NICHOLAS S. SOULELES

Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys

The household data underlying the Michigan Index of Consumer Sentiment are used to test the rationality of consumer expectations and their usefulness in forecasting expenditure. The results can be interpreted as characterizing the shocks that hit different types of households over time. Expectations are found to be biased and inefficient, at least ex post. People underestimated the disinflation of the early 1980s and the severity of recent business cycles. People's forecast errors are also systematically correlated with their demographic characteristics, in part because of time-varying, group-level shocks. Further, sentiment helps forecast consumption growth. Some of this rejection of the permanent income hypothesis is due to the systematic demographic components in forecast errors.

JEL code: E21

Keywords: consumer sentiment, consumer confidence, permanent income hypothesis, excess sensitivity, precautionary saving, subjective expectations, rational expectations, forecast errors, shocks, unobserved heterogeneity.

Debate over the usefulness of consumer sentiment surveys in forecasting economic activity began soon after their introduction in the 1940s. The possibility that a decline in consumer confidence helped cause or worsen the 1990–91 recession renewed interest in the debate. Most recent studies of sentiment have

For helpful comments the author thanks two anonymous referees, Martin Browning, Chris Carroll, Roger Ferguson, Todd Sinai, Sarah Tanner, and seminar participants at Rutgers, the Federal Reserve Bank of Philadelphia, the Michigan/HRS/BLS Conference on Survey Research on Household Expectations and Preferences, Chicago, Princeton, Tilburg, the Federal Reserve Board of Governors, McMaster, York, the NBER summer workshop and the Wharton macro lunch. I am very grateful to Richard Curtin for providing data and feedback, to Hoon Ping Teo for excellent research assistance, and for financial support to the Rodney L. White Center for Financial Research. Of course, none of these people is responsible for any errors herein. An earlier version of this paper circulated under the title "Consumer Sentiment: Its Rationality and Usefulness in Forecasting Expenditure—Evidence from the Michigan Micro Data".

NICHOLAS S. SOULELES *is an associate professor at the Finance Department of the University of Pennsylvania and NBER. E-mail:* souleles@wharton.upenn.edu

Received March 23, 2000; and accepted in revised form May 14, 2002.

Journal of Money, Credit, and Banking, Vol. 36, No. 1 (February 2004)

focused on the time-series relationship between aggregate consumption and the two main aggregate indices of sentiment, the Michigan Index of Consumer Sentiment (ICS) and the Conference Board Consumer Confidence Index. This paper, by contrast, provides perhaps the first comprehensive analysis of the household-level data that underlies the ICS, the Michigan Survey of Consumer Attitudes and Behavior (CAB). The attention that the ICS receives, from policymakers, academics, and the business community, itself warrants an analysis of the underlying data. There are also a number of methodological advantages to such an analysis.

First, with micro data one can assess the rationality of household expectations. Most previous rationality tests have limited their focus to inflation expectations, just one of the many variables that will be examined here. Also, the tests have generally used aggregated data or at most short micro panels. But when agents' information sets differ, aggregation can lead to spurious rejections of rationality. The average of rational individual forecasts need not be a rational forecast conditional on any single information set (Keane and Runkle 1990). And, even if individual forecasts are perfectly rational, it might take a long time—perhaps multiple business cycles—for forecast errors to average out. Hence, to test rationality, it is important to use micro data on expectations over long sample periods. Unfortunately, such data are not usually available. The CAB survey, however, is unique in containing almost 20 years of monthly household expectations data. This paper exploits its panel aspect to test more cleanly than usual whether expectations are unbiased and efficient. The results can also be interpreted as explicitly characterizing the time-series and cross-sectional properties of the shocks that have hit different types of households over time, across business cycles, and policy regimes. In addition to its welfare implications, such a characterization is of methodological interest because both theoretical and empirical models are generally sensitive to the assumptions made about shock processes. In particular, many models assume that "aggregate" shocks affect all households equally.

Second, this paper assesses whether the sentiment surveys are useful in predicting behavior, specifically household spending. The canonical permanent income (or life cycle) hypothesis (PIH) provides a natural setting for this assessment. One of the central implications of the PIH is that current consumption should incorporate all the information available to an agent. However, the econometrician does not independently observe the contents of agents' information sets, so tests of this implication usually need to make strong assumptions, inferring agents' expectations econometrically. This paper instead uses direct measures of expectations from the CAB data. This data is matched, using a rich set of demographic variables, with the Consumer Expenditure Survey (CEX), which has the most comprehensive micro data on expenditure. The resulting test is whether the expectations data contain additional information, beyond that in current consumption, that helps predict future consumption. Previous studies of the excess sensitivity of consumption to sentiment have used aggregate sentiment data but aggregation can induce spurious excess sensitivity even when there is none at the micro level (Attanasio and Weber 1995). The construction of the ICS is not necessarily consistent with the construction of aggregate consumption. For instance, the ICS is an equal-weighted average of the sentiment of the CAB survey respondents, which ignores differences in the scale of consumption across respondents.

With micro data, one can also more readily investigate the sources of any excess sensitivity. One alternative hypothesis that has not previously received much scrutiny is that forecast errors might not be classical but rather contain systematic components correlated with the excess sensitivity regressor. For instance, over the sample period, high-income households might on average have been optimistic about the future, and might have happened to receive disproportionately positive shocks. In this case, increases in their consumption and so a positive correlation between consumption and sentiment would not be inconsistent with the PIH. More broadly, Chamberlain (1984) and others have pointed out that systematic forecast errors can be a potential problem in estimating any rational expectations (or forward-looking) model in a short panel. Because direct measures of households' forecast errors are available here, it is possible to test this point directly.

Third, the aggregate ICS ignores potentially useful information available in the micro CAB data. As already noted, the ICS neglects the cross-sectional distribution of sentiment. This distribution might be useful in predicting the expenditure of different groups of consumers or even aggregate expenditure insofar as the relation between expenditure and sentiment at the household level does not aggregate up. In the ICS, a given respondent's sentiment is in turn the sum of her answers to five very different survey questions, which makes it hard to interpret. This paper examines, separately for each question, whether the survey responses help forecast household spending. This examination also addresses one perennial question in the forecasting literature: Does sentiment provide information useful in forecasting above and beyond the information contained by other available macro variables like stock prices? By controlling for time effects in the micro data, one can exploit purely cross-sectional variation that is orthogonal to any macro variable.

To preview the results, expectations appear to have been biased, at least ex post, in that forecast errors did not average out even over a long sample period lasting almost 20 years. This bias is not constant over time; it is related to the inflation regime and the business cycle. People underestimated the disinflation of the early 1980s and in the 1990s and generally appear to underestimate the severity of business cycles. Expectations are also inefficient, in that people's forecast errors were correlated with their demographic characteristics. That is, forecast errors are systematically heterogeneous. The results suggest an important role for time-varying, group-level shocks—aggregate shocks do not hit all people equally. For instance, during recent expansions high-income households received relatively good shocks but low-income households continued to receive somewhat negative shocks on balance, consistent with ongoing, unexpected skill-biased technical change. Further, sentiment is useful in forecasting future consumption, even beyond lagged consumption and other macro variables, counter to the PIH. Higher confidence is correlated with less saving, consistent with precautionary motives and increases in expected future resources. Some of the rejection of the PIH is found to be due to the systematic demographic components in forecast errors. But even after controlling for these components, some excess sensitivity persists. More broadly, because forecast errors are correlated with household demographic characteristics, they will be correlated with many regressors of interest in forward-looking models. This suggests that systematic heterogeneity in forecast errors is in practice a general problem.

The paper begins by surveying related studies in Section 1. Section 2 describes the data and Section 3, the econometrics. Section 4 tests the rationality of expectations and more generally characterizes the properties of forecast errors. Section 5 tests whether sentiment helps forecast expenditure, and if so, whether this is due to systematic heterogeneity in forecast errors. Section 6 concludes.

1. RELATED STUDIES

Most tests of the rationality of surveyed expectations have focused on inflation expectations of economists (e.g., Keane and Runkle 1990). A few studies have examined the inflation expectations of consumers in general, using the aggregated Michigan data (e.g., Maddala, Fishe, and Lahiri, 1981, Gramlich, 1983, Batchelor, 1986). These studies mostly analyzed the Michigan question that allows only discrete, qualitative responses about the future path of inflation (up/down/no change). To use this question quantitatively, the studies typically made strong assumptions to derive a continuous-valued expectations time-series from the Michigan data. Moreover, as already noted, because of aggregation bias, the implications of these tests for individual rationality are not straightforward. One study, Batchelor and Jonung (1989), examined micro-level data on the inflation expectations of a small and short (one year) Swedish panel, finding evidence of bias and inefficiency. However, rationality does not require that people's expectations be on target over the course of only a single year.

Flavin (1991) and Alessie and Lusardi (1997) used micro-level data on income expectations to predict future income. While they did not formally test the rationality of these expectations, they did find a positive, if not very large, correlation between them and future realizations of income (see also Domnitz 1988). More recently, an interesting paper by Das and van Soest (1999) tested the rationality of income expectations in a Dutch dataset. They found that income expectations were on average too low relative to subsequent realizations. However, their data is also limited to a relatively short panel (1984–88). As shown below, even five years might be too few to allow forecast errors to average out. Expectations might have been rational ex ante, but might not appear rational ex post. For instance, the sample might by chance have received unexpectedly good income realizations over the period. This paper, by contrast, uses almost 20 years of micro data for many different kinds of expectations questions. Of course, even 20 years might not be a long enough period. But such a result would be as significant as a finding of irrationality because most micro studies are limited to datasets with a shorter sample period.

Even if expectations are not fully classical, people might still act on them and so they might help forecast spending. Of particular interest is whether sentiment surveys contain predictive information not available in other variables, most saliently current consumption. Two interesting papers have examined this issue using aggregate time-series data, in an Euler-equation framework.[1] Carroll, Fuhrer, and Wilcox (1994) used the ICS and Acemoglu and Scott (1994) used a similar Gallup poll in Britain. Both found significant excess sensitivity of consumption to sentiment and suggested that sentiment might be picking up precautionary motives. But under this interpretation, the sign of their estimated excess sensitivity is somewhat surprising: increased confidence led to a steeper consumption profile, i.e., to increased saving; whereas the simplest precautionary model would have increased confidence lead to less saving.[2,3] Also, it remains an open question whether other variables might already incorporate the information in aggregate sentiment. While Carroll, Fuhrer, and Wilcox show that the ICS contains additional information beyond that available in aggregate income, other studies have found that financial variables, in particular stock prices, significantly reduce the contribution of aggregate sentiment in forecasting (Friend and Adams, 1964, Ludvigson, 1996). By revisiting the matter using micro data, this paper avoids potential aggregation bias and takes advantage of additional information in the cross-sectional distribution of sentiment.

Only a few papers have used micro-level expectations data in an Euler-equation framework.[4] Two of the most interesting are by Flavin (1991) and Alessie and Lusardi (1997), who used income expectations as instruments for income in the

1. The earliest study of which I am aware that used an Euler equation framework to analyze sentiment is an unpublished Federal Reserve Board working paper by Burch and Gordon (1985), again using aggregate data. The Gulf War triggered a number of additional studies of aggregate sentiment, often by researchers in the Federal Reserve System (e.g., Throop, 1992, and Carroll, Fuhrer, and Wilcox, 1994).

2. A steeper consumption profile implies increased saving under the null hypothesis of the PIH. Outside the PIH, this implication need not hold.

3. Carroll, Fuhrer, and Wilcox (1994) note that frictions in consumption, e.g., due to habits, can potentially explain the sign of their results. Acemoglu and Scott (1994) suggest a different explanation: higher confidence might be correlated with higher levels of income, which in turn might be correlated with a higher variance in income, and so a greater precautionary motive.

4. Some of the earliest studies of sentiment, in the 1950s and 1960s, also used micro data. Their results were mixed. (See McNeil (1974) for a summary.) They generally had small sample sizes and short time horizons. Further, it is often difficult to interpret their results because the models of consumption they used are generally different from current models. Outside the consumption literature, Nicholson and Souleles (2002a, 2002b) find that income expectations of medical students help predict their specialty choice and subsequent practice behavior. They also trace the source of physicians' forecast errors to particular shocks to their practices and health-care market, such as the emergence of HMOs.

related Euler equation for saving. Both rejected the PIH. However, both studies were limited to essentially single cross sections (the 1967 SCF and a 1986 Dutch panel, respectively), leaving systematic heterogeneity in forecast errors a potential problem.[5] To illustrate, Mariger and Shaw (1993) showed that in the Panel Study of Income Dynamics the excess sensitivity coefficient on lagged income growth varies in sign from year to year. For instance, the three-year sample used by Hall and Mishkin (1982) yields a negative coefficient, but other short samples yield a positive coefficient. Mariger and Shaw conjectured that this instability might be due to aggregate shocks. But in contrast to this paper, without an independent measure of these shocks they were unable to test their conjecture directly.

