 100,000

Question: Over the past three months, roughly how much did your household spend on major purchases, in euro?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

[Input field] euro

P5104A | 2024_04 | Purchases over the past twelve months | purchase_12m

Respondent group: all

Input filter: drandom2 = 1 & purchase_b = 1

Range of valid values: 1 to 100,000

Question: Over the past twelve months, roughly how much did your household spend on major purchases, in euro?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

[Input field] euro

P5102B | 2024_04 | Purchases, individual | purchase_individual

Respondent group: all

Input filter: drandom2 = 2

Randomise the order of P5102B und P5103B

{If purchase_household first & purchase_household != -9997 | -9998: You just indicated that your household's spending on major purchases over the past twelve months was {if purchase_household = 1: significantly lower than} {if purchase_household = 2: roughly the same as} {if purchase_household = 3: significantly higher than} in an average year.}

Question: How high was **your individual spending** on major purchases over the past twelve months compared with an average year?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 Significantly lower
2 Roughly the same
3 Significantly higher

P5103B | 2024_04 | Purchases, household | purchase_household

Respondent group: all

Input filter: drandom2 = 2 & hhtype = 2

{If purchase_individual first & purchase_individual != -9997 | -9998: You just indicated that your personal spending on major purchases over the past twelve months was {if purchase_individual = 1: significantly lower than} {if purchase_individual = 2: roughly the same as} {if purchase_individual = 3}: significantly higher than} in an average year.}

Question: How high was **your household's spending** on major purchases over the past twelve months compared with an average year?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 Significantly lower
2 Roughly the same
3 Significantly higher

P5104B | 2024_04 | Purchases, nominal – increase | purchase_increase_nominal

Respondent group: all

Input filter: drandom2 = 2 & (purchase_household = 3 | purchase_individual = 3 & hhtype =1)

{If purchase_household = 3} You stated that your household's spending on major purchases has **increased significantly** over the past twelve months.

{If purchase_individual = 3 & hhtype = 1} You stated that your individual spending on major purchases has **increased significantly** over the past twelve months.

Question: Why do you think your spending on major purchases has increased significantly?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 {if purchase_household = 3: My household} {if purchase_individual = 3 & hhtype = 1: I} made roughly the same number of major purchases, but they were more expensive.
2 {if purchase_household = 3: My household} {if purchase_individual = 3 & hhtype = 1: I} made a larger number of major purchases.
3 The two reasons apply equally.

P5105B | 2024_04 | Purchases, nominal – decrease | purchase_decrease_nominal

Respondent group: all

Input filter: drandom2 = 2 & (purchase_household = 1 | purchase_individual = 1 & hhtype = 1)

{If purchase_household = 1} You stated that your household's spending on major purchases has **decreased significantly** over the past twelve months.

{If purchase_individual = 1 & hhtype = 1} You stated that your individual spending on major purchases has **decreased significantly** over the past twelve months.

Question: Why do you think your spending on major purchases has decreased significantly?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 {if purchase_household = 3: My household} {if purchase_individual = 3 & hhtype = 1: I} made a smaller number of major purchases, but they were more expensive.
2 {if purchase_household = 1: My household} {if purchase_individual = 1 & hhtype = 1: I} made a smaller number of major purchases because {if purchase_household = 1: my household} {if purchase_individual = 1 & hhtype = 1: I} did not need or want to buy as much.
3 The two reasons apply equally.

--- | --- | END SPLIT SAMPLE 2

--- | Split sample 3 | RANDOMISATION 3

Respondent group: all

The sample is split randomly into two groups. Randomisation dummy (drandom3) is used for question P5106.

One factor variable: drandom3
group A3 – one-half of the sample
group B3 – one-half of the sample

P5106 | 2024_04 | Reasons for buying durable goods | reasons_durable_[a-l]

Respondent group: all

Randomise the order of the scale: drandom3 = 1: 1 -> 7, drandom3 = 2: 7 -> 1

Randomise items

We would now like to better understand when and why you make major purchases.

Question: To what extent do you agree or disagree with the following statements?

Note: Major purchases are purchases of cars, furniture, electrical appliances, etc.

1 Strongly disagree
2-6 -> no label
7 Strongly agree

I make major purchases …

a … only when necessary (e.g. current appliance is broken).
b … if I am expecting prices to go up over a fairly short period of time (around twelve months).
c … if I am expecting prices to go up over a fairly long period of time (around five years).
d … when they are on offer (e.g. Black Friday, sales).
e … when I am optimistic about future economic developments (e.g. lower unemployment rate and/or higher economic growth).
f … when I have access to cheap consumer credit.
g … when I want to replace the current appliance with a newer or better one (e.g. higher efficiency, better energy savings, more convenience, better functionality).
h … based on a fixed budget plan, to which I make only slight changes over time.
i … if I am expecting a pay rise.

---

6 Socioeconomic data

CO001 | Core-S | Residence in 1989 | eastwest1989

Respondent group: refresher only

Input filter: year of birth = 1990 or earlier

Question: In which part of Germany were you living shortly before the fall of the Berlin Wall on 9 November 1989?

1 In eastern Germany, the former German Democratic Republic
2 In western Germany, the Federal Republic of Germany
3 I moved to Germany after 1989

CS001 | Core-S | School education | eduschool

Respondent group: refresher only

Question: What is your highest level of school education?

1 Still at school
2 Completed lower secondary school
3 Completed higher secondary school
4 Polytechnical secondary school certificate, 10th grade
5 University of applied sciences entrance diploma/completed technical school
6 General or subject-specific university entrance diploma/senior school-leaving certificate (from a grammar school)/East German secondary school up to 12th grade (also with apprenticeship)
7 Other school-leaving certificate
8 No school-leaving certificate (and currently not a student)

CS002 | Core-S | Professional education | eduwork

Respondent group: refresher only

Input filter: eduschool > 1

Question: What level of vocational training or university degree do you have? Please think of your highest qualification here.

1 Currently in training or studying (bachelor's degree not yet completed)
2 Completed vocational training (apprenticeship)
3 Completed vocational training (vocational school or commercial college)
4 Completed training at a technical or commercial college, school for master craftsmen or engineers or university of cooperative education with shorter preparation time (up to 880 hours)
5 Completed training at a university of cooperative education with longer preparation time (more than 880 hours)
6 Bachelor's degree, applied sciences degree, completed training at an engineering college
7 Diploma or master's degree, completed teacher training degree
8 Doctorate/postdoctoral qualification obtained
9 Other professional qualification
10 No vocational training completed (and currently not in training/higher education)

CS003 | Core-S | Employment status | employ

Respondent group: all

Question: Which of the following best describes your current employment status?

1 In full-time employment (including apprenticeship)
2 In part-time employment (including phased retirement)
3 In casual or irregular employment (including paid internship and integration measures)
4 On maternity leave/parental leave/longer-term sick leave/other leave; planning to return to work
5 Unemployed (officially registered)
6 At school, university or in an unpaid internship
7 Retiree or pensioner
8 Retired early or about to retire (including unfit for work or reduced ability to work)
9 Federal volunteer service/voluntary year
10 Homemaker
11 Other form of non-employment

CS004 | Core-S | Short-time work | employ2

Respondent group: all

Filter: employ1 = 1,2,3,4

Question: Are you currently receiving short-time working benefits?

1 Yes
2 No

CS005 | Core-S | Profession | profession

Respondent group: all

Filter: employ1 = 1,2,3,4

Question: Which professional status currently applies to you?

1 Non-salaried employee, including in agriculture
2 Salaried employee
3 Civil servant (including judge, career and regular soldier)
4 Self-employed or entrepreneur (including self-employed farmer) without employees
5 Self-employed or entrepreneur (including self-employed farmer) with employees
6 Trainee/intern
7 Unpaid family worker
8 Other

CS006 | Core-S | Household size | hhsize

Respondent group: all

Show info box “Household” as in CQ008

Question: How many persons live permanently in your household(i), including yourself? Please also include all children living in your household.

[Input field]

CS007 | Core-S | Household children | hhchildren

Respondent group: all

Show info box “Household” as in CQ008

Input filter: hhsize > 1

Question: And how many children under the age of 18 live in your household (i)?

[Input field]

CS010 | Core-S | Marital status | familystatus

Respondent group: refresher only

Question: What is your marital status?

Note: The question is referring to your legal marital status. “Registered partnership” means that it is or was a same-sex partnership within the meaning of the German Law on Civil Partnerships (*Lebenspartnerschaftsgesetz*).

1 Married/in a registered partnership, living with partner
2 Married/in a registered partnership, living separately
3 Single
4 Single, living with partner
5 Divorced/registered partnership dissolved
6 Widowed/registered partner deceased

CS008 | Core-S | Household income | hhinc

Respondent group: all

Show info box “Household” as in CQ008

Question: What is the total monthly net income of your household (i)?

Note: This refers to the total amount, comprising wages, salaries, income from self-employment and pensions, in each case after deducting tax and social security contributions. In this amount, please include any income received through public aid, earnings from rents and leases, housing allowance, child benefits and any other sources of income.

If your net household income has varied over the last twelve months, please enter the average of the last twelve months.

1 Less than €500
2 €500 to €999
3 €1,000 to €1,499
4 €1,500 to €1,999
5 €2,000 to €2,499
6 €2,500 to €2,999
7 €3,000 to €3,499
8 €3,500 to €3,999
9 €4,000 to €4,999
10 €5,000 to €5,999
11 €6,000 to €7,999
12 €8,000 to €9,999
13 €10,000 or more

CS009 | Core-S | Personal income | pinc

Respondent group: all

Input filter: hhsize > 1

Question: And what is your personal total monthly net income?

Note: If your net personal income has varied over the last twelve months, please enter the average of the last twelve months.

1 Less than €500
2 €500 to €999
3 €1,000 to €1,499
4 €1,500 to €1,999
5 €2,000 to €2,499
6 €2,500 to €2,999
7 €3,000 to €3,499
8 €3,500 to €3,999
9 €4,000 to €4,999
10 €5,000 to €5,999
11 €6,000 or more

CS011 | Core-S | Main shopper | mainshopper_[a-d]

Respondent group: refresher only

Input filter: hhsize > 1

Question: In your household, who is primarily responsible for the following?

1 Mostly me
2 My partner/another member of the household and me together
3 Mostly my partner/another member of the household

a Everyday purchases (e.g. grocery shopping)
b Major purchases (e.g. furniture, car)
c Meal planning and preparation
d Decisions regarding savings and financial assets

CQ007A | Core-Q | Wealth and debt finer categories | netwealth_detail_[a-g]

Respondent group: refresher only

Show info box item d: "If you are the owner of or partner in a business or company, please enter the rough value of your ownership/equity."

Question: How high do you estimate the assets and liabilities of your household to be?

Categories for all asset classes other than real estate	**Real estate**
1. (No bank deposits/no securities/no equity/no other assets)	
2. €1 to less than €2,500	1. No real estate
3. €2,500 to less than €5,000	2. €1 to less than €100,000
4. €5,000 to less than €10,000	3. €100,000 to less than €200,000
5. €10,000 to less than €25,000	4. €200,000 to less than €300,000
6. €25,000 to less than €50,000	5. €300,000 to less than €400,000
7. €50,000 to less than €100,000	6. €400,000 to less than €500,000
8. €100,000 to less than €250,000	7. €500,000 to less than €750,000
9. €250,000 to less than €500,000	8. €750,000 to less than €1 million
10. €500,000 or more	9. €1 million to less than €1.5 million
	10. €1.5 million or more

Collateralised loans	**Other loans**
1. No loans	1. No loans
2. Debts totalling €1 to less than €25,000	2. Debts totalling €1 to less than €1,000
3. €25,000 to less than €50,000	3. €1,000 to less than €2,000
4. €50,000 to less than €100,000	4. €2,000 to less than €5,000
5. €100,000 to less than €150,000	5. €5,000 to less than €10,000
6. €150,000 to less than €200,000	6. €10,000 to less than €20,000
7. €200,000 to less than €300,000	7. €20,000 to less than €40,000
8. €300,000 to less than €500,000	8. €40,000 to less than €60,000
9. €500,000 to less than €750,000	9. €60,000 to less than €100,000
10. €750,000 or more	10. €100,000 or more

Assets

a Bank deposits
b Real estate (estimated market value)
c Securities (shares, bonds including funds/ETFs)
d Ownership of or equity in unlisted businesses or companies (i)
e Other assets

Loans and advances

f Amount of outstanding loans secured by real estate (mortgage loans)
g Amount of other outstanding loans (e.g. overdraft facilities, consumer credit or loans for goods and services, loans to finance an enterprise or a professional activity, loans from friends or family)

B Additional Results and Regressions

B.1 Summary Statistics

	Sample ($N \approx 3,600$)			
	Mean	**SD**	**Min**	**Max**
Attitude towards spending	1.87	0.54	1	3
Individual purchases	2.03	0.65	1	3
Household purchases	2.09	0.66	1	3
Past month's spending (€)	1,068.99	4,924.43	0	55,000
Past three months' spending (€)	1,314.28	5,328.20	0	50,000
Past twelve months' spending (€)	4,914.83	12,069.91	0	90,000
Planned Spending	1.87	0.70	1	3
Inflation Expectation (%)	3.55	3.74	-25.00	30.00
Interest Rate Expectations on Savings Account (%)	1.92	1.46	-4	22
Expects Interest Rate on Savings Account Increase (%)	15.25	.	.	.
Expects Interest Rate on Savings Account Decrease (%)	34.77	.	.	.
Expects Interest Rate on Lending Increase (%)	37.45	.	.	.
Expects Interest Rate on Lending Decrease (%)	24.84	.	.	.
Expects Unemployment Increase (%)	40.71	.	.	.
Expects Unemployment Decrease (%)	12.27	.	.	.
Nominal Wage Growth Expectation (%)	3.42	20.65	-300.10	150.08
Monthly Net Household Income (€)	3,000 - 3,499	.	< 500	$\geq 10,000$
Age	54.84	16.13	16	80
Female (%)	41.01	.	.	.
No College (%)	44.86	.	.	.