2. DATA

2.1. The Michigan Survey of Consumer Attitudes and Behavior

The CAB is a nationally representative survey that since 1978 has been conducted monthly. This paper uses the data from December 1978 through June 1996. In recent years, about 500 households are sampled each month, in the earlier years two to three times as many were sampled. The five questions that comprise the widely followed ICS are as follows. The allowed responses are in brackets (underlining in original).

QFP^r. (Financial Position realization) We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago? [better now, same, worse now]

QFP^e. (Financial Position expectation) Now looking ahead—do you think that <u>a year from now</u> you (and your family living there) will be <u>better off</u> financially or <u>worse off</u>, or just about the same as now? [will be better off, same, will be worse off]

QBC. (Business Conditions) Now turning to business conditions in the country as a whole—do you think that during the next 12 months we'll have <u>good</u> times financially, or bad times, or what? [good times, good times with qualifications, pro-con, bad with qualifications, bad times]

QBC5. (Business Conditions, five-year horizon) Looking ahead, which would you say is more likely—that in the country as a whole we'll have continuous good times during the <u>next 5 years</u> or so, or that we will have periods of widespread unemployment or depression, or what? [good times, good times qualified, pro-con, bad times qualified, bad times]

QDP. (Durables Purchases) About the big things people buy for their homes—such as furniture and refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items? [good, pro-con, bad]

5. A recent paper by Jappelli and Pistaferri (2000) uses a few cross sections of income expectations from an Italian Survey in an Euler equation. While they do not find excess sensitivity, they note this might be due to measurement error, especially in the timing of their expectational questions vis-à-vis the other variables.

Some economists are wary of subjective survey questions. Instead of reviewing their generic advantages and disadvantages or offering an exegesis of these particular questions, this paper will formally test the rationality of the responses to the questions and see whether they are correlated with behavior, specifically whether they help forecast spending.[6,7] In a related paper, Souleles (2001) shows that these same questions help predict household purchases of risky securities. Even controlling for past stock returns, households that are pessimistic about the future buy fewer risky securities, ceteris paribus.[8]

A few additional notes are in order. First, questions QBC, QBC5, and QDP ask the respondent about aggregate economic activity, while QFP^r and QFP^e ask about the household's own financial position. This difference suggests there might be more cross-sectional variation in QFP^r and QFP^e than in the other variables. Second, QFP^e, QBC, and QBC5 ask about the future[9], whereas QFP^r asks about the past year and QDP asks about the present. Third, the wording of QFP^e ("e" for expectation) matches that of QFP^r ("r" for realization). Thus if someone is asked QFP^e this year, and then QFP^r next year, QFP^e provides a forecast of what his answer to QFP^r will be. However, the response to QFP^e is constrained to fall in one of three categories (better, worse, or the same). Therefore the analysis will accommodate the discrete, ordered nature of this and the other variables. For convenience, the better states ("better" or "good" or "good with qualification") are usually coded as $+1$, the intermediate states ("same" or "pro-con") as 0, and the worse states ("worse" or "bad" or "bad with qualification") as -1.[10]

Figures 1 and 2 show the average response for each question month-by-month. All five variables are procyclical. Notably, the forward-looking expectational variable QFP^e appears to lead the backward-looking realization QFP^r. For instance, QFP^e recovers more quickly from both the 1980–81 recession and the 1990 invasion of Kuwait. Nonetheless, the two aggregate time series are highly correlated, at about 0.8.

The CAB survey asks many additional questions. This paper highlights the five questions above because they comprise the ICS but will also consider the most salient of the additional questions, listed in Appendix A. There are two matching

6. As for the particular wording of these questions, they have the virtue of having stayed the same over the sample period. Also, it is worth noting that most household-level data, not just sentiment, is self-reported by households.

7. Carroll, Fuhrer, and Wilcox (1994) and others have shown that the aggregate ICS helps forecast aggregate consumption. Studies of the CAB inflation expectations, described below, have found that they are helpful in predicting CPI inflation, sometimes performing better than inflation forecasts from professional forecasters (Thomas 1999).

8. For another application of the CAB, to tax cuts, see Shapiro and Slemrod (1995).

9. These three questions make up the Expectations subindex of the ICS, which in turn is a component of the Index of Leading Economic Indicators.

10. The ICS uses this coding in a diffusion index. For each question, the aggregate value at a given time is the number of people answering $+1$ at that time minus the number of people answering -1. Such indexes omit the people answering 0, as well as the distribution of the rest of the answers across people of different characteristics.

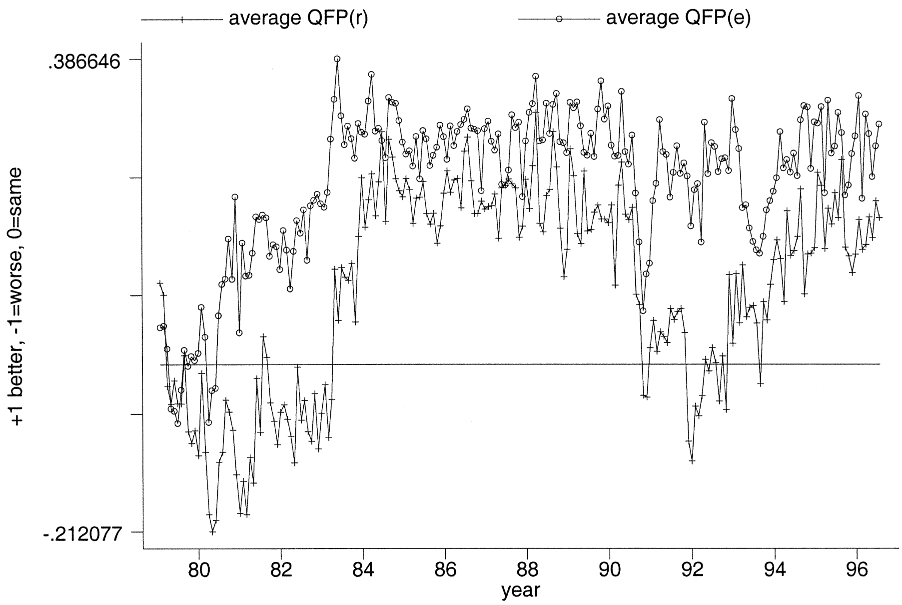


FIG. 1. Monthly means of the household financial position variables QFP^r and QFP^e. For the realizations question QFP^r, responses were coded as +1 if the household's financial position is better now than a year ago, −1 if it is worse now, and 0 if it is about the same. For the expectations question QFP^e, responses were coded as +1 if its financial position is expected to be better a year from now, −1 if expected to be worse, and 0 if expected to be about the same.

questions on business conditions similar to QBC. Since one can be taken as the expectation of the other, they will be denoted QBC^e and QBC^r. There are also matched questions about changes in prices, QP^e and QP^r, and changes in the household's real income, QY^e and QY^r, over the following year and previous year, respectively. QU^e asks whether the respondent expects the national unemployment rate to increase or decrease over the next year. Even though there is no matching realization question about perceived changes in unemployment over the past year, this question is used because precautionary saving might be sensitive to unemployment expectations.[11] The answers to all these questions are again discrete and ordered. For business conditions QBC and household income QY, again +1 denotes the good state. But note that for inflation QP and unemployment QU, +1 denotes the bad state (an increase in inflation or unemployment). There are also matched pairs of continuous-valued (quantitative) questions, which can be used to verify that the discreteness of the previous questions is not driving their results. The continuous questions concern the inflation rate over the next and past 12 months (denoted by $Q\Pi^e$ and $Q\Pi^r$) and the growth rate of the household's income (QGY^e and QGY^r).

11. Carroll (1992) was amongst the first to explicitly link QU to precautionary motives, in an aggregate time-series context. More recently Carroll, Dynan, and Krane (1996) examine the effects of cross-sectional differences in (ex post) unemployment rates on balance sheets in the Survey of Consumer Finances. The results are consistent with precautionary saving.

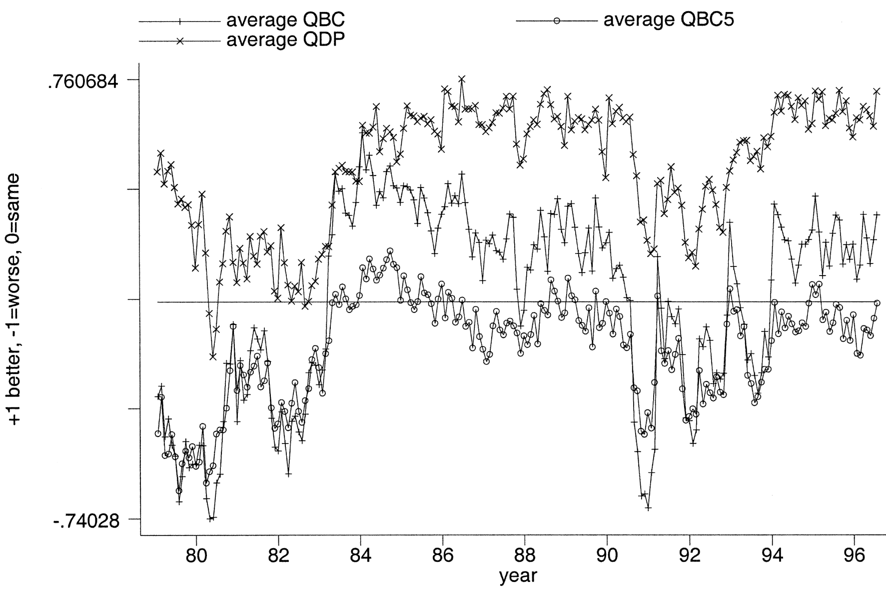


FIG. 2. Monthly means of the aggregate conditions variables QBC, QBC5, and QDP. Questions QBC and QBC5 elicit expectations for business conditions in the next year and next five years, QDP asks whether now is a good time to buy durables. Responses were coded as $+1$ for good times, -1 for bad times, 0 otherwise.

Unlike the five ICS questions (QFP^r to QDP), these additional questions were not always asked in every month of the sample period. They will be used over the periods for which they are available.

Even though the CAB surveys are archived as independent cross sections, there is a short panel aspect to them that has not previously been much exploited: Households are reinterviewed once and re-asked the same sentiment questions. Much effort was expended by the author to create a single, consistent panel dataset from the entire history of CAB cross-sections. Explicit forecast errors could then be calculated for the matched pairs of questions by taking a realization from the second interview (e.g., QY^r_2, where the subscript refers to the interview number) and subtracting the corresponding expectation from the first interview (QY^e_1). Thus, for a given household, the error regarding income is defined as $\varepsilon Y \equiv QY^r_2 - QY^e_1$. Errors for financial position, business conditions, and prices are defined similarly: $\varepsilon FP \equiv QFP^r_2 - QFP^e_1$, $\varepsilon BC \equiv QBC^r_2 - QBC^e_1$, and $\varepsilon P \equiv QP^r_2 - QP^e_1$, respectively. Given the coding of the underlying variables Q in $\{-1,0,1\}$, these errors ε take on values in the set $\{-2,-1,0,1,2\}$. With a few exceptions, since December 1978 the second household interview in the CAB survey has taken place six months after the first interview. For consistency in calculating forecast errors, the sample is started in December 1978 and is limited to households reinterviewed after six months. Since the forecast horizon written into most of the expectational questions is one year, not six months, the timing in forming the errors ε is unavoidably inexact.

Nevertheless, the timing is exogenous and unsystematic, since the sample covers every month over almost two decades.[12] Extensions below will verify that this timing issue does not drive the results.