Table B.1: Summary Statistics

	F1	F2	F3	F4	F5	F6	F7
Panel A							
No Answer (in %)	0.1522	1.1923	0.0507	0.6944	0.4644	0.0508	0.1399
Don't Know (in %)	0.2029	0.6849	0.1268	0.3472	0.9288	0.1523	0.1399
Sum (in %)	0.3552	1.8772	0.1776	1.0417	1.3932	0.2030	0.2797
Panel B							
No Answer	6	47	2	2	3	1	2
Don't Know	8	27	5	1	6	3	2
Sum	14	74	7	3	9	4	4

Note: This table summarizes missing responses for the consumption survey items. Respondents could either report that they have no answer or that they do not know the answer. Panel A is in % and Panel B is in absolute numbers. F1: planned spending; F2/F4/F5: quantitative recall-based (1, 3, 12 months); F3: attitude; F6/F7: qualitative recall-based (individual, household).

Table B.2: Missing Responses

	F4 (treat == 1)	F5 (treat == 0)	Test of mean difference
Mean Age	53.48	51.54	$t(838) = -1.6123$
	(15.38)	(16.22)	$p = 0.107$
Percentage no university	45.02%	42.11%	$\chi^2(1) = 0.6101$
			$p = 0.435$
Percentage low income	5.58%	6.79%	$\chi^2(1) = 0.4308$
			$p = 0.512$
Percentage female	39.44%	38.20%	$\chi^2(1) = 0.1146$
			$p = 0.735$
Percentage unemployed	1.20%	1.36%	$\chi^2(1) = 0.0362$
			$p = 0.849$
Percentage East Germany	16.73%	16.81%	$\chi^2(1) = 0.0007$
			$p = 0.979$
N	251	589	840

Note: The table compares respondents answering the 3-month (F4) and 12-month (F5) quantitative recall-based spending questions in the estimation sample of Hypothesis Test 2. Standard deviations in parentheses. The test of mean difference is a two-sample t-test for age and a Pearson χ^2-test for all other characteristics. Low income is a total monthly net household income of €1,999 or less; no university refers to the lower and medium education categories.

Table B.3: Balance Table: F4 and F5 responses

	F4/F5	F7	Test of mean difference
Mean age	53.99	54.58	$t(2624) = -0.9420$
	(15.90)	(16.02)	$p = 0.346$
Percentage no university	42.31%	41.14%	$\chi^2(1) = 0.3669$
			$p = 0.545$
Percentage low income	3.60%	4.72%	$\chi^2(1) = 2.0577$
			$p = 0.151$
Percentage female	36.83%	39.68%	$\chi^2(1) = 2.2432$
			$p = 0.134$
Percentage unemployed	0.90%	1.01%	$\chi^2(1) = 0.0770$
			$p = 0.781$
Percentage East Germany	17.25%	18.79%	$\chi^2(1) = 1.0529$
			$p = 0.305$
N	1,333	1,293	2,626

Notes: The table compares the two subsamples of Hypothesis Test 4, restricted to multi-person households and to the respondents included in the estimation. F4/F5: Group 1 respondents, for whom quantitative recall-based spending (3 and 12 months) is coded as a binary indicator for positive expenses over the past 12 months; F7: Group 2 respondents answering the qualitative recall-based question at the household level. Standard deviations in parentheses. The test of mean difference is a two-sample t-test for age and a Pearson χ^2-test for all other characteristics. Low income is a total monthly net household income of €1,999 or less; no university refers to the lower and medium education categories.

Table B.4: Balance Table: Sample F4/F5 and Sample F7 (Multi-Person Households)

B.2 Estimation Strategy

To test whether, and along which dimensions, the framing of the consumption question matters for the relationship between inflation expectations and consumption, we conduct pairwise comparisons of the consumption measures. All hypothesis tests are based on Equation (3.1) in Section 3:

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.1)}$$

where C_{im} is the durable consumption response of individual i to consumption measure m, and m indexes the two measures compared in a given hypothesis test. $E_i(\pi_{t+12})$ denotes individual i's twelve-month-ahead inflation expectation. $Treat_{im}$ is a binary indicator for the consumption question the observation comes from; in between-subject comparisons each respondent contributes a single observation, so $Treat_{im}$ reduces to $Treat_i$. The coefficient of interest is β_2, which captures whether the association between inflation expectations and reported consumption differs across the two measures. X_i includes expectations of higher economic growth and sociodemographic characteristics (age, age squared, gender, household income, education level, employment status, an East Germany dummy, household size, and city size). What differs across hypothesis tests is the pair of consumption measures compared, the estimation sample, the coding of the dependent variable, the estimator, and the error structure.

Controlling for expectations of economic growth accounts for the possibility that macroeconomic expectations are related to both inflation expectations and consumption: individuals may anticipate higher inflation because they have a positive outlook on the economy, and they may also consume more for the same reason. It also partially controls for the inherent optimism or pessimism of individuals. This is in line with the estimation strategies in Bachmann et al. (2015); Burke and Ozdagli (2023), D'Acunto et al. (2022), Dräger and Nghiem (2021), Duca-Radu et al. (2021) and Ichiue and Nishiguchi (2015).

Household income is expressed in three categories: low income (a total monthly net household income of €1,999 or less), mid income (between €2,000 and €4,999), and high income (€5,000 or higher). Education is categorized as lower, medium and higher professional education according to the German Qualifications Framework for Lifelong Learning (DQR) (of Education and Research, 2023). Lower education corresponds to individuals with no completed job training or those who are still in school. Medium education corresponds to individuals who completed vocational training. Higher education corresponds

to individuals who completed university or an equivalent.

The seven consumption measures in our survey use different response formats (numeric and categorical), so some of them have to be transformed before they can be compared in pairs. Section B.3 describes the coding for each test. In brief, for reference unit framing (Hypothesis Test 1) both items are three-category ordered items and enter in their original coding. For temporal framing (Hypothesis Tests 2 and 3) the estimation samples contain only strictly positive spending amounts, which we transform into logs. For question-type framing, Hypothesis Tests 4 and 5 compare a quantitative with a categorical measure and therefore require harmonization: we convert both measures into binary indicators and compare them at the extensive margin. In Hypothesis Test 4, for example, the indicator for the qualitative measure equals one if the respondent reports increased spending and zero otherwise, and the indicator for the quantitative measure equals one if the respondent reports positive expenditure and zero otherwise. Hypothesis Tests 6 and 7 compare two three-category ordered items, so no harmonization is required.

B.3 Hypothesis Tests

This section summarizes all hypothesis tests. For each test based on pooled responses, we also report the corresponding disaggregated tests, denoted by subscripts a, b, etc. Throughout, the control question is the measure for which $Treat_{im} = 0$ and the treatment question the measure for which $Treat_{im} = 1$. In all within-subject tests, the model includes individual random effects and standard errors are clustered at the individual level; in between-subject tests, standard errors are robust (Huber-White).

B.3.1 Reference Unit Framing

Hypothesis Test 1 tests whether the relationship between inflation expectations and reported durable spending depends on the reference unit (individual vs. household level).

- Control question: F6 (qualitative recall-based spending at the individual level)
- Treatment question: F7 (qualitative recall-based spending at the household level)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \qquad \text{(B.2)}$$

where C_{im} is the three-category ordered response, $m \in \{\text{F6}, \text{F7}\}$, and $Treat_{im}$ equals one if the observation comes from F7 and zero if it comes from F6. The estimation sample consists of respondents in Group 2 living in multi-person households, who answered both questions in randomized order. Each respondent contributes two observations ($N = 2{,}584$)

and the comparison is within-subject. Both items are three-category ordered items and enter in their original coding. Equation (B.2) is estimated as an ordered probit model with individual random effects.

B.3.2 Temporal Framing

Hypothesis Test 2 tests whether the relationship between inflation expectations and reported durable spending depends on the recall horizon (3 vs. 12 past months).

- Control question: F5 (12-month quantitative recall-based spending)
- Treatment question: F4 (3-month quantitative recall-based spending)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_i + \beta_3 Treat_i + \beta_4' X_i + \epsilon_{im}, \qquad \text{(B.3)}$$

where C_{im} is log spending, $m \in \{\text{F5}, \text{F4}\}$, and $Treat_i$ equals one if the observation comes from F4 and zero if it comes from F5. Each respondent in Group 1 answered only one of the two questions. The questionnaire routed respondents who reported purchases in the last 3 months only to F4 and respondents who reported purchases in both the last 3 months and the preceding 4–12 months or only in the preceding 4–12 months to F5; respondents who reported no purchases were shown neither question. Assignment to F4 or F5 is therefore not random, and we interpret β_2 as a difference between associations estimated in two different groups of consumers (see Section 3 and the balance tests in Online Appendix Table B.3). The estimation sample contains only respondents with strictly positive spending, each contributing one observation ($N = 840$), and the comparison is between-subject. The dependent variable is log spending. Equation (B.3) is estimated as a log-linear model by OLS.

Hypothesis Test 3 tests whether the relationship between inflation expectations and reported durable spending depends on the recall horizon (1 vs. 3/12 past months).

- Control question: F4/F5 (pooled 3- and 12-month quantitative recall-based spending)
- Treatment question: F2 (1-month quantitative recall-based spending)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \qquad \text{(B.4)}$$

where C_{im} is log spending, $m \in \{\text{F4/F5}, \text{F2}\}$, and $Treat_{im}$ equals one if the observation comes from F2 and zero if it comes from F4/F5. The estimation sample is restricted to

respondents who reported strictly positive spending both for F2 and for F4 or F5. Each respondent contributes two observations ($N = 886$) and the comparison is within-subject, with a fixed question order. The dependent variable is log spending. Equation (B.4) is estimated as a log-linear model with individual random effects.

Hypothesis Test 3_a repeats Hypothesis Test 3 for the 3-month horizon only (1 vs. 3 past months).

- Control question: F4 (3-month quantitative recall-based spending)
- Treatment question: F2 (1-month quantitative recall-based spending)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.5)}$$

where C_{im} is log spending, $m \in \{\text{F4}, \text{F2}\}$, and $Treat_{im}$ equals one if the observation comes from F2 and zero if it comes from F4. The estimation sample is restricted to respondents who reported strictly positive spending for both F2 and F4 ($N = 352$). Design, coding and estimator are as in Hypothesis Test 3.

Hypothesis Test 3_b repeats Hypothesis Test 3 for the 12-month horizon only (1 vs. 12 past months).

- Control question: F5 (12-month quantitative recall-based spending)
- Treatment question: F2 (1-month quantitative recall-based spending)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.6)}$$

where C_{im} is log spending, $m \in \{\text{F5}, \text{F2}\}$, and $Treat_{im}$ equals one if the observation comes from F2 and zero if it comes from F5. The estimation sample is restricted to respondents who reported strictly positive spending for both F2 and F5 ($N = 542$). Design, coding and estimator are as in Hypothesis Test 3.

B.3.3 Question Type Framing

Hypothesis Test 4 tests whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question is quantitative or qualitative recall-based.

- Control question: F4/F5 (quantitative recall-based spending over the past 12 months)

  - Indicator equal to one if the respondent reported positive spending over the past 12 months and zero otherwise. Respondents who reported no purchase in the screening item and were therefore not shown F4/F5 are coded as zero.