For the quantitative questions on expected inflation and income growth, $Q\Pi^e$ and QGY^e, continuous forecast errors can be computed analogously, e.g., $\varepsilon GY \equiv QGY_2^r - QGY_1^e$.[13] For inflation, there is more flexibility in computing the errors since the actual consumer inflation rate can be measured independently via the CPI. Three different forecast errors $\varepsilon\Pi$ are computed. The "subjective" error $\varepsilon\Pi^{subj} \equiv Q\Pi_2^r - Q\Pi_1^e$ compares the inflation rate the respondent expects over the next 12 months, taken from the first interview, with the inflation rate the respondent believes was realized over the past 12 months, taken from the second interview six months later. Again, because the realization variable is not elicited exactly 12 months later, the timing is not exact. To avoid this problem, the "objective" error $\varepsilon\Pi^{obj} \equiv \Pi_{12}^r - Q\Pi_1^e$ compares the 12-month inflation rate the respondent expected in the first interview with the actual inflation rate over the next 12 months, according to the CPI (Π_{12}^r). In this case, the timing is exact. The third error $\varepsilon\Pi_6^{obj} \equiv \Pi_6^r - Q\Pi_1^e$ uses for its realization the CPI inflation rate over only the first six months following the first interview, annualized. This error can be contrasted with $\varepsilon\Pi^{obj}$ to investigate the effects of the six-month mistiming in the other forecast errors.

The CAB survey also includes a number of demographic questions. Since some of these changed across surveys, great care was taken to create a single set of demographic variables consistent across the entire sample (and consistent with the CEX). Appendix A provides more details. The main sample exclusion concerns the survey respondent. The sample drops an observation when there is a married couple in the household but the respondent is neither of the spouses. (Most such respondents appear to be grown children of the couple.) This should help make the respondent's answers more representative of the entire household. Demographic variables referring to the reference person were switched to refer to the head of household (i.e., for a married couple, the male, following the convention in the literature). An additional exclusion was adopted in forming the subjective forecast errors (i.e., all but the objective inflation errors): to make the answers in both interviews more comparable, the same person had to be the respondent in both interviews.

12. Suppose a household's first interview is in month t, and Q_1^e refers to the expected change in some variable X between months t and $t+12$, and Q_2^r from the second interview elicits the realized change $X_{t+6}–X_{t-6}$. Then the timing mismatch corresponds to the term $[(X_t–X_{t-6})–(X_{t+12}–X_{t+6})]$. This term can reasonably be assumed to average out over the long sample period, and in the cross section. E.g., events that take place in months 7–12 after the first interview for one household, will appear in months 1–6 before the first interview for other households interviewed later, and hence tend to average out.

13. There is an additional complication regarding the timing of QGY^e. The corresponding realization question elicits the level of household income (not the growth rate) in the previous calendar year. Since the second interview follows after only six months, to compute a nonzero growth rate for income from one year to the next, QGY^r, the sample for this question must be limited to households whose first interview takes place in the second half of the year, so that the second interview takes place in the following calendar year. By contrast, the expectational question $QGY^e{}_1$ asked in the first interview refers to income growth over the next 12 months so its reference period will somewhat lag the reference period of the computed QGY^r.

2.2. *The Consumer Expenditure Survey*

Because the CAB survey does not include much data on expenditures, it is matched with the CEXs, from 1982–93.[14] CEX households are interviewed four times, three months apart (though starting in different months for different households). The reference periods for expenditure cover the three months before each interview. Strictly speaking, the Euler equation used below applies only to nondurable consumption, but for gauging the aggregate effect of sentiment total consumption also matters. Indeed, some analysts have suggested that sentiment matters most for durables purchases. Therefore, for each household-quarter, both real nondurable expenditure and real total expenditure were computed (1982–84 $).

The CEX sample was selected in standard ways to improve the measurement of consumption. A household was dropped from the sample if there were multiple "consumer units" in the household, or the household lived in student housing or the head of household was a farmer. A household-quarter was dropped if no food-expenditure was recorded in the quarter or any food was received as pay in the quarter. Appendix A provides further details about the data.

3. ECONOMETRIC SPECIFICATIONS

The sentiment of the CEX households will be imputed from the sentiment of demographically similar households interviewed at the same time in the CAB survey. Since the surveys contain a rich, overlapping set of demographic variables, the imputation can be made very fine. Table 1 shows the means of the main variables used. The CAB sample is somewhat more highly educated and likely to live in the South. But generally the means are rather similar, as one would expect from two representative datasets. The imputation proceeds in two steps.

The first step takes place in the CAB data. For the discrete sentiment variables, since their responses are ordered, both linear and ordered probit models will be estimated. In the latter, for a given sentiment variable $Q \in \{-1, 0, +1\}$ and household i, let $Q^*_{i,t}$ be the corresponding (continuous) latent index at time t, representing i's underlying sentiment or confidence. $Q^*_{i,t}$ is assumed to take the following form:

$$Q^*_{i,t} = \boldsymbol{a}_0' \,\mathbf{time}_t + \boldsymbol{a}_1' \,\mathbf{Z}_{it} + u_{i,t}\,. \tag{1}$$

Except for the questions on inflation and unemployment, larger values of Q^* reflect better states. **Z** is the vector of demographic instruments used to link the two datasets, from Table 1. The vector **time** includes a full set of month dummies (a different dummy for each month of each year). These dummies will allow for changes in the average level of sentiment from month to month. Since the cross-sectional distribution of sentiment around the average can also change over time, some of the demographic variables are interacted with the month dummies.

14. The first wave of the CEX, 1980–81, is not used because its data are generally poorer than the data from the following waves.

TABLE 1

Sample Means, 1982–93

	CAB	CEX
Age	45.5	48.7
ln(income)	9.93	9.97
Married	0.567	0.583
Separated	0.269	0.288
Nonwhite	0.093	0.116
Female head	0.252	0.282
No high school	0.165	0.243
Some college	0.261	0.217
College	0.279	0.237
1 kid	0.154	0.155
2 kids	0.150	0.145
3+ kids	0.081	0.083
2 adults	0.526	0.552
3+ adults	0.101	0.125
Midwest	0.284	0.262
South	0.322	0.287
West	0.201	0.237
# obs	54488	97993

Notes: For comparison the Michigan Survey of Consumer Attitudes and Behavior (CAB) sample period is restricted to the Consumer Expenditure Survey (CEX) sample period, 1982–93. The omitted categorical variables are: single, white, male head, high school graduate, no kids, one adult, northeast. Income is real household income (1982–84 $). The actual samples used in the different analyses in the paper can differ somewhat due to missing data or additional sample restrictions, as explained in the text and the following tables.

Because there are well over 100 months in the sample, to keep the computational requirements tolerable only a few variables could be interacted simultaneously. Preliminary analysis found that for most sentiment questions, the effects of age and income varied the most significantly over time, so **Z** also includes month-interactions for these two variables. Ordered logits were also estimated, but since the results were quite similar they are not reported. For the continuous variables QGY and QΠ, the same functional form in Equation (1) is estimated by OLS.

The second step takes place in the CEX. The estimated coefficients from the first step, $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$, are used to impute the (continuous index value) level of sentiment $\hat{Q}$ of the CEX households with the same demographic characteristics **Z**:

$$\hat{Q}_{i,t} = \hat{\mathbf{a}}_0' \, \mathbf{time}_t + \hat{\mathbf{a}}_1' \, \mathbf{Z}_{it} + u_{i,t} \, . \tag{2}$$

Lagged $\hat{Q}$ is then added to a standard linearized Euler equation for consumption. For household i the change in log consumption between periods $t+1$ and t is specified as

$$d \ln C_{i,t+1} = \mathbf{b}_0' \, \mathbf{time}_t + \mathbf{b}_1' \, \mathbf{W}_{i,t+1} + b_2 \hat{Q}_{i,t} + \eta_{i,t+1} \, . \tag{3}$$

Following Zeldes (1989), Dynan (1993), Lusardi (1996), and Souleles (1999), **W** will include the age of the household head and changes in the number of adults and in the number of children. These variables help control for the most basic changes in household preferences over time.[15]

For a given household, the consumption changes in Equation (3) are taken over successive three-month periods. To keep the sentiment data timely, the time-varying components of $\hat{Q}_{i,t}$ are estimated from the CAB survey corresponding to the first of the three months covered by $C_{i,t}$. For instance, consider the case in which $C_{i,t}$ records consumption in November 1990 to January 1991 (and $C_{i,t+1}$ covers February–April 1991). In Equations (1) and (2), the month dummies $\mathbf{time_t}$ and the month-interacted variables in $\mathbf{Z}_{it}$ would then correspond to the November 1990 CAB survey. $\hat{Q}_{i,t}$ is therefore predetermined in Equation (3), and so under the PIH, the coefficient b_2 should be zero. That is, given current consumption, current sentiment should not help predict future consumption.

OLS estimation of Equation (3) would neglect the fact that $\hat{Q}$ is a generated regressor. To take this into account the two-sample instrumental-variables technique of Angrist and Krueger (1992) will be used, although here the technique is not required for consistency but only to adjust the standard errors for the additional variation arising from the first estimation step. This technique requires that both estimation steps be linear, so for the reported excess sensitivity tests Equation (1) is estimated by OLS even for the discrete sentiment questions. A previous version of this paper reported instead the ordered probit results for the discrete questions. Comparing the results shows that the discreteness of Q makes very little difference to the excess sensitivity tests; the signs and significance of the estimated coefficients in Equation (3) are quite similar.[16] The standard errors in Equation (3) are also corrected for general heteroscedasticity and serial correlation by household.

The month dummies in Equation (3) control for all (perfectly uniform) aggregate effects, including seasonality, aggregate interest rates, and any other macro variables like stock prices that might incorporate some of the same information available in the aggregate time series of sentiment. Since the same time dummies are used in the first step in Equation (1), in Equation (3) they effectively partial out the monthly average level of sentiment, leaving only cross-sectional variation in $\hat{Q}$. Although using these time dummies makes it harder to find a significant effect of sentiment in predicting consumption, they provide a crisp test of whether the micro data contains useful information not available in the aggregated data.

15. As Deaton (1992) notes, by restricting the variables in **Z** or expanding the variables in **W**, it would be possible to eliminate most any excess sensitivity. Therefore **W** is restricted to this commonly used set of controls (age and changes in family size), for comparison with previous studies and to retain power to test for excess sensitivity and for systematic heterogeneity in forecast errors. See the survey of specifications in Table 5.1 of Browning and Lusardi (1996).

16. Jappelli, Pischke, and Souleles (1998) also applied this two-sample estimator to excess sensitivity tests. They too imposed linearity on a first-step specification that was originally discrete, and found that the final excess sensitivity results were not sensitive to this imposition. Alternatively, Equations (1) and (3) can be jointly bootstrapped, estimating Equation (1) by ordered probit. However, each ordered probit takes many hours, making bootstrapping infeasible for the full set of results below. The bootstrap standard errors were, however, computed for the first specification in Table 3 (for QFP^{r} for nondurable consumption). The resulting significance levels for the coefficients in Equation (3) were similar to those reported using the two-sample estimator.

This paper also tests the rationality of people's forecasts, namely their unbiasedness and efficiency. The results can also be interpreted as characterizing the shocks that have ex post hit different types of households over time. Efficiency requires that forecast errors be uncorrelated with any variable in an agent's information set at the time of forecast; otherwise the forecast does not take advantage of all available information. Time-series analyses of the efficiency of inflation expectations often test for serial correlation in inflation forecast errors. However, for each sentiment question the CAB data contains only one forecast error per household, so it is impossible to test for serial correlation at the micro level.[17] This paper instead tests for systematic demographic components in households' forecast errors. The focus is on cross-sectional heterogeneity, because that is the variation available in the CAB data, and the variation exploited in most excess sensitivity tests in micro data.

Specifically, heterogeneity in forecast errors will be analyzed using a specification similar to Equation (1), but with the errors ε (defined above) as the dependent variable:

$$\varepsilon_{i,t+1}{*} = \mathbf{d_0'}\,\mathbf{time}_t + \mathbf{d_1'}\,\mathbf{Z}_{it} + v_{i,t+1}\,, \tag{4}$$

where t refers to the first household interview in the CAB data, $t+1$ to the second interview. For instance, for income the error is $\varepsilon Y_{i,t+1} \equiv QY^{r}_{i,t+1} - QY^{e}_{it}$. Since the demographic variables $\mathbf{Z}_{it}$ are known to agent i at the time t of forecast, efficiency requires that $\mathbf{d_1} = 0$. The time dummies control for cross-sectional correlation due to (perfectly uniform) aggregate shocks. When ε is restricted to $\{-2,-1,0,1,2\}$ the estimation is by ordered probit, but for the continuous variables εGY and εΠ OLS is used.