- Treatment question: F7 (qualitative recall-based spending at the household level)
  - Indicator equal to one if the respondent reported increased spending and zero otherwise.

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_i + \beta_3 Treat_i + \beta_4' X_i + \epsilon_{im}, \qquad \text{(B.7)}$$

where C_{im} is the binary indicator defined above, $m \in \{\text{F4/F5}, \text{F7}\}$, and $Treat_i$ equals one if the observation comes from F7 and zero if it comes from F4/F5. The quantitative questions were administered to Group 1 and F7 to Group 2, and assignment to the two groups is random. Because F7 was administered only to respondents in multi-person households, we apply the same restriction to Group 1. Each respondent contributes one observation (N = 2,626) and the comparison is between-subject. The comparison contrasts a recall-based measure of spending incidence with a recall-based measure of spending change, both over the past 12 months. Equation (B.7) is estimated as a logit model.

Hypothesis Test 5 tests whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or quantitative recall-based spending.

- Control question: F3 (attitude)
  - Indicator equal to one if the attitude is positive or neither positive nor negative, and zero otherwise.

- Treatment question: F4/F5 (pooled 3- and 12-month quantitative recall-based spending)
  - Indicator equal to one if the respondent reported positive spending and zero otherwise. Respondents who reported no purchase in the screening item are coded as zero.

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \qquad \text{(B.8)}$$

where C_{im} is the binary indicator defined above, $m \in \{\text{F3}, \text{F4/F5}\}$, and $Treat_{im}$ equals one if the observation comes from F4/F5 and zero if it comes from F3. F3 was administered

to the full sample and the quantitative questions to Group 1. Every Group 1 respondent contributes two observations (N = 3,600) and the comparison is within-subject, with a fixed question order. The comparison contrasts an evaluative assessment of current buying conditions with a recall-based measure of past spending incidence. Equation (B.8) is estimated as a logit model with individual random effects.

Hypothesis Test 5_a repeats Hypothesis Test 5 for the 3-month horizon only.

- Control question: F3 (attitude), coded as in Hypothesis Test 5
- Treatment question: F4 (3-month quantitative recall-based spending)
  - Indicator equal to one if the respondent reported positive spending over the past 3 months and zero otherwise.

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.9)}$$

where C_{im} is the binary indicator defined above, $m \in \{\text{F3}, \text{F4}\}$, and $Treat_{im}$ equals one if the observation comes from F4 and zero if it comes from F3. Design and estimator are as in Hypothesis Test ($N = 3,612$). Given that the estimation considers extensive margins, individuals who indicated that they did not consume in the past three months are included in the sample.

Hypothesis Test 5_b repeats Hypothesis Test 5 for the 12-month horizon only (excluding the first three months).

- Control question: F3 (attitude), coded as in Hypothesis Test 5
- Treatment question: F5 (12-month quantitative recall-based spending)
  - Indicator equal to one if the respondent reported positive spending over the previous 4 to 12 months and zero otherwise.

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.10)}$$

where C_{im} is the binary indicator defined above, $m \in \{\text{F3}, \text{F5}\}$, and $Treat_{im}$ equals one if the observation comes from F5 and zero if it comes from F3. Design and estimator are as in Hypothesis Test 5 ($N = 3,604$).

Hypothesis Test 5_c repeats Hypothesis Test 5 with 1-month quantitative recall-based spending.

- Control question: F3 (attitude), coded as in Hypothesis Test 5
- Treatment question: F2 (1-month quantitative recall-based spending)
  - Indicator equal to one if the respondent reported positive spending over the past month and zero otherwise.

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.11)}$$

where C_{im} is the binary indicator defined above, $m \in \{\text{F3}, \text{F2}\}$, and $Treat_{im}$ equals one if the observation comes from F2 and zero if it comes from F3. Design and estimator are as in Hypothesis Test 5 ($N = 7,098$)

Hypothesis Test 6 tests whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or qualitative recall-based spending.

- Control question: F3 (attitude)
- Treatment question: F6/F7 (pooled qualitative recall-based spending at the individual and household level)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.12)}$$

where C_{im} is the three-category ordered response, $m \in \{\text{F3}, \text{F6/F7}\}$, and $Treat_{im}$ equals one if the observation comes from F6/F7 and zero if it comes from F3. F6 was administered to all respondents in Group 2 and F7 only to those in multi-person households. Where a respondent answered both, we use the household-level response (F7), so that every Group 2 respondent contributes one qualitative observation alongside the attitudinal response ($N = 3{,}580$). The comparison is within-subject, with a fixed question order. All items are three-category ordered items and enter in their original coding. Equation (B.12) is estimated as an ordered probit model with individual random effects.

Hypothesis Test 6_a repeats Hypothesis Test 6 for the individual level only.

- Control question: F3 (attitude)
- Treatment question: F6 (qualitative recall-based spending at the individual level)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.13)}$$

where C_{im} is the three-category ordered response, $m \in \{\text{F3}, \text{F6}\}$, and $Treat_{im}$ equals one if the observation comes from F6 and zero if it comes from F3. The estimation sample consists of all respondents in Group 2 ($N = 3,578$). Design, coding and estimator are as in Hypothesis Test 6.

Hypothesis Test 6_b repeats Hypothesis Test 6 for the household level only.

- Control question: F3 (attitude)
- Treatment question: F7 (qualitative recall-based spending at the household level)

$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.14)}$$

where C_{im} is the three-category ordered response, $m \in \{\text{F3}, \text{F7}\}$, and $Treat_{im}$ equals one if the observation comes from F7 and zero if it comes from F3. The estimation sample consists of respondents in Group 2 living in multi-person households ($N = 2,586$).

Hypothesis Test 7 tests whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or planned spending.

- Control question: F3 (attitude)
- Treatment question: F1 (planned spending)

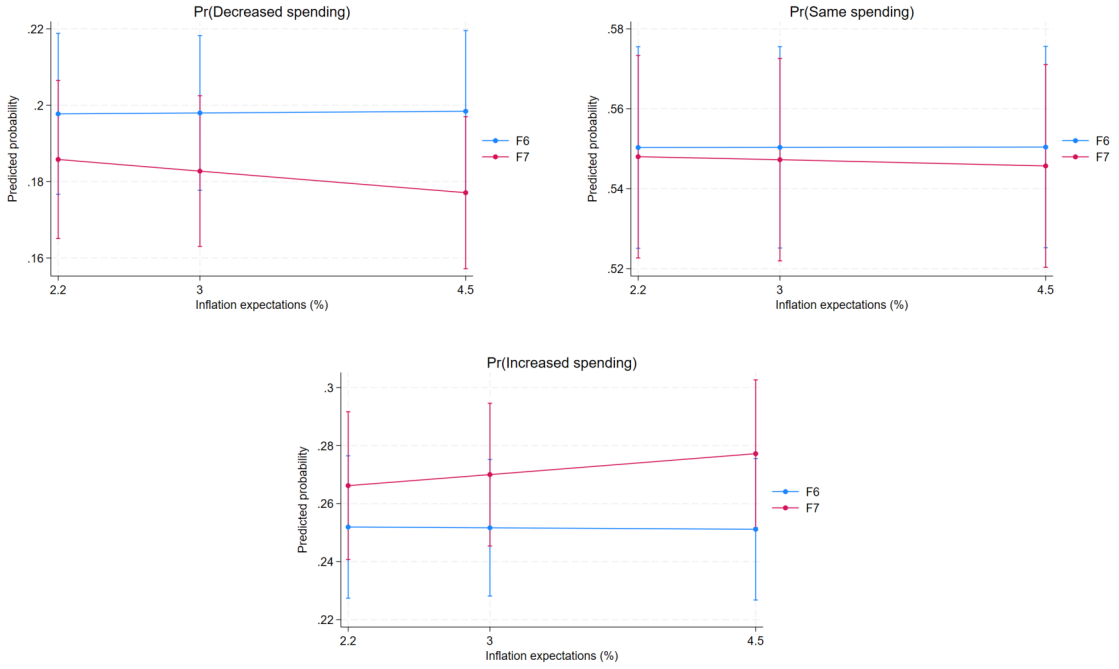
$$C_{im} = \beta_0 + \beta_1 E_i(\pi_{t+12}) + \beta_2 E_i(\pi_{t+12}) \times Treat_{im} + \beta_3 Treat_{im} + \beta_4' X_i + \epsilon_{im}, \quad \text{(B.15)}$$

where C_{im} is the three-category ordered response, $m \in \{\text{F3}, \text{F1}\}$, and $Treat_{im}$ equals one if the observation comes from F1 and zero if it comes from F3. Both items were administered to the full sample, so every respondent contributes two observations ($N = 7{,}196$) and the comparison is within-subject. The question order is fixed, with F1 preceding F3. Both items are three-category ordered items and enter in their original coding. Equation (B.15) is estimated as an ordered probit model with individual random effects.

B.4 Predicted Outcomes and Average Marginal Effects

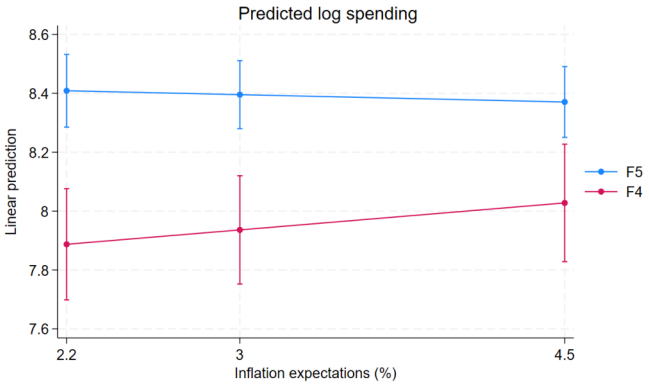
This section presents, for each hypothesis test, the predicted outcomes under the control and treatment questions at selected values of inflation expectations. Parallel lines indicate that the relationship between inflation expectations and reported spending does not depend on the framing, while vertical distance between the lines reflects differences in levels only.

Furthermore, this section presents average marginal effects (AMEs) for Hypothesis Test 1 to 7 (Tables B.5 and B.6).



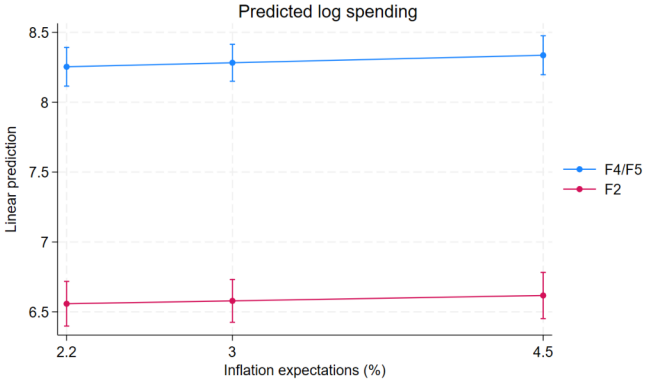
Notes: This figure reports average predicted probabilities of each response category for Hypothesis Test 1, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample. Panel A shows the probability of reporting decreased spending, Panel B of reporting the same spending, and Panel C of reporting increased spending, each separately for the individual-level (F6, control) and the household-level (F7, treatment) qualitative recall-based spending question. Hypothesis Test 1 examines whether the relationship between inflation expectations and reported durable spending depends on the reference unit (individual vs. household), using within-subject variation of respondents in multi-person households (N = 2,584), estimated with a random effects ordered probit model. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on standard errors clustered at the individual level.

Figure B.1: Predicted Probabilities by Reference Unit: Hypothesis Test 1



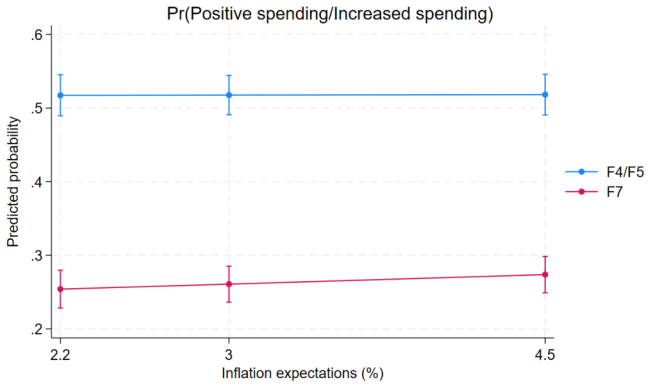
Notes: This figure reports average predicted log spending for Hypothesis Test 2, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for quantitative recall-based spending over the past 12 months (F5, control) and the past 3 months (F4, treatment). Hypothesis Test 2 examines whether the relationship between inflation expectations and reported durable spending depends on the time horizon of the consumption question, using a between-subjects design ($N = 840$), estimated by OLS. The difference in levels between the two lines reflects the different time horizons of the questions; the relationship with inflation expectations is captured by the slope of each line. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on robust (Huber-White) standard errors.

Figure B.2: Predicted Log Spending by Recall Horizon: Hypothesis Test 2



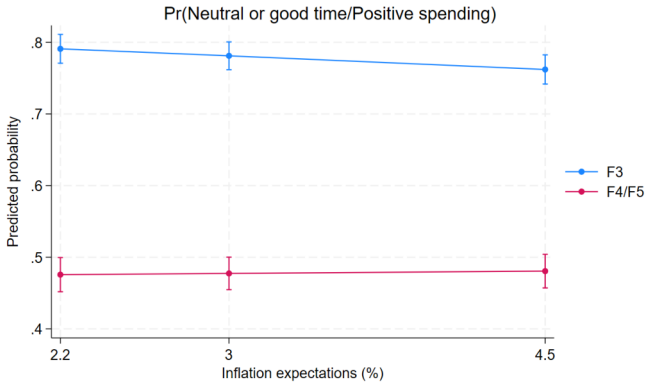
Notes: This figure reports average predicted log spending for Hypothesis Test 3, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for the pooled quantitative recall-based spending over the past 3 and 12 months (F4/F5, control) and over the past month (F2, treatment). Hypothesis Test 3 examines whether the relationship between inflation expectations and reported durable spending depends on the time horizon of the consumption question, using within-subject variation of respondents who reported strictly positive spending for both questions ($N = 886$), estimated with a random effects GLS model. The difference in levels between the two lines reflects the different time horizons of the questions; the relationship with inflation expectations is captured by the slope of each line. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on robust standard errors clustered at the individual level.

Figure B.3: Predicted Log Spending by Recall Horizon: Hypothesis Test 3



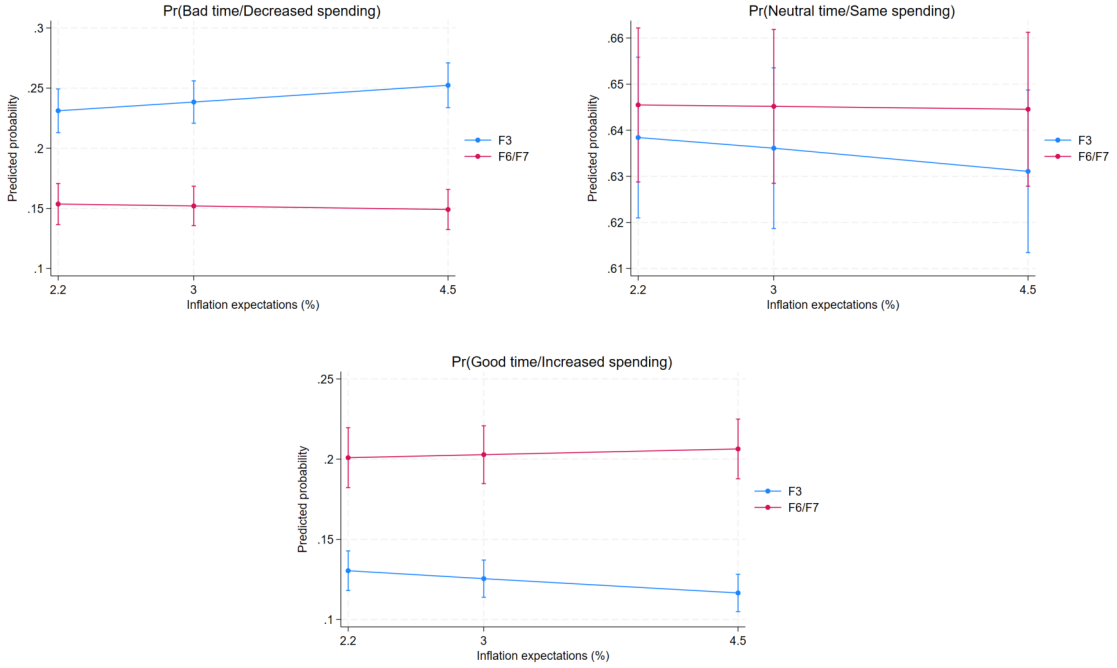
Notes: This figure reports average predicted probabilities for Hypothesis Test 4, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for quantitative (F4/F5, control) and qualitative (F7, treatment) recall-based spending. Hypothesis Test 4 examines whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question is quantitative or qualitative, comparing the two measures at the extensive margin in a between-subject comparison restricted to multi-person households ($N = 2{,}626$), estimated by logit. Quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item; qualitative recall-based spending (F7) is coded as a binary indicator for increased spending. The figure therefore shows the probability of reporting positive spending for F4/F5 and of reporting increased spending for F7; the difference in levels between the two lines reflects these different outcomes, and the relationship with inflation expectations is captured by the slope of each line. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on robust (Huber-White) standard errors.

Figure B.4: Predicted Probabilities by Question Type: Hypothesis Test 4



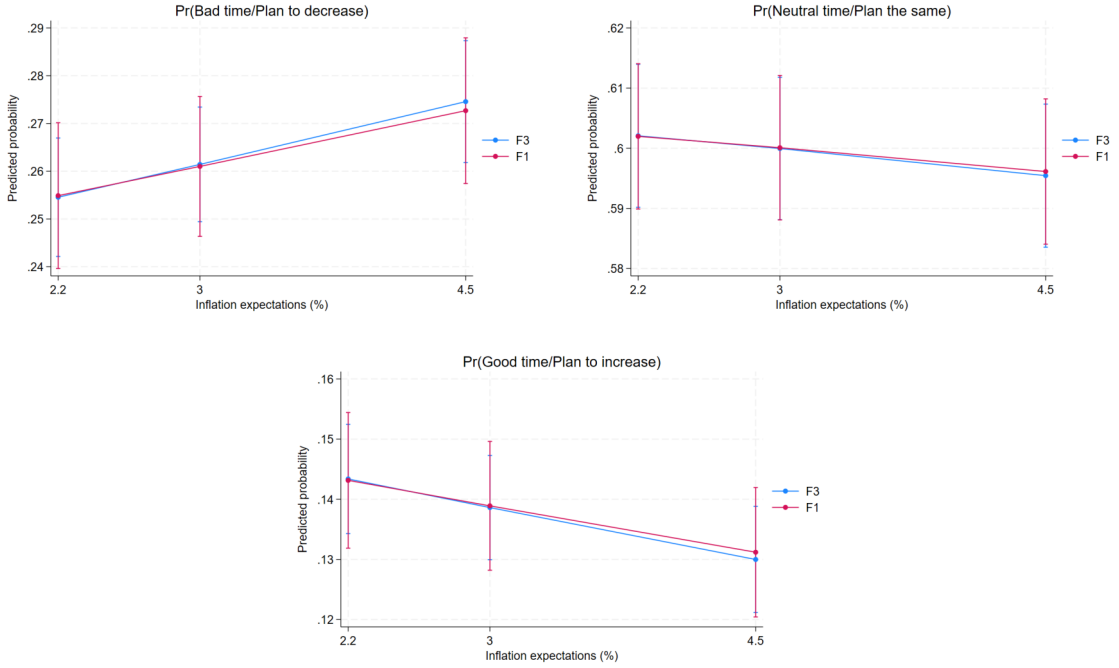
Notes: This figure reports average predicted probabilities for Hypothesis Test 5, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for attitude (F3, control) and quantitative recall-based spending (F4/F5, treatment). Hypothesis Test 5 examines whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or quantitative recall-based spending, comparing the two measures at the extensive margin in a within-subject comparison of all Group 1 respondents (N = 3,600), estimated by random-effects logit. Attitude (F3) is coded as a binary indicator for a positive or neutral spending attitude; quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item. The figure therefore shows the probability of reporting that it is a good or neither a good nor a bad time to buy durable goods for F3 and of reporting positive spending for F4/F5; the difference in levels between the two lines reflects these different outcomes, and the relationship with inflation expectations is captured by the slope of each line. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on standard errors clustered at the individual level.

Figure B.5: Predicted Probabilities by Question Type: Hypothesis Test 5



Notes: This figure reports average predicted probabilities of each response category for Hypothesis Test 6, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for attitude (F3, control) and qualitative recall-based spending (F6/F7, treatment). Panel A shows the probability of reporting that it is a bad time to buy durable goods (F3) or decreased spending (F6/F7), Panel B of a neutral time or the same spending, and Panel C of a good time or increased spending. Hypothesis Test 6 examines whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or qualitative recall-based spending, in a within-subject comparison of all Group 2 respondents ($N = 3{,}580$), estimated by random-effects ordered probit. Qualitative recall-based spending pools the individual- (F6) and household-level (F7) responses, taking the household-level response where both were answered. All items are three-category ordered items entering without transformation. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on standard errors clustered at the individual level.

Figure B.6: Predicted Probabilities by Question Type: Hypothesis Test 6



Notes: This figure reports average predicted probabilities of each response category for Hypothesis Test 7, based on regression specification (3.1), evaluated at inflation expectations of 2.2, 3 and 4.5 percent, the first quartile, the median and the third quartile in the full sample, separately for attitude (F3, control) and planned spending (F1, treatment). Panel A shows the probability of reporting that it is a bad time to buy durable goods (F3) or planning to decrease spending (F1), Panel B of a neutral time or planning the same spending, and Panel C of a good time or planning to increase spending. Hypothesis Test 7 examines whether the relationship between inflation expectations and reported durable spending depends on whether the consumption question measures an attitude or planned spending, in a within-subject comparison of the full sample (N = 7,196), estimated by random-effects ordered probit. Both items are three-category ordered items entering without transformation. Predictions are averaged over the observed values of all other covariates. 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. Vertical bars show 95% confidence intervals, computed with the delta method based on standard errors clustered at the individual level.

Figure B.7: Predicted Probabilities by Question Type: Hypothesis Test 7

Outcome	Control question (1)	Treatment question (2)	Difference (3)
Panel A: Reference unit framing (percentage points)			
$H1$: F6 (control) vs. F7 (treatment), RE ordered probit, $N = 2{,}584$			
Pr(Decreased spending)	0.03	−0.38	−0.41
	(0.27)	(0.24)	(0.24)
Pr(Same spending)	0.00	−0.10	−0.11
	(0.04)	(0.07)	(0.06)
Pr(Increased spending)	−0.03	0.48	0.51
	(0.31)	(0.31)	(0.29)
Panel B: Temporal framing (percent)			
$H2$: F5 (control) vs. F4 (treatment), OLS, $N = 840$			
Log spending	−1.66	6.11*	7.76*
	(1.87)	(2.66)	(3.24)
$H3$: F4/F5 (control) vs. F2 (treatment), RE GLS, $N = 886$			
Log spending	3.58	2.57	−1.01
	(1.98)	(2.42)	(2.37)

Notes: This table reports the relationship between inflation expectations and reported durable spending under the control and the treatment question for Hypothesis Tests 1–3, based on regression specification (3.1). Column 1 reports the effect of inflation expectations under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which corresponds to the interaction between inflation expectations and the treatment indicator. Panel A examines whether the relationship depends on the reference unit (individual vs. household), based on responses to individual- (F6) and household-level (F7) qualitative recall-based spending, estimated with a random effects ordered probit model. Entries in Panel A are AMEs of a 1pp increase in inflation expectations on the probability of each response category, in pp. Panel B assesses whether the time horizon of the consumption question affects the relationship. Hypothesis Test 2 compares the log of quantitative recall-based spending over the past 3 months (F4) and the past 12 months (F5) using a between-subjects design, estimated by OLS. Hypothesis Test 3 contrasts the log of quantitative recall-based spending over the past month (F2) with the pooled responses from F4 and F5 using within-subject variation, estimated with a random effects GLS model. Entries in Panel B are coefficients multiplied by 100 and can be read as the approximate percentage change in spending associated with a 1pp increase in inflation expectations. 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. Robust standard errors in parentheses, clustered at the individual level in random effects models; standard errors of marginal effects are computed with the delta method. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.5: Reference Unit and Temporal Framing Effects: Hypothesis Tests 1–3; ; Average Marginal Effects