Returning to Euler equation (3), the residual η can potentially include many factors, such as measurement error, approximation error from linearizing the Euler equation, or unobserved heterogeneity in discount rates. Other studies have already analyzed the complications such factors pose in estimating Euler equations, including the possibility of spurious excess sensitivity. (For a review, see Deaton 1992 or Browning and Lusardi 1996.) The focus here is instead on a different component of η: the difference between realized and expected consumption growth, resulting from forecast errors (shocks) regarding variables like household income, financial position, and the other sentiment variables. Systematic heterogeneity in forecast errors has not received much empirical scrutiny, even though it can lead to spurious inference in Euler equations and more generally in any forward-looking model. In Equation (3), for consistent estimates of b_2, the forecast errors in η need to be uncorrelated with the excess sensitivity regressor $\hat{Q}$. Most studies rely on the time dummies to soak up all systematic components of forecast errors, such as shocks due to the business cycle. But this makes the strong assumption that such shocks hit all people equally.

17. One could test for serial correlation in the aggregated sample data, but as already explained that could lead to aggregation bias.

The problem can be illustrated with a simple example. Suppose there are two groups of households in the population, those with high education and those with low education. Suppose further that in addition to aggregate and idiosyncratic shocks, there are group-level shocks that hit all members within an education group the same way but hit each group differently. In this case, time dummies will capture the aggregate shocks but will not control for the group-level shocks. Thus, even if each household is behaving according to the PIH, a regression of household consumption growth on time dummies and household education status would produce a significant coefficient for education. If the regression does not control for education but includes an excess sensitivity regressor correlated with education, this regressor will be found to be significant even if the PIH is true, resulting in spurious excess sensitivity. More generally, if forecast errors are correlated with household demographic characteristics, they are likely to be correlated with most regressors of interest in forward-looking models.

Unlike previous studies, with direct measures of forecast errors, this paper is uniquely able to test the implications of systematic heterogeneity in the errors. Shocks to variables like household income and financial position, as well as to aggregate business conditions and inflation, must be among the most important sources of the overall innovation in consumption in η.[18,19] If $\mathbf{d}_1 \neq 0$ in Equation (4), the errors ε are not uniform across households, and then any excess sensitivity estimated in Equation (3) might be spurious. The aggregate time dummies in Equation (3) would not control for such heterogeneity. To assess this possibility, the forecast errors $\hat{\varepsilon}$ of the CEX households will be imputed from the forecast errors of the CAB households with the same demographic characteristics $\mathbf{Z}$, in another two-step process. Then the term $b_3\ \hat{\varepsilon}_{i,t+1}$ will be added to Equation (3). Under the alternative hypothesis that excess sensitivity is being generated by the demographic components in forecast errors, one would expect to find $b_2 = 0$ and $b_3 > 0$ ($b_3 < 0$ for inflation and unemployment), since the PIH allows consumption to respond to the innovations represented by $\hat{\varepsilon}$.[20]

18. Indeed shocks to overall financial position εFP might be more representative of innovations to household consumption and welfare than shocks to just current income, which are more commonly analyzed.

19. Even if the residual η in Equation (3) contains more than the forecast errors ε for income, financial position, etc., orthogonality of ε is a necessary condition for orthogonality of η. E.g., if people's forecast errors for future income are correlated with their demographic characteristics, then so will their innovation in consumption. Of course, other factors in η can also generate excess sensitivity, but under the null hypothesis that these forecast errors are classical these factors will be independent of ε. Thus other factors alone cannot explain the effects of controlling for ε in Equation (3).

20. For instance, Deaton (1992) discusses a model in which income innovations are generated according to $\Delta y_{it} = e_t + g_i e_t + w_{i,t} - w_{i,t-1}$, where e_t is a common permanent shock, w_{it} is an idiosyncratic transitory shock, and g_i is a mean-zero loading factor capturing the nonuniform effect of the aggregate shock across different households. Under the PIH, then $\Delta c_{it} = e_t + g_i e_t + w_{it} r/(1 + r)$, for interest rate r. Hence innovations to household income feed directly into consumption, according to their persistence and cross-sectional loadings, generating $b_3 > 0$. Equation (4) can be thought of as the empirical generalization of this model for income innovations Δy. Analogously one would expect positive innovations to household financial position and aggregate business conditions to lead on average to increases in consumption, generating $b_3 > 0$ for these variables as well. Note that in this model for Δc, time dummies will control for only the first term, the common shock e_t. If the other two terms are correlated with the excess sensitivity regressor Q, as is likely if forecast errors are inefficient, then this would generate spurious excess sensitivity even conditional on the time dummies.

The errors $\hat{\varepsilon}$ can be imputed in two different ways. First, Equation (4) can be estimated directly on the forecast errors $\varepsilon_{t+1} \equiv Q^{\mathrm{r}}_{t+1} - Q^{\mathrm{e}}_{t}$ in the CAB data and then used to impute $\hat{\varepsilon}_{t+1} \equiv \widehat{Q^{\mathrm{r}}_{t+1} - Q^{\mathrm{e}}_{t}}$ in the CEX. Here again, the first household interview t in the CAB data is chosen to correspond to C_{it} in the CEX. Alternatively, in an extension, Equation (1) is first used to impute the levels of sentiment in the CEX, both realized $\hat{Q}^{\mathrm{r}}_{t+1}$ and expected $\hat{Q}^{\mathrm{e}}_{t}$. The difference between these variables then gives the forecast errors $\hat{\varepsilon}_{t+1} \equiv \hat{Q}^{\mathrm{r}}_{t+1} - \hat{Q}^{\mathrm{e}}_{t}$, with the timing matching the quarterly consumption change in Equation (3).[21]

4. RESULTS: THE RATIONALITY OF EXPECTATIONS AND THE PROPERTIES OF FORECAST ERRORS

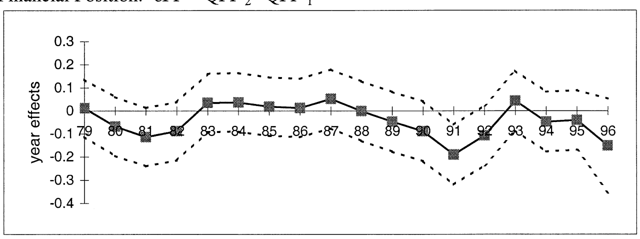
This section analyzes the time-series and cross-sectional properties of households' forecast errors. The working-paper version of this paper presented 3×3 cross-tabulations of the matched pairs of discrete CAB variables, the expectational variables Q^{e}_{1} with their corresponding realizations Q^{r}_{2}, both coded in $\{-1,0,1\}$. Following Manski (1990), Das, Dominitz, and van Soest (1999) derive nonparametric tests of whether a given pair of realizations and expectations is drawn from the same underlying distribution. These rationality tests explicitly accommodate the discreteness of the paired variables, assuming that the expectational variable represents the category containing either the median or the mode of the respondent's subjective distribution for the underlying variable at issue. Applying the test for the median assumption (not reported), rationality is significantly rejected for three of the four discrete expectational variables, $\mathrm{QFP^e}$, $\mathrm{QY^e}$, and $\mathrm{QP^e}$. (The corresponding sample periods are reported in Figure 3.) For $\mathrm{QBC^e}$, rationality is rejected for most of the sample years separately, not the pooled data. The pattern of rejection varies over time in a striking way. In the early 1980s and early 1990s, i.e., around the two recessions in the sample period, business condition realizations $\mathrm{QBC^r}$ were systematically worse than expected (relative to $\mathrm{QBC^e}$); whereas in expansions they were generally better than expected.[22] The other "nonprice" realizations, $\mathrm{QFP^r}$ and $\mathrm{QY^r}$, exhibit similar patterns over the business cycle. The inflation realization $\mathrm{QP^r}$ systematically turned out higher than expected, in the pooled data and for most years at the beginning of the sample period. The results are similar using the mode assumption.

To summarize the signs and magnitudes of the forecast errors, it is convenient to compare the probability of a realization turning out worse than expected with the

21. Despite the complications that the six-month mistiming in forming the forecast errors poses for testing the rationality of forecast errors, the timing has an advantage for the Euler equation tests: The errors ε_{t+1} cover the same six-month period as $\mathrm{dln}C_{t+1}$. Therefore ε_{t+1} will appropriately incorporate any news that comes out over the six months.

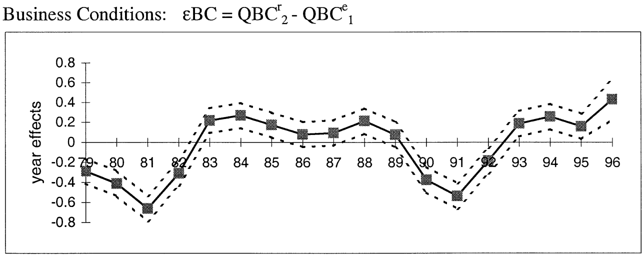
22. E.g., in 1980, conditional on $\mathrm{QBC}^{e}_{1} = 0$ (no change expected), 67% of respondents subsequently reported $\mathrm{QBC}^{r}_{2} = -1$ (worse now); conditional on $\mathrm{QBC}^{e}_{1} = +1$ (improvement expected), 68% reported $\mathrm{QBC}^{r}_{2} = -1$ or 0. Both of these figures are significantly greater than the maximum 50% allowed under the median assumption.

A Financial Position: $\varepsilon FP = QFP^r{}_2 - QFP^e{}_1$



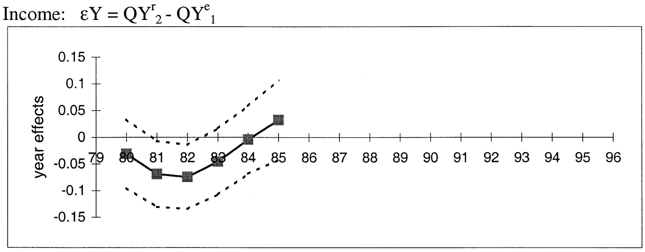
obs = 42767
Pseudo R^2 = .01
$\chi^2(18)$ = 158, pval=0.00
μ = -.064 (.007)

B Business Conditions: $\varepsilon BC = QBC^r{}_2 - QBC^e{}_1$



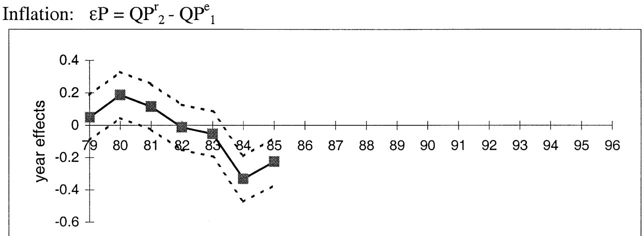
obs = 42645
Pseudo R^2 = .03
$\chi^2(18)$ = 3325, pval=0.00
μ = -.105 (.026)

C Income: $\varepsilon Y = QY^r{}_2 - QY^e{}_1$



obs = 17503
Pseudo R^2 = .01
$\chi^2(6)$ = 17.2, pval=0.01
μ = -.085 (.008)

D Inflation: $\varepsilon P = QP^r{}_2 - QP^e{}_1$



obs = 19577
Pseudo R^2 = .01
$\chi^2(7)$ = 349, pval=0.00
μ = .056 (.012)

FIG. 3. Time effects in forecast errors: discrete CAB variables. For variable and sample definitions, see the notes for Table 2. The graphed results come from an ordered probit of the forecast errors ε in $\{-2,-1,0,1,2\}$ on year dummies. The middle line gives the estimated coefficients on the year dummies (in the latent index function). The outside dashed lines represent 95% confidence intervals. χ^2 tests the joint significance of the year effects. To calculate the mean forecast error μ, the errors ε are regressed by OLS on a constant, correcting the standard errors for heteroscedasticity and cross-correlation within the month.

probability of its turning out better than expected. In 3×3 tables, this requires that one specify how much worse it is to end up two places (cells) off the diagonal than one place off. However, one can avoid taking a stand on this tradeoff by collapsing the 3×3 tables into 2×2 tables, by either dropping the middle (0's) responses or by merging them into one of the other two responses ($+1$ or -1).[23] Nonparametric sign tests can then be used to test whether the probability of falling into the single northeast cell significantly differs from the probability of falling into the single southwest cell, a form of bias. Whichever way one handles the middle responses, these tests (not reported) reject unbiasedness for all four matched pairs of discrete sentiment questions. In all four cases, "bad" shocks (with financial position, business conditions, and income growth turning out worse than expected, and inflation turning out greater than expected) were more common than "good" shocks over the sample period. However, dropping or merging the middle responses wastes a good deal of information.

Alternatively one can parameterize the errors, most simply by treating their values in $\{-2,-1,0,1,2\}$ as cardinal; i.e., by assuming that being two places off the diagonal is twice as bad as being one place off. Then one can summarize the average forecast error μ by regressing the errors ε on a constant by OLS. The reported standard errors are corrected for the fact that the errors across households in a given month can be correlated by common shocks. In Figure 3, the resulting estimates of μ for εFP, εBC, and εY are all significantly negative, while the average inflation error εP is positive. (Recall that for inflation, $+1$ represents the bad state, the reverse of the other variables.) Again, the realizations are disproportionately biased towards being worse than expected.[24]

However, as suggested by the nonparametric rationality tests above, one should not conclude from these results that people are generally over-optimistic or overconfident, at all times. One can test for significant time effects in the forecast errors ε, even without cardinalizing them, by using ordered probits. Equation (4) was first estimated using only the time dummies as independent variables. The resulting coefficients and 95% confidence intervals are graphed in Figure 3, for the sample periods over which each pair of variables is available. For clarity, year dummies are presented, but the conclusions are the same using the full set of month dummies. For all four discrete forecast errors, the chi-squared tests indicate that the year dummies are jointly very significant. That is, there is significant variation in households' forecast errors from year to year. The nonprice errors εFP, εBC, and εY

23. E.g., one could test the rationality of binary variables such as "(a) Will conditions improve or at least stay the same, or (b) will conditions worsen?" This variable would correspond to grouping the 0's with the $+1$'s.

24. While the reported standard errors control for contemporaneous cross-sectional correlation from common shocks, they do not reflect the fact that the forecast horizons for households interviewed in successive months partially overlap, potentially generating serial correlation in the residuals. Since the forecast periods cover six months, this correlation could extend up to five months. To control for this, the regressions were rerun limiting the sample to nonoverlapping forecast periods (e.g., in one regression using only CAB interviews in January and July, in another regression using only February and August, etc.). Even though this throws away 5/6 of the data, for εFP and εY the average μ remained significantly negative for all samples considered (i.e., for all six pairs of months). For εBC and εP the means remained negative and positive, respectively, and were significant in about half of the nonoverlapping samples.

are most negative throughout the early 1980s and the early 1990s. Consistent with the results above, it appears that people were negatively surprised by the recessions, repeatedly over their duration.[25,26] However, recalling the procyclicality of sentiment in Figures 1 and 2 (in particular, the fact that QFP^e is a leading indicator), one should not conclude that people altogether fail to foresee the business cycle. Rather, it appears that people understate the amplitude or duration of the cycle, in both downturns and upturns. Nonetheless, the pseudo R^2's in Figure 3 suggest that time effects explain only a small part of the variation in the forecast errors. The time effects are more significant and produce a larger R^2 for the forecast error for aggregate activity, εBC, than for the household-specific errors εFP and εY.

Figure 3D records the results for the discrete inflation forecast errors εP. The year effects swing from positive to negative. Evidently inflation was higher than expected at the end of the 1970s but then people were surprised by how quickly it abated in the early 1980s.[27] Figure 4B presents analogous OLS results for the continuous, subjective inflation error, $\varepsilon\Pi^{subj}$. This error also dramatically declines from positive to negative in the early 1980s, and is positive on average over the sample period. Figure 4C shows the objective forecast error $\varepsilon\Pi^{obj}$, which uses as its realization the actual CPI inflation rate Π_{12} over the next 12 months (as opposed to the respondent-supplied realization $Q\Pi^r$ used in Figure 4B, which is not available after 1985).[28] The errors again decline with the disinflation in the early 1980s.

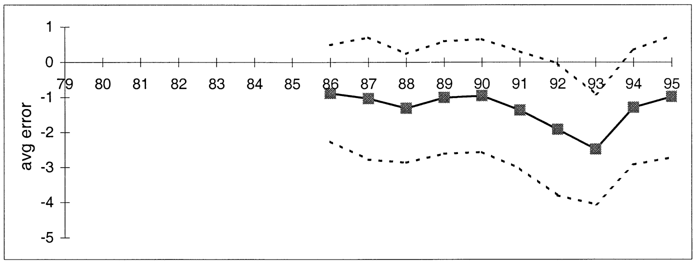
25. Under basic models of rational expectations, efficiency requires that individual agents' forecast errors be independent across time. Even though only one forecast error is available per household, one can test for serial correlation in the aggregated forecast errors, i.e., in the estimated month dummies underlying Figures 3 and 4. There is significant autocorrelation through over 10 months for εFP, εBC, εP, $\varepsilon\Pi^{subj}$, and $\varepsilon\Pi_6^{obj}$, and through over 20 months for $\varepsilon\Pi^{obj}$ (whose forecast period covers 12 months). However, recall that tests of efficiency on aggregated data are subject to aggregation bias.

26. The overlapping forecast periods described above could generate some autocorrelation through five months. But Figure 3 and the results in the previous note show that the autocorrelation in the time effects lasts much longer than this. The diagrams and conclusions remain qualitatively the same on limiting the sample to nonoverlapping periods as above, or on using the full set of month dummies in Equation (4). Estimating (4) by OLS produces year effects that are similar to those in Figure 3.

27. There is a small discrepancy in the wording of QP^e and QP^r. QP^e asks about prices in general, whereas QP^r asks about the prices of goods the household itself buys. This distinction should not matter much here. First, the CAB data are representative, so the average price of goods bought should be relatively close to the consumer price level. Second, any discrepancy is unlikely to explain the dramatic shift in forecast errors from positive to negative during the disinflation in the early 1980s. Third, the continuous questions $Q\Pi^e$ are about aggregate prices and so Figure 4C is not subject to the discrepancy, yet yields a similar pattern. Fourth, Croushore (1998) documents a similar pattern using the aggregate time series for inflation expectations from the Livingston survey and the Survey of Professional Forecasters, where again there is no discrepancy in the wording of the survey questions. Fifth, the results are similar on using the regional CPI for the census region in which the household lives, instead of the national CPI, or using instead the PCE and GDP deflators. Finally, even though QP^e does not specifically mention the CPI, both the results of the previous literature and the staff at the Institute for Social Research suggest that the CPI is the appropriate benchmark. The ISR surveyors prod respondents for the prices of "the things people buy," intending to capture consumer prices, although they deliberately avoid using jargon like "CPI-U".

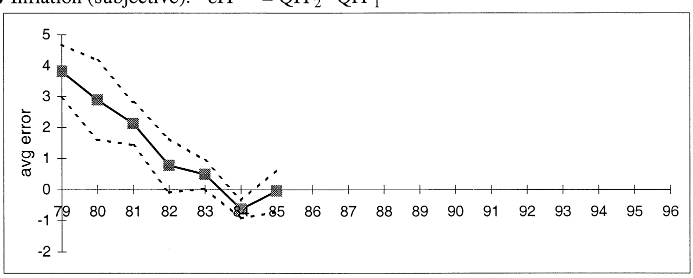
28. The drawbacks to using actual inflation emphasized by Keane and Runkle (1990) do not apply here. First, unlike the GDP deflator the CPI is not revised. (The seasonal adjustment can be changed, but this is unlikely to be important. To avoid any problem the reported results use the nonseasonally adjusted CPI. Using the seasonally adjusted CPI instead made extremely little difference.) Second, revisions are a problem only if the revised variable is used as a regressor to test efficiency but was not in agents' information sets. The efficiency tests here do not use revised variables as regressors.

A Income: $\varepsilon GY = QGY^r_2 - QGY^e_1$



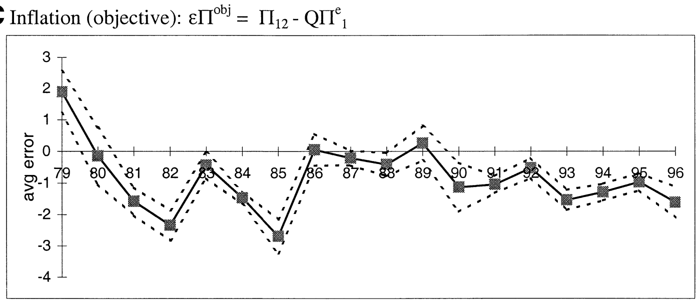
obs = 6757
$R^2 = .01$
F = 2.9, pval=0.01
$\mu = -1.31$ (.24)

B Inflation (subjective): $\varepsilon\Pi^{subj} = Q\Pi^r_2 - Q\Pi^e_1$



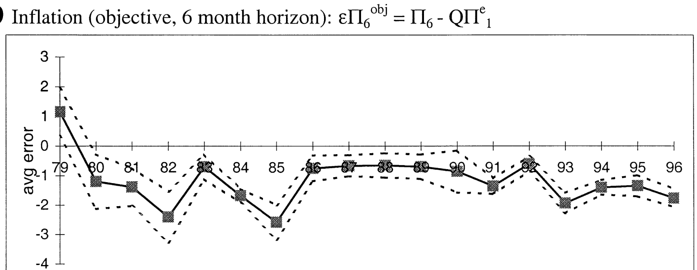
obs = 17650
$R^2 = .04$
F = 23, pval=0.00
$\mu = 1.53$ (.25)

C Inflation (objective): $\varepsilon\Pi^{obj} = \Pi_{12} - Q\Pi^e_1$



obs = 124724
$R^2 = .04$
F = 51, pval=0.00
$\mu = -0.50$ (.13)

D Inflation (objective, 6 month horizon): $\varepsilon\Pi_6^{obj} = \Pi_6 - Q\Pi^e_1$



obs = 125178
$R^2 = .03$
F = 29 , pval=0.00
$\mu = -0.95$ (.13)

FIG. 4. Time effects in forecast errors: continuous CAB variables. For variable and sample definitions, see the notes for Table 2. The graphed results come from an OLS regression of the continuous forecast errors ε on year dummies, correcting the standard errors for heteroscedasticity and cross-sectional correlation within the month. The middle line records the year effects, the outside dashed lines the 95% confidence intervals. *F* tests the joint significance of the year effects. To calculate the mean forecast error μ, the errors ε are regressed by OLS on a constant, correcting the standard errors for heteroscedasticity and cross-correlation within the month.

The magnitude of this decline is both statistically and economically significant, with inflation starting about two percentage points higher than expected in 1979 but falling to 2.5 percentage points lower than expected by 1982—a large, 4.5 percentage point change. More recently, throughout the 1990s households were repeatedly surprised by the low levels of inflation, by about 1–2 percentage points. Such negative errors dominate in the longer sample period, making the overall average error μ significantly negative for $\varepsilon\Pi^{obj}$, whereas it was positive for $\varepsilon\Pi^{subj}$ over the shorter sample period. These results vividly illustrate how sensitive estimates of bias can be to the sample period, even for long samples. Figure 4D shows the objective forecast errors $\varepsilon\Pi_6^{obj}$ using instead the CPI inflation rate over only the first six months after households' first interviews (annualized), to see the effects of the six-month mismatch between expectations and realizations in the other variables. Reassuringly, the results do not much differ from Figure 4C, suggesting that the mismatch is not driving the conclusions.[29] More generally, the results for inflation are robust across different definitions of inflation and its forecast error.

Figure 4A displays the forecast errors εGY for income growth, which are available only in the later part of the sample period. Despite the larger standard errors (reflecting the smaller sample size), the forecast errors still significantly vary over time. They start declining in 1990, and rebound only after 1993, when income growth was 2.5 percentage points lower than expected. Again, people seem to have been surprised by the recession, and perhaps also by the weakness of the subsequent recovery.[30,31]

As a further check that the six-month timing mismatch is not driving the results, forecast errors with the correct timing can be estimated for each household i. For each matched expectational question $Q^e_{1,i}$, the realization exactly *12 months later* $\hat{Q}^r_{12,i}$ can be estimated from the corresponding realizations Q^r_j of other households j with the same characteristics $\mathbf{Z}$ that are interviewed 12 months later. The resulting forecast errors $\varepsilon_{12} \equiv \hat{Q}^r_{12} - Q^e_1$ exhibit very similar means and time effects as those graphed in Figures 3 and 4A,B. The other conclusions given below also persist,

29. In addition to having similar time-series properties, the cross-sectional properties of $\varepsilon\Pi_6^{obj}$ are similar to those for $\varepsilon\Pi^{obj}$ in Table 2 below, both qualitatively and quantitatively. Further, their six-month difference does not materially change the Euler equation results for $\varepsilon\Pi^{obj}$ below in Table 4.

30. Though part of the reason εGY troughs as late as 1993 might be the small lag in its reference period, discussed above.