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
$H4$: F4/F5 (control) vs. F7 (treatment), logit, $N = 2{,}626$			
Pr(Positive spending/Increased spending)	0.04	0.86**	0.82
	(0.38)	(0.31)	(0.49)
$H5$: F3 (control) vs. F4/F5 (treatment), RE logit, $N = 3{,}600$			
Pr(Neutral or good time/Positive spending)	−1.25***	0.22	1.46***
	(0.27)	(0.31)	(0.40)
$H6$: F3 (control) vs. F6/F7 (treatment), RE ordered probit, $N = 3{,}580$			
Pr(Bad time/Decreased spending)	0.92***	−0.20	−1.12***
	(0.24)	(0.19)	(0.30)
Pr(Neutral time/Same spending)	−0.32***	−0.04	0.28**
	(0.09)	(0.04)	(0.10)
Pr(Good time/Increased spending)	−0.60***	0.24	0.84**
	(0.16)	(0.23)	(0.27)
$H7$: F3 (control) vs. F1 (treatment), RE ordered probit, $N = 7{,}196$			
Pr(Bad time/Plan to decrease)	0.87***	0.77***	−0.10
	(0.17)	(0.20)	(0.22)
Pr(Neutral time/Plan the same)	−0.29***	−0.25***	0.03
	(0.06)	(0.07)	(0.08)
Pr(Good time/Plan to increase)	−0.58***	−0.52***	0.06
	(0.11)	(0.13)	(0.15)

Notes: This table reports AMEs of a 1pp increase in inflation expectations on the probability of each response category, in pp, for Hypothesis Tests 4–7 on question-type framing, based on regression specification (3.1). Column 1 reports the AME under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which corresponds to the interaction between inflation expectations and the treatment indicator on the probability scale. Hypothesis Test 4 compares quantitative (F4/F5) and qualitative (F7) recall-based spending at the extensive margin, in a between-subject comparison restricted to multi-person households, estimated by logit. Hypothesis Test 5 compares attitude (F3) with quantitative recall-based spending (F4/F5) at the extensive margin, in a within-subject comparison of all Group 1 respondents, estimated by random-effects logit. In both tests, quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item; qualitative recall-based spending (F7) is coded as a binary indicator for increased spending, and attitude (F3) as a binary indicator for a positive or neutral spending attitude. Hypothesis Test 6 compares attitude (F3) with qualitative recall-based spending (F6/F7), taking the household-level response (F7) where both were answered, and Hypothesis Test 7 compares attitude (F3) with planned spending (F1); both are within-subject comparisons of three-category ordered items entering without transformation, estimated by random-effects ordered probit. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method; they are robust (Huber-White) in Hypothesis Test 4 and clustered at the individual level in the random-effects models. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.6: Question-Type Framing Effects: Hypothesis Tests 4–7; Average Marginal Effects

B.5 Additional Regression Tables

This section presents additional regressions that go beyond the baseline estimations. Tables B.8, B.11 and B.12 disaggregate the pooled measures in Hypothesis Tests 3, 5 and 6, Tables B.7 and B.13 recode qualitative recall-based spending to reflect real changes, and Table B.10 (with the balance check in Table B.9) includes single-person households in Hypothesis Test 4.

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
$H1$: F6 (control) vs. F7 (treatment), RE logit, $N = 2{,}584$			
Pr(Real increase in spending)	0.70	1.16	0.45
	(0.69)	(0.78)	(0.31)

Notes: This table reports how the average predicted probability of reporting a real increase in spending changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for Hypothesis Test 1, based on regression specification (3.1), with qualitative recall-based spending recoded to reflect real changes: responses are classified as real increases when nominal spending rises, and as real decreases otherwise. Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the reference unit. Hypothesis Test 1 compares qualitative recall-based spending at the individual (F6) and household level (F7), using within-subject variation of Group 2 respondents in multi-person households, estimated by random-effects logit. 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. Standard errors in parentheses are obtained by the delta method and clustered at the individual level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.7: Hypothesis Test 1: Real Change in Consumption

Outcome	Control question (1)	Treatment question (2)	Difference (3)
Temporal framing (percent)			
H3a: F4 (control) vs. F2 (treatment), RE GLS, $N = 352$			
Log spending	13.82	14.50*	0.68
	(7.32)	(7.39)	(6.26)
H3b: F5 (control) vs. F2 (treatment), RE GLS, $N = 542$			
Log spending	2.60	1.94	−0.65
	(5.95)	(6.46)	(5.97)

Notes: This table reports how the average predicted outcome changes when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, under the control and the treatment question for Hypothesis Tests 3a and 3b, based on regression specification (3.1). Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the framing used. Both tests disaggregate Hypothesis Test 3 and assess whether the time horizon of the consumption question affects the relationship. Hypothesis Test 3a contrasts the log of quantitative recall-based spending over the past month (F2) with the log of spending over the past 3 months (F4); Hypothesis Test 3b contrasts F2 with the log of spending over the past 12 months (F5). Both use within-subject variation of respondents who reported strictly positive spending for both questions and are estimated with a random effects GLS model. Entries are changes in predicted log spending multiplied by 100 and can be read as the approximate percentage change in spending. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors clustered at the individual level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.8: Temporal Framing Effects by Recall Horizon: Hypothesis Tests 3a and 3b

	F4/F5	F6/F7	Test of mean difference
Mean age	54.69	55.00	$t(3599) = -0.5851$
	(16.17)	(16.09)	$p = 0.559$
Percentage no university	45.38%	44.31%	$\chi^2(1) = 0.4215$
			$p = 0.516$
Percentage low income	11.50%	12.44%	$\chi^2(1) = 0.7648$
			$p = 0.382$
Percentage female	39.91%	42.13%	$\chi^2(1) = 1.8341$
			$p = 0.176$
Percentage unemployed	1.49%	1.62%	$\chi^2(1) = 0.0930$
			$p = 0.760$
Percentage East Germany	17.08%	18.47%	$\chi^2(1) = 1.1897$
			$p = 0.275$
Percentage single-person households	26.31%	27.85%	$\chi^2(1) = 1.0716$
			$p = 0.301$
N	1,809	1,792	3,601

Notes: The table compares the two subsamples of the robustness check for Hypothesis Test 4 that includes single-person households, restricted to the respondents included in the estimation. F4/F5: Group 1 respondents, for whom quantitative recall-based spending (3 and 12 months) is coded as a binary indicator for positive expenses over the past 12 months; F6/F7: Group 2 respondents answering the qualitative recall-based question at the household level (F7) or, if F7 was not answered, mainly in single-person households, at the individual level (F6). Standard deviations in parentheses. The test of mean difference is a two-sample t-test for age and a Pearson χ^2-test for all other characteristics. Low income is a total monthly net household income of €1,999 or less; no university refers to the lower and medium education categories.

Table B.9: Balance Table: Sample F4/F5 and Sample F6/F7 (All Households)

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
$H4$: F4/F5 (control) vs. F6/F7 (treatment), logit, $N = 3{,}601$			
Pr(Positive spending/Increased spending)	−0.19	1.18*	1.37
	(0.72)	(0.57)	(0.91)

Notes: This table reports how the average predicted probability changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for Hypothesis Test 4, based on regression specification (3.1), including single-person households. Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the question type. Hypothesis Test 4 compares quantitative (F4/F5) and qualitative recall-based spending at the extensive margin, in a between-subject comparison of Group 1 and Group 2 respondents, estimated by logit. In contrast to the main specification, the sample is not restricted to multi-person households: qualitative recall-based spending is measured by the household-level response (F7) and, for respondents who did not answer F7 because they live in single-person households by the individual-level response (F6). Quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item; qualitative recall-based spending as a binary indicator for increased spending. The row label gives the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method and are robust (Huber-White). Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.10: Hypothesis Test 4: Including Single-Person Households

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
*H*5*a*: F3 (control) vs. F4 (treatment), RE logit, $N = 3{,}612$			
Pr(Neutral or good time/Positive spending)	-2.93^{***}	-0.07	2.86^{**}
	(0.64)	(0.63)	(0.87)
*H*5*b*: F3 (control) vs. F5 (treatment), RE logit, $N = 3{,}604$			
Pr(Neutral or good time/Positive spending)	-2.82^{***}	1.08	3.90^{***}
	(0.63)	(0.67)	(0.85)
*H*5*c*: F3 (control) vs. F2 (treatment), RE logit, $N = 7{,}098$			
Pr(Neutral or good time/Positive spending)	-2.68^{***}	-0.05	2.64^{***}
	(0.42)	(0.50)	(0.67)

Notes: This table reports how the average predicted probability changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for Hypothesis Tests 5a–5c, based on regression specification (3.1). Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the question type. The three tests disaggregate Hypothesis Test 5 by recall horizon and compare attitude (F3) with quantitative recall-based spending at the extensive margin: over the past 3 months (F4, Hypothesis Test 5a), over the past 12 months (F5, Hypothesis Test 5b) and over the past month (F2, Hypothesis Test 5c). All are within-subject comparisons estimated by random-effects logit. Attitude (F3) is coded as a binary indicator for a positive or neutral spending attitude. Quantitative recall-based spending is coded as a binary indicator for positive expenses over the respective horizon; for F4 and F5, it is taken from the screening item preceding these questions, so that respondents who reported no purchase are recorded as zero. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method and clustered at the individual level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.11: Question-Type Framing Effects by Recall Horizon: Hypothesis Tests 5a–5c

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
H6a: F3 (control) vs. F6 (treatment), RE ordered probit, $N = 3{,}578$			
Pr(Bad time/Decreased spending)	2.14^{***}	0.10	-2.03^{**}
	(0.55)	(0.47)	(0.65)
Pr(Neutral time/Same spending)	-0.75^{***}	0.01	0.76^{***}
	(0.21)	(0.05)	(0.22)
Pr(Good time/Increased spending)	-1.39^{***}	-0.11	1.27^{*}
	(0.36)	(0.52)	(0.58)
H6b: F3 (control) vs. F7 (treatment), RE ordered probit, $N = 2{,}586$			
Pr(Bad time/Decreased spending)	1.92^{**}	-0.99	-2.91^{**}
	(0.73)	(0.52)	(0.90)
Pr(Neutral time/Same spending)	-0.67^{*}	-0.35	0.32
	(0.27)	(0.20)	(0.33)
Pr(Good time/Increased spending)	-1.25^{**}	1.35	2.59^{**}
	(0.48)	(0.70)	(0.85)

Notes: This table reports how the average predicted probability of each response category changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for Hypothesis Tests 6a and 6b, based on regression specification (3.1). Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the question type. The two tests disaggregate Hypothesis Test 6 by reference unit and compare attitude (F3) with qualitative recall-based spending at the individual level (F6, Hypothesis Test 6a, all Group 2 respondents) and at the household level (F7, Hypothesis Test 6b, Group 2 respondents in multi-person households). Both are within-subject comparisons of three-category ordered items entering without transformation, estimated by random-effects ordered probit. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method and clustered at the individual level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.12: Question-Type Framing Effects by Reference Unit: Hypothesis Tests 6a and 6b

Outcome	Control question	Treatment question	Difference
	(1)	(2)	(3)
H6: F3 (control) vs. F6/F7 (treatment), RE logit, $N = 3{,}580$			
Pr(Neutral or good time/Real increase in spending)	-2.92^{***}	1.55^{**}	4.48^{***}
	(0.56)	(0.57)	(0.84)

Notes: This table reports how the average predicted probability changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for Hypothesis Test 6, based on regression specification (3.1), with qualitative recall-based spending recoded to reflect real changes: responses are classified as real increases when nominal spending rises, and as real decreases otherwise. Column 1 reports the change under the control question, Column 2 under the treatment question, and Column 3 the difference between the two, which tests whether the relationship between inflation expectations and reported durable spending depends on the question type. Hypothesis Test 6 compares attitude (F3) with the pooled responses of qualitative recall-based spending at the individual (F6) and household level (F7), taking the household-level response where both were answered, using within-subject variation of all Group 2 respondents, estimated by random-effects logit. Attitude (F3) is coded as a binary indicator for a positive or neutral spending attitude, and qualitative recall-based spending as a binary indicator for a real increase in spending. The row label gives the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are obtained by the delta method and clustered at the individual level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.13: Hypothesis Test 6: Real Change in Consumption

B.6 Robustness Checks

This section presents robustness checks of the baseline estimations. Tables B.14, B.15, B.16, B.17, B.18, B.19 and B.20 compare the results of each hypothesis test with separate regressions for each consumption measure. Tables B.21 and B.22 report the baseline results estimated using linear estimations. In addition, we re-estimate Hypothesis Test 2 (Tables B.23, B.24) and Hypothesis Tests 3, 3_a, and 3_b (Tables B.25, B.26) using Tobit and two-part models.