31. Again, these conclusions persist on dropping the overlapping forecast periods. The average errors μ for $\varepsilon\Pi^{subj}$ and $\varepsilon\Pi_6{}^{obj}$ remain significant in all six nonoverlapping samples. (For $\varepsilon\Pi^{obj}$ μ is less significant, but with a 12 month horizon, twice as much of the data (11/12) had to be dropped. Even so $\varepsilon\Pi^{obj}$ still varies significantly over time, in all the nonoverlapping samples.) For εGY μ remains significant in over half of the nonoverlapping samples. To allow for one month's delay in the release of the CPI, this analysis was redone dropping 6/7 of the data for $\varepsilon\Pi_6^{obj}$ and 12/13 for $\varepsilon\Pi^{obj}$. The conclusions are the same. As already noted, the serial correlation in the aggregated inflation errors lasts well over a year.

confirming that the timing mismatch is not a problem.[32,33] Further, the consistency of the results in Figures 3 and 4 suggests that the results in Figure 3 are not driven by the discreteness of its variables.

In sum, consumer forecasts appear to have been biased. However, it is very difficult to distinguish whether they were biased ex ante, or just ex post, requiring many years—perhaps many business cycles—to meet their targets on average. In either case, the bias is problematic for empirical studies with short sample periods. In particular, the business cycle and inflation regime induce low-frequency systematic patterns in average forecast errors.

Turning to the cross-sectional properties of forecast errors, the demographic variables **Z** were added to the models of the forecast errors ε using Equation (4), along with the full set of month dummies (but not yet interacting age and income by month). Table 2 records the results, starting with ordered probit models of the discrete errors in columns (1) to (4). The pseudo R^2's are small, implying that the forecast errors are largely unsystematic, as expected. Nonetheless, according to the chi-squared statistics the demographic variables are jointly very significant, for all four discrete errors. Hence the forecasts appear to have been inefficient. While it is difficult to interpret individual coefficients in this context, there are some interesting patterns. As regards financial position in column (1), the errors εFP tend to be more positive on average for older, higher income, and higher education households, more negative for divorcees and minorities. Since the overall average error μ was negative (Figure 3A), the bias in the forecasts QFP^e tends to decrease in magnitude with age, income, and education.

The pattern of results is roughly similar for business conditions εBC and income εY in columns (2) and (3), and often reversed in sign for inflation εP (which has the opposite coding) in column (4). Columns (5)–(7) show analogous results for the continuous income and inflation variables, estimated by OLS. In all cases, the demographic variables are again jointly quite significant, counter to the requirement of efficiency. They are also economically significant. For instance, in column (7), the inflation forecast error is about 0.4 percentage points larger in magnitude (more negative) for those without high school education, relative to those with high school education. The error is about 1.0 percentage point larger as real (1982–84 $) household income declines from $50,000 to $10,000, and for minorities and females relative to whites and males.

32. Because ε_{12} is continuous, the estimation is by OLS. While this changes the magnitude of the time effects compared to the ordered probit time effects for ε graphed in Figure 3, the differences are small even quantitatively. Overall, the differences in the time effects for ε_{12} versus ε are generally comparable in scope to the differences for $\varepsilon\Pi_6^{obj}$ versus $\varepsilon\Pi^{obj}$ graphed in Figures 4C, D. The cross-sectional properties of ε_{12} are also similar to those for ε in Table 2.

33. The similarity of the results for ε and ε_{12} also suggests that recall bias is not driving the conclusions, because ε and ε_{12} are calculated using different realization questions with only partly overlapping reference periods. Severe recall bias would imply little overlap between these realization questions. Further, recall bias is unlikely to be correlated with monetary policy, the business cycle, and skill-biased technical change, so is unlikely to explain the results in Figures 3–5. Finally, if the systematic components in the measured forecast errors simply reflected recall bias, not actual shocks, they should not help predict consumption changes below.

TABLE 2

CROSS-SECTIONAL HETEROGENEITY IN FORECAST ERRORS: CAB SURVEYS, 1978–96

	(1) $\varepsilon FP = QFP_2^r - QFP_1^e$		(2) $\varepsilon BC = QBC_2^r - QBC_1^e$		(3) $\varepsilon Y = QY_2^r - QY_1^e$		(4) $\varepsilon P = QP_2^r - QP_1^e$		(5) $\varepsilon GY = GY_2^r - QGY_1^e$		(6) $\varepsilon\Pi^{subj} = Q\Pi_2^r - Q\Pi_1^e$		(7) $\varepsilon\Pi^{obj} = \Pi_{12}^r - Q\Pi_1^e$	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Age	−0.018	0.003*	−0.010	0.003*	−0.016	0.005*	0.003	0.005	0.754	0.147*	−0.029	0.050	−0.037	0.013*
$Age^2/100$	0.020	0.003*	0.005	0.003#	0.018	0.005*	0.007	0.005	−0.675	0.142*	0.106	0.052*	0.071	0.013*
ln(income)	0.067	0.011*	−0.007	0.011	0.071	0.019*	−0.050	0.020*	−5.291	0.705*	−0.370	0.221#	0.539	0.053*
Married	−0.016	0.041	0.037	0.041	0.097	0.065	0.090	0.069	2.111	2.106	−0.226	0.609	−0.165	0.123
Separated	−0.074	0.026*	0.051	0.026*	−0.018	0.040	−0.027	0.043	−1.167	1.240	−0.067	0.416	0.123	0.114
Nonwhite	−0.119	0.026*	−0.031	0.025	−0.065	0.041	0.128	0.043*	−1.289	1.261	0.126	0.535	−0.815	0.140*
Female	0.005	0.023	−0.025	0.023	−0.001	0.036	0.146	0.038*	−1.360	1.075	1.032	0.338*	−0.970	0.098*
No high school	−0.021	0.023	−0.025	0.023	−0.008	0.035	0.038	0.037	−0.970	1.169	−0.337	0.391	−0.396	0.112*
Some college	−0.010	0.019	−0.043	0.019*	−0.015	0.031	−0.080	0.033*	1.234	0.816	−0.051	0.297	0.269	0.073*
College	0.057	0.019*	−0.032	0.019#	0.082	0.031*	−0.111	0.033*	1.303	0.837	−0.336	0.250	0.099	0.067
1 kid	−0.029	0.021	−0.023	0.021	−0.081	0.035*	−0.005	0.037	−0.634	0.963	−0.361	0.308	−0.152	0.086#
2 kids	−0.030	0.022	−0.018	0.022	−0.007	0.037	−0.034	0.039	0.194	0.962	0.017	0.341	−0.319	0.086*
3+ kids	−0.049	0.028#	−0.014	0.027	−0.047	0.046	−0.043	0.049	−0.416	1.186	−0.304	0.431	−0.504	0.120*
2 adults	0.037	0.038	−0.029	0.038	−0.105	0.061#	−0.007	0.065	−0.007	1.945	0.822	0.552	0.097	0.108
3+ adults	0.017	0.044	−0.022	0.044	−0.075	0.072	−0.054	0.076	−1.093	2.173	1.271	0.624*	0.026	0.128
Midwest	0.018	0.020	0.053	0.020*	−0.023	0.033	−0.038	0.035	0.012	0.844	−0.532	0.295#	0.116	0.081
South	0.019	0.020	0.042	0.020*	0.019	0.032	−0.024	0.034	−1.167	0.860	−0.215	0.306	0.001	0.081
West	−0.043	0.022*	0.025	0.022	−0.055	0.036	−0.081	0.038*	−1.577	0.939#	−0.531	0.328#	−0.237	0.086*
log likelihood	−31068		−32366		−11252		−8919							
# obs	23798		23775		9295		9405		4856		8788		60695	
Pseudo R^2	0.01		0.05		0.01		0.04							
R^2									0.03		0.05		0.05	
χ^2 [pval]	228	[0.00]	158	[0.00]	74	[0.00]	401	[0.00]						
F [pval]									5.52	[0.00]	9.51	[0.00]	35.3	[0.00]

NOTES: This table analyzes the systematic demographic components of forecast errors ε, using Equation (4). In columns (1) to (4), the forecast errors are discrete and the estimation is by ordered probit. For each matched pair of sentiment questions, Q_1^e represents an expectation from a household's first interview, Q_2^r the corresponding realization from its second interview. Except for inflation, $Q = +1$ represents the better states (e.g., better or good), 0 the intermediate states, −1 the worse states. Forecast errors are the difference between expectations and realizations: $\varepsilon = (Q_2^r - Q_1^e)$, with values in {−2,−1,0,1,2}. In columns (5)–(7), the forecast errors are constructed analogously but are continuous-valued; estimation is by OLS, correcting the standard errors for heteroscedasticity. Coefficients on month dummies are not shown. The omitted categorical variables are: single, white, male head, high school graduate, no kids, one adult, northeast. χ^2 and F test the joint significance of the demographic variables, with p-values in the brackets. In columns (1)–(6) the sample is limited to households interviewed twice. Further, the same person (either head or spouse) must have been the respondent in both interviews. In column (7), which uses the actual CPI inflation rate Π_{12} as the realization, the household may have been interviewed only once; and if interviewed twice, the respondent need not be the same person in both interviews. # represents significance at the 10% level, * represents significance at the 5% level.

Whether one should interpret these results as evidence of "irrationality" is a subtle issue. It could be that young, low income, and low education people have perfectly rational expectations ex ante, but ex post happened to have received disproportionately bad shocks over the sample period. This is consistent with the literature finding increased inequality over the period, in part due to skill-biased technical change (e.g., Cutler and Katz, 1991, Attanasio and Davis, 1996). But even the ex post interpretation of the results is problematic for empirical studies that assume that time dummies capture all systematic components of forecast errors. Further, the inefficiency of the forecasts of aggregate variables (QBC, QP, and QΠ) is harder to explain, and more likely represents ex ante inefficiency. Even if people receive different shocks to their own income and financial position, household-specific shocks should have less effect on their forecasts of aggregate economic activity and prices.

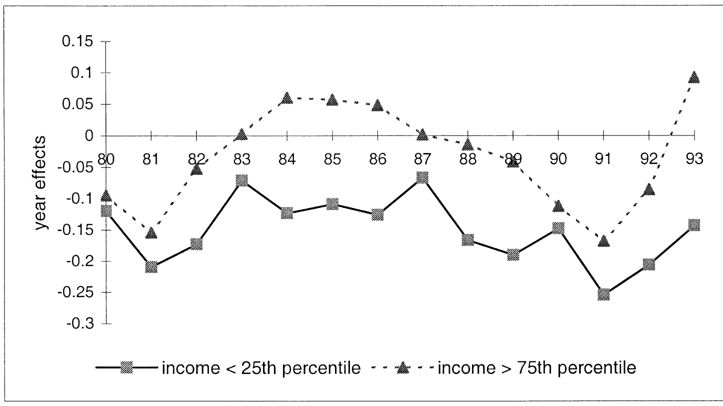
The cross-sectional distribution of forecast errors can change over time. To illustrate, Figure 5 shows the sample average of the errors in financial position εFP, year-by-year for different demographic groups. Since income and age are the variables interacted with time below, the figures contrast the histories of the top and bottom quartiles of the income and age distributions. In Figure 5A for income, the errors are always more negative for low-income households than for high-income households, though they are more cyclical for the high-income households. One interpretation is that during the expansions high-income households received relatively good shocks but low-income households continued to receive somewhat negative shocks on balance, consistent with ongoing skill-biased technical change. These results go beyond most of the literature on technical change by implying that the increased inequality was repeatedly unexpected, year after year, which has additional welfare consequences. In Figure 5B for age, the errors for young households are both more negative and more cyclical than for older households. This suggests that long-run and business cycle shocks disproportionately hit young households.

5. RESULTS: EXCESS SENSITIVITY AND SYSTEMATIC HETEROGENEITY IN FORECAST ERRORS

Even if expectations are not fully rational, they might still help forecast spending. To test for excess sensitivity of consumption to sentiment, the sentiment variables $\hat{Q}$ were first imputed into the CEX using an OLS regression of Equation (1). For brevity, these results are not reported but are available in the working-paper version.[34] In this first-step, most of the demographic variables were significant, and jointly they were very significant. In Table 3, column (1) shows the resulting adjusted R^2's from the first-step regressions. More of the level of sentiment is explained than that of the associated forecast errors (in Table 2), as expected. The dynamic variables

34. The working paper reported the first-step results using ordered probit models for the discrete sentiment questions. Those results are qualitatively similar to the OLS results.