	F6 (control)		F7 (treatment)	
Outcome	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Pr(Decreased spending)	0.07	0.01	−0.87	−0.84
	(0.63)	(0.62)	(0.56)	(0.55)
Pr(Same spending)	0.01	0.00	−0.23	−0.22
	(0.10)	(0.10)	(0.15)	(0.15)
Pr(Increased spending)	−0.08	−0.02	1.10	1.05
	(0.72)	(0.72)	(0.71)	(0.69)
Estimator	RE ordered probit	Ordered probit	RE ordered probit	Ordered probit
N	2,584	1,292	2,584	1,292

Notes: This table reports how the average predicted probability of each response category changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 1: individual-level (F6) and household-level (F7) qualitative recall-based spending. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated with a random effects ordered probit model on both measures jointly (Columns 1 and 2 of Table 3, Panel A). Columns 2 and 4 report the changes from separate ordered probit regressions of each measure on inflation expectations and the controls. All estimations use the same sample of Group 2 respondents in multi-person households who answered both questions. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in the random effects model. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.14: Hypothesis Test 1: Comparison and Separate Regressions

Outcome	F5 (control)		F4 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Log spending	−3.81	−3.71	14.05*	14.59*
	(4.31)	(4.33)	(6.12)	(6.47)
Estimator	OLS	OLS	OLS	OLS
N	840	589	840	251

Notes: This table reports how average predicted log spending changes when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 2: quantitative recall-based spending over the past 12 months (F5) and the past 3 months (F4). Entries are changes in predicted log spending multiplied by 100 and can be read as approximate percentage changes in spending. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated by OLS on both measures jointly (Columns 1 and 2 of Table 3, Panel B). Columns 2 and 4 report the changes from separate OLS regressions of each measure on inflation expectations and the controls, estimated on the respondents who answered the respective question. Each respondent answered only one of the two questions, so the separate samples add up to the comparison sample. 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. Standard errors in parentheses are robust (Huber-White) and computed with the delta method. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.15: Hypothesis Test 2: Comparison and Separate Regressions

Outcome	F4/F5 (control)		F2 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Log spending	8.23	8.87	5.90	5.26
	(4.55)	(4.68)	(5.56)	(5.69)
Estimator	RE GLS	OLS	RE GLS	OLS
N	886	443	886	443

Notes: This table reports how average predicted log spending changes when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 3: the pooled quantitative recall-based spending over the past 3 and 12 months (F4/F5) and over the past month (F2). Entries are changes in predicted log spending multiplied by 100 and can be read as approximate percentage changes in spending. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated with a random effects GLS model on both measures jointly (Columns 1 and 2 of Table 3, Panel B). Columns 2 and 4 report the changes from separate OLS regressions of each measure on inflation expectations and the controls. All estimations use the same respondents, who reported strictly positive spending for both questions; each contributes two observations to the comparison and one to each separate regression. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in the random effects model. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.16: Hypothesis Test 3: Comparison and Separate Regressions

Outcome	F4/F5 (control)		F7 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Pr(Positive spending/Increased spending)	0.10	0.42	1.97**	1.86*
	(0.87)	(0.89)	(0.72)	(0.73)
Estimator	Logit	Logit	Logit	Logit
N	2,626	1,333	2,626	1,293

Notes: This table reports how the average predicted probability changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 4: quantitative (F4/F5) and qualitative (F7) recall-based spending at the extensive margin. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated by logit on both measures jointly (Columns 1 and 2 of Table 4). Columns 2 and 4 report the changes from separate logit regressions of each measure on inflation expectations and the controls, estimated on the respondents who answered the respective question. The sample is restricted to multi-person households; each respondent answered only one of the two questions, so the separate samples add up to the comparison sample. Quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item; qualitative recall-based spending (F7) is coded as a binary indicator for increased spending. The row label gives the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are robust (Huber-White) and computed with the delta method. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.17: Hypothesis Test 4: Comparison and Separate Regressions

Outcome	F3 (control)		F4/F5 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Pr(Neutral or good time/Positive spending)	−2.88***	−2.41***	0.50	0.05
	(0.63)	(0.62)	(0.72)	(0.73)
Estimator	RE logit	Logit	RE logit	Logit
N	3,600	1,800	3,600	1,800

Notes: This table reports how the average predicted probability changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 5: attitude (F3) and quantitative recall-based spending (F4/F5) at the extensive margin. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated by random-effects logit on both measures jointly (Columns 1 and 2 of Table 4). Columns 2 and 4 report the changes from separate logit regressions of each measure on inflation expectations and the controls. All estimations use the same Group 1 respondents; each contributes two observations to the comparison and one to each separate regression. Attitude (F3) is coded as a binary indicator for a positive or neutral spending attitude; quantitative recall-based spending is coded as a binary indicator for positive expenses over the past 12 months, with respondents who reported no purchase recorded as zero from the preceding screening item. The row label gives the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in the random-effects model. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.18: Hypothesis Test 5: Comparison and Separate Regressions

Outcome	F3 (control)		F6/F7 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Pr(Bad time/Decreased spending)	2.12***	1.79**	−0.45	−0.21
	(0.55)	(0.55)	(0.44)	(0.46)
Pr(Neutral time/Same spending)	−0.73***	−0.81**	−0.09	−0.03
	(0.20)	(0.26)	(0.09)	(0.07)
Pr(Good time/Increased spending)	−1.39***	−0.98**	0.55	0.25
	(0.36)	(0.30)	(0.53)	(0.52)
Estimator	RE ordered probit	Ordered probit	RE ordered probit	Ordered probit
N	3,580	1,790	3,580	1,790

Notes: This table reports how the average predicted probability of each response category changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 6: attitude (F3) and qualitative recall-based spending (F6/F7), taking the household-level response (F7) where both were answered. Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated with a random effects ordered probit model on both measures jointly (Columns 1 and 2 of Table 4). Columns 2 and 4 report the changes from separate ordered probit regressions of each measure on inflation expectations and the controls. All estimations use the same Group 2 respondents; each contributes two observations to the comparison and one to each separate regression. Both items are three-category ordered items entering without transformation. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in the random effects model. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.19: Hypothesis Test 6: Comparison and Separate Regressions

Outcome	F3 (control)		F1 (treatment)	
	Comparison	Separate	Comparison	Separate
	(1)	(2)	(3)	(4)
Pr(Bad time/Plan to decrease)	2.00***	1.80***	1.78***	1.94***
	(0.39)	(0.40)	(0.46)	(0.45)
Pr(Neutral time/Plan the same)	−0.66***	−0.82***	−0.58***	−0.47***
	(0.14)	(0.19)	(0.16)	(0.12)
Pr(Good time/Plan to increase)	−1.34***	−0.97***	−1.19***	−1.47***
	(0.26)	(0.22)	(0.31)	(0.34)
Estimator	RE ordered probit	Ordered probit	RE ordered probit	Ordered probit
N	7,196	3,598	7,196	3,598

Notes: This table reports how the average predicted probability of each response category changes, in pp, when inflation expectations increase from 2.2 to 4.5 percent, the interquartile range in the full sample, for the two consumption measures compared in Hypothesis Test 7: attitude (F3) and planned spending (F1). Columns 1 and 3 report the changes from the comparison in regression specification (3.1), estimated with a random effects ordered probit model on both measures jointly (Columns 1 and 2 of Table 4). Columns 2 and 4 report the changes from separate ordered probit regressions of each measure on inflation expectations and the controls. All estimations use the same respondents from the full sample; each contributes two observations to the comparison and one to each separate regression. Both items are three-category ordered items entering without transformation. Row labels give the response category of the control and the treatment question, separated by a slash. 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. Standard errors in parentheses are computed with the delta method, based on robust standard errors, clustered at the individual level in the random effects model. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.20: Hypothesis Test 7: Comparison and Separate Regressions

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 $E_{i,t}(\pi_{t+1})$	-0.0001	-0.0166	0.0358
	(0.0058)	(0.0187)	(0.0198)
$E_{i,t}(\pi_{t+1}) \times Treatment$	0.0081	0.0776*	-0.0101
	(0.0050)	(0.0324)	(0.0237)
$Treatment$	0.0058	-0.6923***	-1.6735***
	(0.0214)	(0.1526)	(0.1122)
Individual Random Effects	+		+
Constant	+	+	+
Controls	+	+	+
N	2,584	840	886
R^2	0.0325	0.0917	0.2687

Note: This table reports the results for Hypothesis Tests 1–3. Column 2 is estimated by OLS with robust standard errors; Columns 1 and 3 are estimated by random-effects GLS with standard errors clustered at the individual level, and the reported R^2 is the overall R^2. Column 1 (Hypothesis Test 1) compares qualitative recall-based spending at the individual (F6) and household level (F7) in a within-subject comparison of Group 2 respondents living in multi-person households, who answered both questions (1,292 respondents). Both outcomes are coded on a three-point scale (1 = decreased, 2 = same, 3 = increased). The treatment is F7. Columns 2 and 3 use the log of quantitative recall-based spending as the dependent variable. Column 2 (Hypothesis Test 2) compares 3-month (F4) and 12-month quantitative recall-based spending (F5) in a between-subject comparison. The treatment is F4. Column 3 (Hypothesis Test 3) compares 1-month quantitative recall-based spending (F2) with the pooled responses of 3-month (F4) and 12-month quantitative recall-based spending (F5) in a within-subject comparison, restricted to respondents reporting strictly positive spending in both questions (443 respondents). The treatment is F2. 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.21: Hypothesis Tests 1-3: OLS Estimation

Framing	Question Type			
	$H4$	$H5$	$H6$	$H7$
Measures	F4/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 $E_{i,t}(\pi_{t+1})$	0.0006	-0.0137***	-0.0150***	-0.0140***
	(0.0038)	(0.0030)	(0.0038)	(0.0026)
$E_{i,t}(\pi_{t+1}) \times Treatment$	0.0084	0.0156***	0.0194***	0.0015
	(0.0051)	(0.0041)	(0.0055)	(0.0035)
$Treatment$	-0.2811***	-0.3479***	0.1047***	-0.0042
	(0.0254)	(0.0200)	(0.0266)	(0.0182)
Individual Random Effects		+	+	+
Constant	+	+	+	+
Controls	+	+	+	+
N	2,626	3,600	3,580	7,196
R^2	0.0913	0.1243	0.0442	0.0338

Note: This table reports the results for Hypothesis Tests 4–7. Column 1 is estimated by OLS with robust standard errors; Columns 2–4 are estimated by random-effects GLS with standard errors clustered at the individual level, and the reported R^2 is the overall R^2. Column 1 (Hypothesis Test 4) compares quantitative recall-based spending (F4/F5) with qualitative recall-based spending at the household level (F7) in a between-subject comparison restricted to multi-person households. F4/F5 is coded as a binary indicator for positive expenses over the past 12 months, while F7 is coded as a binary indicator for increased spending. The treatment is F7. Columns 2–4 are within-subject comparisons. Column 2 (Hypothesis Test 5) compares attitude (F3) with quantitative recall-based spending (F4/F5) for Group 1 respondents. F3 is coded as a binary indicator for a positive or neutral spending attitude, and F4/F5 as a binary indicator for positive expenses over the past 12 months. The treatment is F4/F5. Column 3 (Hypothesis Test 6) compares attitude (F3) with qualitative recall-based spending at the individual (F6) or household level (F7) for Group 2 respondents. The treatment is F6/F7. Column 4 (Hypothesis Test 7) compares attitude (F3) with planned spending (F1). The treatment is F1. In Columns 3 and 4, all outcomes are coded on a three-point scale (1 = bad/decreased, 2 = neutral/same, 3 = good/increased). 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.22: Hypothesis Tests 4-7: OLS Estimation

	$H2$
	Tobit
	(1)
Inflation Expectations $E_{i,t}(\pi_{t+1})$	11.34
	(145.22)
$E_{i,t}(\pi_{t+1}) \times Treatment$	-162.07
	(248.93)
$Treatment$	-14072.07***
	(1238.85)
Constant	+
Controls	+
N	2,787
Pseudo R^2	0.0173

Note: The table reports the results for Hypothesis 2, which tests whether the effect of expected inflation varies with the time frame of consumption by comparing responses to quantitative recall-based spending over the past 3 (F4) and 12 months (F5). The treatment is F4. The estimation method is Tobit. 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.23: Hypothesis Test 2 - Tobit Estimation

	$H2$	
	Logit	OLS
	(1)	(2)
Inflation Expectations $E_{i,t}(\pi_{t+1})$	-0.0054	0.0093
	(0.0147)	(0.0152)
$E_{i,t}(\pi_{t+1}) \times Treatment$	-0.0243	0.0517
	(0.0253)	(0.0321)
$Treatment$	-1.4731***	-0.4605**
	(0.1224)	(0.1471)
Constant	+	+
Controls	+	+
N	2,787	1,091
Pseudo R^2	0.1309	0.0499