A by Income



B by Age

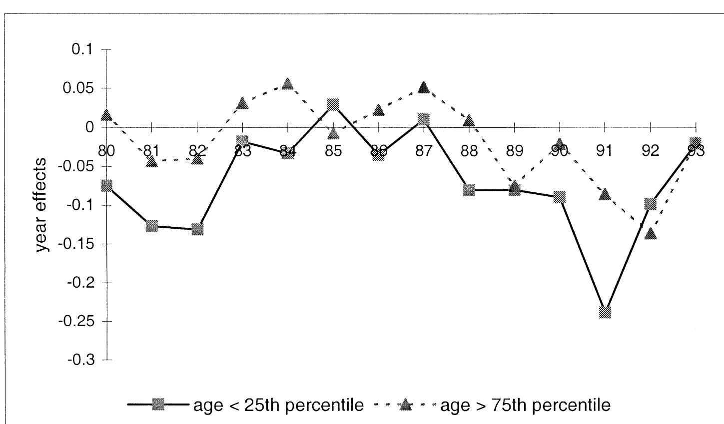


FIG. 5. Time effects in forecast errors, by demographic groups: financial position εFP. The figures graph the sample means of the error εFP in financial position (See Table 2.), for the bottom and top quarters of the income and age distributions.

in Equation (1), namely the month dummies and their interactions with age and income, were always significant.[35] The "static R^2's" in brackets in column (1) come from redoing the estimation without the dynamic variables. For all the household-specific variables (QFP^r, QFP^e, QY^e, and QGY^e), the static R^2 is well over half

35. To ease the computational demands, the quadratic term in age has been dropped. Preliminary analysis suggested that for most sentiment questions the quadratic term did not vary as significantly across time.

the size of the original R^2, suggesting that while the dynamic variables help explain some of the variation in sentiment, the static demographic variables in **Z** are themselves quite important. The static R^2's for the aggregate variables (QBC, QBC5, QDP, QP^e, QU^e, and $Q\Pi^e$) are relatively smaller. Not surprisingly, respondents' expectations of aggregate variables vary less with their own (head's) demographic characteristics than do their expectations of their own financial position and income; i.e., the aggregate variables contain relatively less cross-sectional variation.

Given $\hat{Q}$ one can then estimate Euler equation (3). The resulting excess sensitivity coefficients b_2 appear in columns (2) and (3) of Table 3, for both nondurable and total consumption.[36] Over half of the coefficients are significant, counter to the PIH. While the coefficients are usually larger in magnitude for total consumption, they are generally as significant for nondurable consumption. The signs on b_2 are always negative, except for inflation and unemployment for which the coding was reversed. Thus, in all cases, the better states are associated with less steep consumption profiles; that is, higher confidence is associated with less saving. This outcome is consistent with precautionary motives for saving (e.g., Deaton, 1992, Carroll, 1992, Lusardi, 1998) as well as increases in expected future resources.[37]

TABLE 3

Excess Sensitivity of Consumption to Sentiment: CEX, 1982–93

	Independent variable	(1) 1st Stage R^2 [static R^2]	(2) $\Delta\ln(\text{nondurables})_{t+1}$ Coefficient	(2) S.E.	(3) $\Delta\ln(\text{total consumption})_{t+1}$ Coefficient	(3) S.E.
(1)	QFP^r_t	0.10 [.08]	−0.0134	0.0046*	−0.0210	0.0062*
(2)	QFP^e_t	0.11 [.10]	−0.0376	0.0103*	−0.0376	0.0141*
(3)	QBC_t	0.13 [.02]	−0.0079	0.0050	−0.0109	0.0068#
(4)	$QBC5_t$	0.06 [.03]	−0.0148	0.0042*	−0.0173	0.0056*
(5)	QDP_t	0.07 [.02]	−0.0049	0.0065	−0.0073	0.0086
(6)	QY^e_t	0.11 [.10]	−0.0190	0.0046*	−0.0278	0.0063*
(7)	QP^e_t	0.05 [.01]	0.0049	0.0173	0.0448	0.0243#
(8)	QU^e_t	0.09 [.01]	0.0132	0.0084	0.0050	0.0112
(9)	QGY^e_t	0.08 [.08]	−0.0010	0.0007	−0.0021	0.0010*
(10)	$Q\Pi^e_t$	0.03 [.02]	0.0017	0.0008*	0.0032	0.0011*
	#obs		97993		97874	

Notes: This table tests for excess sensitivity of consumption to sentiment. Each row-column cell in columns (2) and (3) represents a separate regression of Euler Equation (3) in the CEX. The sentiment variables Q are the predicted values from a first-step OLS regression of Equation (1) in the CAB, with adjusted R^2 as shown in column (1). The static R^2 is for the same regression without time dummies and their interactions with demographic characteristics. Increases in Q represent worse states for inflation and unemployment (rows 7,8,10). In all other rows, increases in Q represent better states. Demographic control variables W and month dummies are not shown. Standard errors are corrected for heteroscedasticity and serial correlation by household, as well as the generated regressors. # represents significance at the 10% level, * represents significance at the 5% level.

36. The coefficients on the demographic variables **W** in Equation (3) are similar to those in related studies using the CEX, e.g., Souleles (1999), and so are not reported. In short, the coefficients on changes in family size are generally positive; the coefficients on age are less significant.

37. It remains unclear whether people's answers to the sentiment questions (other than QY and QGY) reflect expected future *uncertainty* or expected future *levels* of income and other resources. But Carroll, Fuhrer, and Wilcox (1994) show that the aggregate ICS reflects more than just the level of expected income. Also, as already noted, outside the PIH a flatter consumption profile need not necessarily imply less saving.

Most of the insignificant excess sensitivity coefficients are for questions referring to aggregate variables: QBC, QDP, QP^e, and QU^e. In part, this is the result of their having less cross-sectional variation, conditional on the time dummies, as evidenced by their smaller first-step static R^2's. Conversely, almost all the household-specific variables generate significant excess sensitivity.[38] Thus, the cross-sectional information in sentiment appears to help predict future consumption.

There are many possible sources of this excess sensitivity. One possibility is unobserved differences in discount factors and other household fixed effects. Following Runkle (1991), lagged consumption growth from households' first interview was added to Equation (3) to control for household fixed effects, at the cost of reduced sample size and power. Nonetheless, over half of the significant coefficients for nondurable consumption in Table 3 remain significant, including QFP^e, QBC5, and QY^e. While QFP^r and $Q\Pi^e$ become less significant, QP^e becomes more significant.[39] Hence, heterogeneous discount factors and other fixed effects cannot alone be generating the estimated excess sensitivity. Further, Hausman tests and autocorrelation tests of the residuals produce little evidence for the presence of fixed effects, consistent with the previous literature (Browning and Lusardi 1996).[40]

Another possible, but understudied, explanation for the results is time-varying, systematic heterogeneity in forecast errors. This is especially likely to be a problem since both sentiment and forecast errors have just been found to be correlated with the same household demographic characteristics. These findings suggest that even a long sample period and a full set of time dummies might not be enough to ensure orthogonality of the forecast errors in the residual with the sentiment regressors. Since the forecast errors are likely to be correlated with many regressors of interest, this would be a general problem.

To verify this suggestion directly, estimates of the forecast errors $\hat{\varepsilon}$ were added to Euler equation (3), for the variables for which there are matching realization and expectation questions. Table 4 shows the results, after imputing the forecast

38. Even though QFP^r does not ask about the future, time-series studies have similarly found the coincident component of the aggregate ICS index (QFP^r +QDP) to be useful in forecasting (e.g., Throop 1992).

39. These results do not correct the standard errors for the generated sentiment regressors, because the Angrist and Krueger (1992) estimator requires that the independent variables in Equation (3) be available in the first-step dataset, but the CAB does not measure consumption growth.

40. Most studies assume that preferences are identical across agents, in which case the time dummies control for the net discount factors $[(r\text{-}\rho)_t]$. Omitted fixed effects should lead to positive autocorrelation in the residuals of individual households' consumption growth. However, in all regressions in Tables 3 and 4 the residuals are negatively correlated at both the first and second household lags. The Hausman test is motivated by the fact that adding lagged consumption growth provides consistency in the presence of fixed effects, but inefficiency in their absence. In 17 of the 20 Euler equations in Table 3, the Hausman test fails to find evidence for fixed effects. Also, the three exceptions are all for sentiment questions regarding aggregate variables (QDP, QP, $Q\Pi$), yet it would be surprising if households' discount factors were more correlated with aggregate sentiment questions than with household-specific sentiment questions. (Further, the excess-sensitivity coefficients for two of these exceptions, QDP and QP, are already insignificant in Table 3, before adding lagged consumption growth. Hence their Hausman test results do not imply that failure to control for fixed effects generated any spurious excess sensitivity.) Similarly, in Table 4, most of the Hausman tests fail to find evidence for fixed effects.

TABLE 4

EXCESS SENSITIVITY AND SYSTEMATIC HETEROGENEITY IN FORECAST ERRORS: CEX, 1982–93

	Independent variables	(1) $\Delta\ln(\text{nondurables})_{t+1}$ Coefficient	S.E.	(2) $\Delta\ln(\text{total consumption})_{t+1}$ Coefficient	S.E.
(1)	QFP^e_t	−0.0374	0.0105*	−0.0375	0.0142*
	$\varepsilon FP_{t+1}(=QFP^r_{t+1}-QFP^e_t)$	0.0186	0.0075*	0.0105	0.0101
	# obs		97993		97874
(2)	QBC^e_t	−0.0065	0.0089	0.0091	0.0119
	$\varepsilon BC_{t+1}(=QBC^r_{t+1}-QBC^e_t)$	0.0186	0.0072*	0.0115	0.0100
	# obs		97993		97874
(3)	QY^e_t	−0.0031	0.0098	−0.0087	0.0135
	$\varepsilon Y_{t+1}(QY^r_{t+1}-QY^e_t)$	0.0269	0.0158#	0.0255	0.0219
	# obs		29528		29504
(4)	QP^e_t	0.0020	0.0342	0.0318	0.0450
	$\varepsilon P_{t+1}(=QP^r_{t+1}-QP^e_t)$	−0.0124	0.0148	−0.0033	0.0200
	# obs		29528		29504
(5)	QGY^e_t	−0.0047	0.0017*	−0.0045	0.0023*
	$\varepsilon GY_{t+1}(=QGY^r_{t+1}-QGY^e_t)$	0.0014	0.0004*	0.0010	0.0005*
	# obs		30292		30255
(6)	$Q\Pi^e_t$	−0.0010	0.0015	−0.0001	0.0020
	$\varepsilon\Pi_{t+1}(=Q\Pi^r_{t+1}-Q\Pi^e_t)$	−0.0024	0.0013#	−0.0035	0.0018*
	# obs		29528		29504

NOTES: This table tests whether the excess sensitivity in Table 3 is due to systematic heterogeneity in forecast errors. See the notes to Table 3. Each row-column cell represents a separate regression of Euler Equation (3) in the CEX. The forecast errors $\varepsilon = Q^r_{t+1} - Q^e_t$ are first estimated in the CAB using OLS regressions of Equation (4), then imputed in the CEX. Demographic control variables W and month dummies are not shown. Standard errors are corrected for heteroscedasticity and serial correlation by household, as well as the generated regressors. # represents significance at the 10% level, * represents significance at the 5% level.

errors into the CEX using Equation (4) (now interacting age and income by month): $\hat{\varepsilon}_{t+1} \equiv \widehat{\hat{Q}^r_{t+1} - \hat{Q}^e_t}.$ Despite the time dummies in Equation (3), the coefficients b_3 on the errors $\hat{\varepsilon}$ are often significant. Except for inflation, when they are significant they are positive: positive innovations in financial position, income, etc., are correlated with increases in consumption, as expected. For the inflation questions QP and QΠ, with the opposite coding, the coefficients are negative. But even controlling for the forecast errors, the excess sensitivity regressor b_2 remains significantly negative for two of the household-specific variables, QFP^e and QGY^e (in rows (1) and (5)). That is, some excess sensitivity persists and so is not due to heterogeneity in forecast errors alone.[41] On the other hand, b_2 has become insignificant for the third household-specific variable QY^e, as well as for QBC^e, QP^e, and $Q\Pi^e$. Hence some, though not all, of the excess sensitivity appears to be due to systematic heterogeneity in forecast errors. This suggests the possibility that previous excess sensitivity tests might have made spurious inferences.

41. Of course, it is possible that even the remaining excess sensitivity is spurious, due to other systematic heterogeneity in forecast errors that matters for consumption but is not controlled for by the available sentiment variables (or due to other sources of mispecification, such as intertemporal nonseparability or liquidity constraints). However, shocks to income, financial position, aggregate economic activity and prices must be among the most important sources of innovations to consumption. Further, as already noted, under the null hypothesis that forecast errors are classical, they should be orthogonal to other factors in agents' information sets, including discount rates. Hence the effect in Table 4 of adding forecast errors to the Euler equation cannot be due to such factors.