Note: The table reports the results for Hypothesis 2, which tests whether the effect of expected inflation varies with the time frame of consumption by comparing responses to quantitative recall-based spending over the past 3 (F4) and 12 months (F5). The estimation follows a two-part model: the first part is estimated using a logit model for the probability of positive spending, and the second part uses a log-linear specification for the conditional amount of spending. 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.24: Hypothesis Test 2 - Two-Part Estimation

	$H3$	$H3_a$	$H3_b$
	(1)	(2)	(3)
Inflation Expectations $E_{i,t}(\pi_{t+12})$	45.34	−137.16	56.86
	(134.62)	(175.24)	(172.96)
$E_{i,t}(\pi_{t+12}) \times Treatment$	−206.92	−147.05	−176.22
	(170.07)	(146.09)	(233.31)
$Treatment$	−12,139.11***	−4,379.02***	−15,935.68***
	(820.61)	(670.66)	(1,149.20)
Constant	+	+	+
Controls	+	+	+
Individual random effects	+	+	+
N	3,082	2,436	2,584
Individuals	1,541	1,218	1,292
Pseudo R^2	0.0191	0.0138	0.0246

Notes: The table reports the results for Hypothesis Tests 3, 3_a and 3_b, which test whether the relationship between inflation expectations and reported durable spending depends on the time horizon of the consumption question, comparing quantitative recall-based spending over the past month (F2) with spending over the past 3 and 12 months pooled (F4/F5, Column 1), over the past 3 months (F4, Column 2) and over the past 12 months (F5, Column 3). The treatment is F2. Zero spending is included, and the estimation method is a random effects Tobit model. The sample is restricted to individuals who contribute an observation for both questions, so that all comparisons are within-subject. 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. Pseudo R^2 is McFadden's R^2 relative to a model with a constant and individual random effects only. Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.25: Hypothesis Test 3 - Tobit Estimation

	$H3$		$H3_a$		$H3_b$	
	Logit	OLS	Logit	OLS	Logit	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Inflation Expectations $E_{i,t}(\pi_{t+12})$	0.0001	0.0358	−0.0874	0.0601	0.0574	0.0113
	(0.0905)	(0.0231)	(0.0651)	(0.0371)	(0.0860)	(0.0281)
$E_{i,t}(\pi_{t+12}) \times Treatment$	−0.1394	−0.0101	−0.0317	0.0030	−0.1441	−0.0028
	(0.0912)	(0.0250)	(0.1125)	(0.0279)	(0.1041)	(0.0322)
$Treatment$	−6.0208***	−1.6735***	−3.7514***	−0.7257***	−6.6056***	−2.3610***
	(0.6452)	(0.1140)	(0.5225)	(0.1266)	(0.7078)	(0.1467)
Constant	+	+	+	+	+	+
Controls	+	+	+	+	+	+
Individual random effects	+	+	+	+	+	+
N	3,082	886	2,436	352	2,584	542
Individuals	1,541	443	1,218	176	1,292	271
(Pseudo) R^2	0.1722	0.2687	0.0884	0.1539	0.1835	0.4458

Notes: The table reports the results for Hypothesis Tests 3, 3_a and 3_b, which test whether the relationship between inflation expectations and reported durable spending depends on the time horizon of the consumption question, comparing quantitative recall-based spending over the past month (F2) with spending over the past 3 and 12 months pooled (F4/F5, Columns 1–2), over the past 3 months (F4, Columns 3–4) and over the past 12 months (F5, Columns 5–6). The treatment is F2. The estimation follows a two-part random effects model: the first part is a logit model for the probability of positive spending, estimated on all individuals who contribute an observation for both questions; the second part is a log-linear model for the spending amount, estimated on individuals who report positive spending for both questions. Both parts are therefore within-subject comparisons. 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. For the logit models, pseudo R^2 is McFadden's R^2 relative to a model with a constant and individual random effects only; for the log-linear models, the overall R^2 is reported. Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.26: Hypothesis Test 3 - Two-Part Estimation

C Validation Exercises I–II

C.1 Validation Exercise I: Aggregate Consumption Benchmarking

This section outlines the methodology used to compute the correlations between survey-based measures of consumption and aggregate consumption. In our dataset, repeated observations are available for planned spending (F1) and the 1-month quantitative recall-based measure (F2), while the remaining survey items were collected in a single wave. To supplement our analysis, we additionally draw on the Michigan Survey, which provides a time series of the attitudinal measure (University of Michigan, Survey Research Center, 2025).

Monthly Spending. We estimate the correlation between average reported quantitative recall-based spending on durable goods and aggregate consumption (International Monetary Fund, 2025; Real Final Consumption Expenditure for Germany [NCRSAXDCDEQ], retrieved from FRED). The current spending variable, elicited at time t, reflects individual spending over the preceding month, i.e., from $t-1$ to t. To align with the quarterly frequency of the national accounts data, we compute the quarterly average of current quantitative recall-based spending at period q and match it to aggregate consumption in the same quarter. The resulting correlation coefficient is 0.84, indicating that self-reported durable goods spending closely tracks aggregate consumption dynamics.

Planned Spending. We estimate the correlation between the share of individuals reporting plans to increase spending and subsequent aggregate consumption growth, measured by the quarterly real final consumption expenditure for Germany (International Monetary Fund, 2025; Real Final Consumption Expenditure for Germany [NCRSAXDCDEQ], retrieved from FRED). Since aggregate consumption is reported at a quarterly frequency, we average the share of respondents indicating higher planned spending within each quarter q and match this to consumption growth in the corresponding quarter one year ahead, i.e., period $q+4$. The resulting correlation coefficient is 0.2597, suggesting that planned spending is positively associated with realized consumption one year later, although the strength of this relationship is moderate.

Attitude. The Michigan Survey data contains monthly observations on attitude towards durable consumption at the individual level for the time period 1978 to 2025. The attitude is elicited using the following survey question:

1. *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

1 - Good

2 - Pro-con

3 - Bad

We compute the share of individuals who report that it is a good time to purchase durable goods and correlate this measure with the year-over-year percent 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).

We consider two timing conventions for the correlation analysis. First, we match the average attitude in period t with the year-over-year percentage change in durable goods consumption at the same time, i.e., from $t-12$ to t. For example, consumer attitude in January 2015 is matched with the percentage change in consumption between January 2014 and January 2015. Second, we lead the consumption measure by 12 months and correlate average attitude at time t with the year-over-year change from t to $t+12$. In this case, attitude in January 2015 is matched with the change in consumption between January 2015 and January 2016.

The contemporaneous correlation yields a coefficient of 0.3709, consistent with the magnitudes reported in Bachmann et al. (2015), Andrade et al. (2023), and Marenčák (2023). In contrast, the forward-looking correlation yields a coefficient of 0.0720, suggesting that positive attitude does not reliably translate into higher realized consumption one year later.

C.2 Validation Exercise II: Confirmatory Factor Analysis

Model / Panel	Indicators included	Standardized loadings (λ)	Interpretation
Panel A: Attitude + Quantitative Realized Spending (1m, 3m)			
	Attitude	0.17	Very weak link
	Past month's spending	0.74	Moderate link
	Past 3 months' spending	0.95	Very strong link
			Attitude does not cohere with realized spending.
Panel B: Attitude + Quantitative Realized Spending (1m, 12m)			
	Attitude	0.61	Moderate link
	Past month's spending	0.26	Weak link
	Past 12 months' spending	0.17	Weak link
			Distinct constructs: realized spending weakly related to attitude.
Panel C: Attitude + Qualitative Realized Spending (Individual vs. Household)			
	Attitude	Non-convergent	
	Change in spending (individual level)		
	Change in spending (household level)		*Heywood case: conceptual mismatch.*

Note: The table summarizes three confirmatory factor analyses. Panel A and Panel B combine respondents' attitude with quantitative measures of realized spending over different horizons (past 1, 3, or 12 months) to test whether attitudes and realized spending can be represented by a single latent construct. Panel C includes attitude together with qualitative realized spending measures (changes in individual and household spending). Standardized loadings (λ) from one-factor confirmatory factor analyses estimated using WLSMV. "Heywood" indicates a negative residual variance or non-convergence. Models with $df = 0$ reproduce observed covariances exactly; fit indices are omitted because they are uninformative for just-identified models. The results show that willingness-to-spend attitudes cluster with forward-looking or planning-type variables, but not with realized spending behavior.

Table C.1: Confirmatory Factor Analyses of Attitude, Planned, and Realized Spending Measures

D Replications

This section explains the replications of the papers by Andrade et al. (2023); Ichiue and Nishiguchi (2015) and Dräger and Nghiem (2021), presents the results using our data and compares them to the original results.

D.1 Replication of Dräger and Nghiem (2021)

Evaluating a new survey on German households, Dräger and Nghiem (2021) test whether individual consumption spending decisions are formed according to a Euler equation model. They find that consumers are more likely to increase current spending if they plan to increase spending in the future and if they expect higher inflation.

Survey Questions. For their analysis, they employ the following questions for inflation expectations:

1. *How do you think prices in general have developed over the past 12 months? They have...*

   1 - increased strongly

   2 - increased moderately

   3 - increased slightly

   4 - stayed about the same

   5 - fallen

2. *By what percentage do you think prices in general will increase or decrease on average over the next 12 months?*

For consumption measures, they employ the following measures:

3. *How would you say your total expenditures in the past 12 months compare to an average year in the past? They were...*

   1 - considerably higher

   2 - about the same

   3 - considerably lower

4. *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?*

1 - Now is a good time

2 - Neither a good nor a bad time

3 - Now is a bad time

5. *How would you say your total expenditures will be in the next 12 months compared to an average year in the past? They will be...*

1 - considerably higher

2 - about the same

3 - considerably lower

The third question corresponds to our questions on qualitative recall-based spending at the individual (F6) and household level (F7), the fourth to attitude (F3), and the fifth to planned spending (F1). However, while our measures focus exclusively on durable consumption, Dräger and Nghiem (2021) consider both overall and durable consumption.

Specifications. Dräger and Nghiem (2021) estimate ordered probit models in the following regression setup:

$$Pr(c_{it}^{current} = 3) = Pr(\kappa_2 < \beta_1 c_{it}^e + \beta_2 i_{it}^e + \beta_3 \pi_{it}^e + X_{it}^{controls'}\Gamma + u_{it} \leq \kappa_3) \tag{D.1}$$

where $\kappa_{2,3}$ denote the estimated cut-points between the three answer categories for qualitatively measured current consumption. c_{it}^e measures consumers' reported qualitative measure of expected spending in the next twelve months and i_{it}^e and π_{it}^e are individual expectations reported in t on nominal interest rates and the inflation rate over the next twelve months. X_{it} includes demographic variables (age, squared age, income, employment status, education, and risk aversion), and further macroeconomic expectations, namely, expectations on the general economic situation and on employment. We replicate the estimation in the same way, without controlling for risk aversion, and with inflation instead of price expectations.

Results. Dräger and Nghiem (2021) find that higher inflation expectations are associated with stronger increases in past overall spending and lower willingness to purchase durable goods. Our results differ with respect to realized consumption, likely reflecting our focus on durable spending versus their measure of overall consumption (Table D.2). For attitudes toward purchasing, the findings are broadly consistent, except for a significant positive estimate for planned consumption in our data, whereas their estimate remains insignificant (Table D.1).

	Attitude					
	Our Results			Original Results		
	(1)	(2)	(3)	(4)	(5)	(6)
C^e_{it}	0.0366***	0.0360***	0.0386***	0.029	0.034	0.065
	(0.0043)	(0.0043)	(0.0073)	(0.038)	(0.036)	(0.073)
$i^e_{qual,it}$	0.0007	0.0005		-0.006	-0.009	
	(0.0038)	(0.0038)		(0.036)	(0.039)	
$i^e_{quant,it}$			0.0014			0.010
			(0.0038)			(0.037)
$\pi^e_{qual,it}$	-0.0070*			-0.022		
	(0.0033)			(0.027)		
$\pi^e_{quant,it}$		-0.0026**	-0.0028*		-0.041**	-0.037***
		(0.0009)	(0.0012)		(0.012)	(0.013)
Constant	+	+	+	+	+	+
Controls	+	+	+	+	+	+
N	3,597	3,597	1,207	271	243	149
Pseudo R^2	0.0578	0.0590	0.0834	0.129	0.201	0.300

Note: The table reports marginal effects from ordered probit estimations. The dependent variable is the individual's attitude toward spending. The first three columns are based on our results, Column 4-6 are based on the results by Dräger and Nghiem (2021). Short-term inflation expectations are measured both qualitatively and quantitatively and are truncated at ±30. In addition, the models include expectations regarding interest rates on savings accountsmeasured both qualitatively and quantitativelyas well as planned spending. The quantitative interest rate is truncated at 10%. The analysis is based on a single wave of survey data. All specifications include controls for expected economic growth (binary indicator), expected unemployment (binary indicator) and a standard set of sociodemographic characteristics: age, age squared, gender, household income, education, employment status, a dummy for residence in East Germany, household size, and city size. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.1: Replication of Dräger and Nghiem (2021) - Attitude

	Individual-Level Qualitative Recall-Based Spending					
	Our Results			Original Results		
	(1)	(2)	(3)	(4)	(5)	(6)
C^e_{it}	-0.0510***	-0.0516***	-0.0847***	0.151*	0.153*	0.103
	(0.0126)	(0.0126)	(0.0226)	(0.062)	(0.076)	(0.096)
$i^e_{qual,it}$	0.0173	0.0172		-0.016	-0.021	
	(0.0105)	(0.0105)		(0.042)	(0.047)	
$i^e_{quant,it}$			0.0041			-0.027
			(0.0113)			(0.062)
$\pi^e_{qual,it}$	-0.0111			0.063*		
	(0.0093)			(0.025)		
$\pi^e_{quant,it}$		-0.0024	-0.0071		0.011*	0.021
		(0.0023)	(0.0040)		(0.006)	(0.014)
Constant	+	+	+	+	+	+
Controls	+	+	+	+	+	+
N	1,790	1,790	607	278	246	150
Pseudo R^2	0.0216	0.0215	0.0354	0.130	0.119	0.181

Note: The table reports marginal effects from ordered probit estimations. The dependent variable is the individual-level qualitative recall-based spending. The first three columns are based on our results, Column 4-6 are based on the results by Dräger and Nghiem (2021). Short-term inflation expectations are measured both qualitatively and quantitatively and are truncated at ±30. In addition, the models include expectations regarding interest rates on savings accountsmeasured both qualitatively and quantitativelyas well as planned spending. The quantitative interest rate is truncated at 10%. The analysis is based on a single wave of survey data. All specifications include controls for expected economic growth (binary indicator), expected unemployment (binary indicator) and a standard set of sociodemographic characteristics: age, age squared, gender, household income, education, employment status, a dummy for residence in East Germany, household size, and city size. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.2: Replication of Dräger and Nghiem (2021) - Change in Past Spending

D.2 Replication of Andrade et al. (2023)

Using survey data on French households over the time period of 2004 to 2018, Andrade et al. (2023) show that households expecting stable prices have a lower propensity to buy durable goods than those expecting positive inflation. In contrast, for households expecting positive inflation, variation in expectations does not translate into significant differences in durable consumption.

Survey Questions. For their analysis, Andrade et al. (2023) employ the following questions for inflation expectations:

1. *In comparison with the past 12 months, how do you expect consumer prices will develop in the next 12 months? They will ...*

   1 - Increase more rapidly

   2 - Increase at the same rate

   3 - Increase at a slower rate

   4 - Stay about the same

   5 - Fall

2. *By how many percents do you think consumer prices will go up/down over the next 12 months?*

   Consumer prices will increase/decrease by XX.X%

The survey design is such that the quantitative question is not asked to individuals who previously answered that prices "stay about the same". Therefore, Andrade et al. (2023) impute 0% to the quantitative inflation rate that these households expect. It is important to note that our data include only inflation expectations while their survey elicits price expectations.

For consumption, they employ the following two survey questions:

3. *Have you made any major purchase over the last 12 months? (such as furniture, household appliances, electronic or computer equipment, etc.)*

   1 - Yes

   2 - No

4. *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.)?*

   1 - Yes, now is the right time.

   2 - It is neither the right time nor the wrong time.

   3 - No, it is the wrong time.

The third question aligns with our survey question on quantitative recall-based spending over the past 12 months (F5). The fourth question aligns with our question on attitude (F3).

Specification. Andrade et al. (2023) consider the decision to buy durable goods between $t-1$ and t, $b_{i,t}$, a binary process that follows:

$$b_{i,t} = \begin{cases} 1 & if z^*_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{D.2}$$

where $z^*_{i,t}$ is a latent variable which evolves according to

$$z^*_{i,t} = \beta_0 + \beta_1 \pi^e_{i,t|t+1} + \gamma X_{i,t} + \lambda_t + \mu z_i + \epsilon_{i,t} \tag{D.3}$$

where $\pi^e_{i,t|t+1}$ is the expected inflation formed by household i at date t for the next year. X_{it} is a set of controls for other individuals' beliefs such as perceived inflation, planned spending, assessment about their own financial situation, and whether they think it is the right time to save. λ_t are time fixed-effects and z_i is a set of fixed household characteristics such as age, compositions of the household, occupation, income, working regime, education, gender, region and city size. We employ the same controls, except for perceived inflation, the assessment about their own financial situation and whether they think it is the right time to save as we either have insufficient data on these variables or none at all.

Methodology. Andrade et al. (2023) decompose inflation expectations into an extensive margin—households expecting prices to remain stable or increase—and an intensive margin—the expected rate of inflation, conditional expecting positive inflation. As we do not have data on price expectations, we use our inflation expectations data to create an artificial variable of the extensive margin of price expectations. The variable equals one if a household expects positive inflation and zero if they expect zero inflation.

Results. Andrade et al. (2023) show that households expecting stable prices are significantly less likely to have purchased durable goods in the past year than those expecting

inflation, with the extensive margin driving the link between inflation expectations and spending. Our results for durable goods are directionally consistent but statistically insignificant, likely due to smaller sample size (Table D.3). In contrast, our findings on consumption attitudes differ from their French results but align with their robustness check using German data (Table D.4). For attitudes, the intensive margin appears more relevant in our data, whereas Andrade et al. (2023) emphasize the extensive margin. Nonetheless, we share the finding—both in terms of direction and significance—that inflation expectations matter for households' attitude towards durable consumption.

	Spending over the Past 12 Months					
	Our Results			Original Results		
Independent Var.:	All π^e	Intensive	Extensive	All π^e	Intensive	Extensive
	(1)	(2)	(3)	(4)	(5)	(6)
π^e	-0.0011	-0.0061		0.005	-0.058	
	(0.0032)	(0.0040)		(0.028)	(0.0039)	
$1(\pi^e > 0)$			0.0572			1.047***
			(0.0350)			(0.337)
Constant	+	+	+	+	+	+
Controls	+	+	+	+	+	+
N	1,801	1,591	1,801	135,337	90,765	135,337

Note: The table reports marginal effects from logit estimations examining the relationship between expected inflation and realized spending over the past 12 months. The first three columns are based on our data, the Columns 4-6 are the original results from Andrade et al. (2023). Spending is measured as a binary indicator for positive expenditures. Short-term inflation expectations are measured quantitatively and truncated at ±30. In Columns 2 and 5, the independent variable includes only strictly positive values of expected inflation (intensive margin). In Columns 3 and 6, expected inflation is captured by a binary indicator equal to one for strictly positive expectations. The analysis is based on a single survey wave. All models include controls for expected economic growth (binary indicator), planned spending 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.3: Replication of Andrade et al. (2023) - Past Spending

D.3 Replication of Ichiue and Nishiguchi (2015)

Using survey data on Japanese households from 2006 to 2013, Ichiue and Nishiguchi (2015) find evidence supporting the standard theoretical model: consumers with higher expected inflation tend to report higher real household spending compared to one year ago, but anticipate reducing it in the future.

Survey Questions. For their analysis, Ichiue and Nishiguchi (2015) employ qualitative questions for expected and perceived inflation:

1. *What is your outlook for prices 1 year from now?*

    - Will go up significantly

	Attitude					
	Our Results			Original Results		
Independent Var.:	All π^e	Intensive	Extensive	All π^e	Intensive	Extensive
	(1)	(2)	(3)	(4)	(5)	(6)
π^e	-0.0037***	-0.0062***		0.006	-0.028	
	(0.0009)	(0.0013)		(0.016)	(0.020)	
$1(\pi^e > 0)$			0.0058			0.620**
			(0.0091)			(0.187)
Constant	+	+	+	+	+	+
Controls	+	+	+	+	+	+
N	3,598	3,210	3,598	132,897	89,346	132,897

Note: The table reports marginal effects from ordered probit estimations assessing the impact of expected inflation on attitudes toward spending. The first three columns are based on our data, the Columns 4-6 are the original results from Andrade et al. (2023). Short-term inflation expectations are measured quantitatively and truncated at ±30. In Columns 2 and 5, the independent variable includes only strictly positive values of expected inflation (intensive margin). In Columns 3 and 6, expected inflation is captured by a binary indicator equal to one for strictly positive expectations. The analysis is based on a single survey wave. All models include controls for expected economic growth (binary indicator), planned spending 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. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.4: Replication of Andrade et al. (2023) - Attitude

- Will go up slightly
- Will remain almost unchanged
- Will go down slightly
- Will go down significantly

2. *How do you think prices have changed compared with 1 year ago?*

- Have gone up significantly
- Have gone up slightly
- Have remained almost unchanged
- Have gone down slightly
- Have gone down significantly

For consumption, they employ the following two survey questions:

3. *How does your household plan to change its spending within the next 12 months?*

1 - Will increase

2 - Will neither increase nor decrease

3 - Will decrease

4. *How has your household changed its spending compared with 1 year ago?*

1 - Has increased

2 - Has neither increased nor decreased

3 - Has decreased

The third question aligns with our question on planned spending (F1). The fourth question aligns with our question on qualitative recall-based spending at the household level (F7).

Methodology. Ichiue and Nishiguchi (2015) estimate the effect of expected inflation on changes in real durable consumption. To do so, they construct an artificial measure of real spending. For planned spending, they grade question 3 with +1, 0, −1 for "increase", "neither" and "decrease". They also assign for questions 1, "up significantly" to −1, "up slightly" to −0.5, and so on. The responses to the artificial question about real spending are then defined as "increase significantly" if the total points are 2, "increase" if the total points are in the range of 1 to 1.5, and so on. The composite real spending measure is then categorized into ordered response levels depending on the total points. An analogous procedure is applied for past real spending, combining questions 2 and 4.

Specification. Ichiue and Nishiguchi (2015) assume that there is an underlying unobserved variable of the expected change in real spending which can be represented as follows:

$$y_i^* = \beta' X_i + \epsilon_i \tag{D.4}$$

where X_i is a vector of dummies for expected inflation and dummies of controls variables, including outlook for economic conditions, expected income, sex, age, employment status, income, household composition and time dummies. The specification is estimated using ordered probit.

In our replication, we focus on planned spending as the sole dependent variable, since we lack sufficient data on inflation perceptions for the construction of the past real spending variable. We otherwise follow their methodology and control variables, excluding only income expectations. Because our data capture inflation expectations directly (rather than price expectations), we map expected inflation into price expectations for consistency: expectations below −4 are coded as "prices decrease significantly", between −4 and −0.5 as "prices decrease slightly", between −0.5 and 0.5 as "prices remain stable", and so on.

Results. Ichiue and Nishiguchi (2015) show that higher inflation expectations are associated with decreased planned spending for the future. Our results are largely consistent with their findings in terms of direction, although not all estimates are statistically

significant—possibly due to limited sample size (Table D.5).

	Real Change in Planned Spending	
	Our Results	Original Results
	(1)	(2)
Prices 1 year from now		
Up significantly	-0.0024**	-2.43**
	(0.0009)	(0.04)
Up slightly	-0.0012	-0.12**
	(0.0009)	(0.01)
Down slightly	0.0000	1.28**
	(0.0007)	(0.03)
Down significantly	-0.0014*	2.08**
	(0.0007)	(0.08)
Constant	+	+
Controls	+	+
N	3,601	56,267
Pseudo R^2	0.0498	0.3112

Note: The table reports the marginal effects of an ordered probit estimation. The dependent variables capture the real change in planned spending. The first column is based on our data, and the second column is based on the results by Ichiue and Nishiguchi (2015). The key independent variables are categorical indicators based on short-term inflation expectations: (i) significantly increasing prices (expected inflation > 4%), (ii) slightly increasing prices (expected inflation between 0.5% and 4%), (iii) slightly decreasing prices (expected inflation between -0.5% and 4%), and (iv) significantly decreasing prices (expected inflation < 4%). The reference category is stable prices (expected inflation between -0.5% and 0.5%). The analysis is based on a single wave of data. All specifications include a control for expected economic growth (binary indicator), as well as sociodemographic controls: age, age squared, gender, household income, education, employment status, an East Germany dummy, household size, and city size. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D.5: Replication of Ichiue and Nishiguchi (2015) - Real Planned Spending