The forecast errors were also computed by first separately imputing the realizations and expectations $\hat{Q}^{\mathrm{r}}_{t+1}$ and $\hat{Q}^{\mathrm{e}}_{t}$ in the CEX using Equation (1), and then taking their difference: $\hat{\varepsilon}_{t+1} = \hat{Q}^{\mathrm{r}}_{t+1} - \hat{Q}^{\mathrm{e}}_{t}$. The results are generally similar, and appear in the working paper.[42]

6. CONCLUSION

This paper provided perhaps the first comprehensive analysis of the household data underlying the Michigan Index of Consumer Sentiment. This data allowed for a cleaner test of the rationality of consumers' expectations than in most previous studies. The results can also be interpreted as characterizing the time-series and cross-sectional properties of the shocks that hit different types of households over time. Expectations appear to have been biased, at least ex post, in that forecast errors did not average out even over the long sample period lasting almost 20 years. This bias is not constant over time; it is related to the inflation regime and the business cycle. People underestimated the disinflation of the early 1980s and in the 1990s and generally appear to underestimate the severity of business cycles. Expectations are also inefficient, in that people's forecast errors were correlated with their demographic characteristics. That is, forecast errors are systematically heterogeneous. The results suggest an important role for time-varying, group-level shocks—aggregate shocks do not hit all people equally. For instance, during recent expansions high-income households received relatively good shocks but low-income households continued to receive somewhat negative shocks on balance, consistent with ongoing, unexpected skill-biased technical change. Whether one interprets these results as evidence of ex ante irrationality or not, they are problematic for empirical studies that have short sample periods or assume that time dummies control for all systematic components of forecast errors. Empirical implementations of forward-looking models need to recognize that forecast errors are more complex than usually assumed.

Attention then turned to whether the sentiment data helps predict household expenditure. Significant evidence of excess sensitivity was found, counter to the PIH. Higher confidence was correlated with less saving, consistent with precautionary motives and increases in expected future resources. Further, this paper provided a unique test of the specific alternative hypothesis that systematic heterogeneity in forecast errors explains the rejection of the PIH. Previous studies, lacking explicit measures of these errors, have not been able to consider this hypothesis directly. Demographic components of forecast errors were found to explain some, though

42. When the realization questions Q^{r} are not available for much of the sample, the change in the estimated expectational variable was used instead: $\varepsilon_{t+1} = \hat{Q}^{\mathrm{e}}_{t+1} - \hat{Q}^{\mathrm{e}}_{t}$. A change in expectations over time still represents an innovation. The results are somewhat less significant than those reported in Table 4, perhaps because the imputed variables $\hat{Q}$ do not vary enough across quarters. Still, $\hat{\varepsilon}$ is significantly positive for QY^{e}. Also the excess sensitivity coefficients b_2 for QFP^{e} and QGY^{e} remain significant, and now are significant for QY^{e}. Again the excess sensitivity coefficients are generally insignificant for the aggregate variables, with the exception of $Q\Pi^{\mathrm{e}}$.

not all, of the excess sensitivity. More broadly, because forecast errors are correlated with household demographic characteristics, they will be correlated with many regressors of interest in forward-looking models, suggesting that nonclassical forecast errors are in practice a general problem. Finally, the cross-sectional variation in sentiment, net of time dummies, was itself found to be informative. This is information lost in the aggregated ICS time series for sentiment; nor is it contained in other macro variables used in forecasting. Of the Michigan survey questions, those asking specifically about the household, rather than the aggregate economy, were generally found to contain the most useful cross-sectional information.

This analysis can be extended in a number of ways. First, given the significance of the cross-sectional distribution of sentiment, new sentiment time-series might be created to better incorporate this distribution, for instance by taking weighted averages of sentiment across households.[43] Second, one can similarly examine many other economic decisions in addition to spending for which expectations matter, such as portfolio choice. Cross-sectional data is especially well suited to studying the effects of one-time events, like the 1987 stock market crash. Third, durables purchases might be modeled more explicitly, taking into account their discreteness.

APPENDIX A: THE DATA

A.1 The CAB Survey

The additional sentiment questions, not part of the aggregate ICS index, include the following. The allowed responses are in brackets (underlining in original).

QBC^r. Would you say that at the present time business conditions are better or worse than they were a year ago? [better now, about same, worse now]

QBC^e. And how about a year from now, do you expect that in the country as a whole, business conditions will be <u>better</u> or <u>worse</u> than they are at present, or just about the same? [better a year from now, about same, worse a year from now]

QY^r. During the last year or two, would you say that your (family) income went up more than prices, went up about the same as prices, or went up less than prices? [more, same, less]

QY^e. During the next year or two, do you expect that your (family) income will go up <u>more than prices</u> will go up, <u>about the same</u>, or <u>less than prices</u> will go up? [more, same, less]

QP^r. During the <u>last 12 months</u>, have prices of the things you buy remained unchanged, or have they gone up, or have they gone down? [gone up, remained unchanged, gone down]

QP^e. During the <u>next 12 months</u>, do you think that <u>prices in general</u> will go up, or down, or stay where they are now? [go up, will not go up, go down]

43. The weights could reflect e.g., the scale of spending by different groups of people, or the sensitivity of their spending to their sentiment.

QU^e. How about people out of work during the coming 12 months—do you think that there will be more unemployment than now, about the same, or less? [more, about same, less]

QGY^r. [The growth rate is computed from changes in the level of income taken from the following question:] Now, thinking about your (family's) total income from all sources (including your job), how much did you (your family) receive in [the previous calendar year]?

QGY^e. By what percent do you expect your (family) income to (increase/decrease) during the next 12 months?

$Q\Pi^r$. By about what percent do you think prices have gone (up/down) on the average, during the last 12 months?

$Q\Pi^e$. By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

Other answers such as "Don't Know" are also allowed, but are not used here. When the answers to $Q\Pi$ and QGY were topcoded, they were not used.

For CAB interviews that took place in more than one installment, if these installments spanned two different calendar months, the second month is used to date the observation. If any demographic variable used in a regression is missing, topcoded, or flagged (e.g., Don't Know), the observation is not used. For the demographic variables $\mathbf{Z}$, when the continuous measure of total household income was missing, the midpoint of the bracketed income variable was used instead. (But the bracketed variable is not used in computing the growth rate of income.) The reference period for realized income (used in computing the growth rate QGY^r) is the previous calendar year, whereas for the CEX it is the past 12 months. For consistency, CAB income was deflated using the CPI (1982-84 $) for the past 12 months. Since the original CAB income variable is constrained to be positive, for consistency total income in the CEX was used only when positive and not flagged. Sample selection is discussed in the text.

A.2 The CEX Survey

In aggregating individual expenditures, if any component of total consumption or nondurable consumption was topcoded or missing its cost, the whole consumption group was set to missing. If any component was missing its date or dated before the reference period, the group was dropped for all interviews for the household at issue. A large number of expenditures are dated in the month of the interview. Following the recommendation of the staff at the BLS, for consistency such expenditures were accrued to the following reference period.

In addition to the sample restrictions in the text, an observation is dropped if the age of the head increases by more than one year, or decreases, on moving into the next quarter. An observation is also dropped if the age of any other member changes in this way and thereby results in the member's switching between being a kid (less than 16 years old) and an adult (at least 16). If any variable used in a regression is missing, the observation is not used. Other sample restrictions are described in the text.

LITERATURE CITED

Acemoglu, Daron, and Andrew Scott (1994). "Consumer Confidence and Rational Expectations: Are Agents' Beliefs Consistent with the Theory?" *Economic Journal* 104, 1–19.

Alessie, Rob, and Annamaria Lusardi (1997). "Saving and Income Smoothing: Evidence from Panel Data." *European Economic Review* 41, 1251–1279.

Angrist, Joshua, and Alan B. Krueger (1992). "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." *Journal of the American Statistical Association* 87, 328–336.

Attanasio, Orazio P., and Steven J. Davis (1996). "Relative Wage Movements and the Distribution of Consumption." *Journal of Political Economy* 104, 1227–1262.

Attanasio, Orazio P., and Guglielmo Weber (1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 95, 1121–1157.

Batchelor, Roy (1986). "Quantitative v. Qualitative Measures of Inflation Expectations." *Oxford Bulletin of Economic and Statistics* 48, 99–120.

Batchelor, Roy, and Lars Jonung (1989). "Cross-Sectional Evidence on the Rationality of the Mean and Variance of Inflation Expectations." In *Understanding Economic Behavior*, edited by Klaus G. Grunert, pp. 93–105. Boston: Kluwer Press.

Browning, Martin, and Annamaria Lusardi (1996). "Household Saving: Micro Theories and Micro Facts." *Journal of Economic Literature* 34, 1797–1855.

Burch, Susan W., and Stephen E. Gordon (1985). "The Problem of Inference in Consumer Surveys." Federal Reserve Board of Governors working paper.

Carroll, Christopher D. (1992). "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity* 2, 61–156.

Carroll, Christopher D., Karen E. Dynan, and Spencer D. Krane (1996). "Balance Sheets and Precautionary Saving." Federal Reserve Board of Governors working paper, January.

Carroll, Christopher D., Jeffrey C. Fuhrer, and David W. Wilcox (1994). "Does Consumer Sentiment Forecast Household Spending? If So, Why?" *American Economic Review* 84, 1397–1408.

Chamberlain, Gary (1984). "Panel Data." In *Handbook of Econometrics*, edited by Zvi Griliches and Michael D. Intriligator, pp. 1248–1318. Amsterdam: North-Holland.

Croushore, Dean (1998). "Evaluating Inflation Forecasts." Federal Reserve Bank of Philadelphia working paper.

Cutler, David M., and Lawrence F. Katz (1991). "Macroeconomic Performance and the Disadvantaged." *Brookings Papers on Economic Activity* 1–61.

Das, Marcel, Jeff Dominitz, and Arthur van Soest (1999). "Comparing Predictions and Outcomes: Theory and Application to Income Changes." *Journal of the American Statistical Association* 94, 75–85.

Das, Marcel, and Arthur van Soest (1999). "A Panel Data Model for Subjective Information on Household Income Growth." *Journal of Economic Behavior and Organization* 40, 409–426.

Deaton, Angus (1992). *Understanding Consumption*. Oxford, UK: Oxford University Press.

Dominitz, Jeff (1998). "Earnings Expectations, Revisions, and Realications." *Review of Economics and Statistics* 82, 374–388.

Dynan, Karen E. (1993). "How Prudent Are Consumers?" *Journal of Political Economy*, 101, 1104–1113.

Flavin, Marjorie (1991). "The Joint Consumption/Assets Demand Decision: A Case Study in Robust Estimation." NBER Working Paper No. 3802.

Friend, Irwin, and F. Gerard Adams (1964). "The Predictive Ability of Consumer Attitudes, Stock Prices, and Non-Attitudinal Variables." *Journal of the American Statistical Association* 59, 987–1005.

Gramlich, Edward M. (1983). "Models of Inflation Expectations Formation." *Journal of Money, Credit, and Banking* 15, 155–173.

Hall, Robert E., and Frederic S. Mishkin (1982). "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica* 50, 461–482.

Jappelli, Tullio, Jorn-Steffen Pischke, and Nicholas S. Souleles (1998). "Testing for Liquidity Constraints in Euler Equations with Complementary Data Sources." *The Review of Economics and Statistics* 80, 251–262.

Jappelli, Tullio, and Luigi Pistaferri (2000). "Using Subjective Income Expectations to Test for Excess Sensitivity of Consumption to Predicted Income Growth." *European Economic Review* 44, 337–358.

Keane, Michael P., and David E. Runkle (1990). "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." *American Economic Review* 80, 714–735.

Ludvigson, Sydney C. (1996). "Consumer Sentiment and Household Expenditure: Reevaluating the Forecasting Equations." Federal Reserve Bank of New York working paper.

Lusardi, Annamaria (1996). "Permanent Income, Current Income and Consumption: Evidence from Two Panel Data Sets." *Journal of Business and Economic Statistics* 14, 81–90.

Lusardi, Annamaria (1998). "On the Importance of the Precautionary Saving Motive." *American Economic Review* 88, 449–453.

Maddala, G.S., Raymond P.H. Fishe, and Kajal Lahiri (1983). "A Time Series Analysis of Popular Expectations Data." In *Applied Time Series Analysis of Economic Data*, edited by Arnold Zellner, pp. 278–290. Washington, DC: U.S. Department of Commerce.

Manski, Charles F. (1990). "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis." *Journal of the American Statistical Association* 85, 934–940.

Mariger, Randall P., and Kathryn Shaw (1993). "Unanticipated Aggregate Disturbances and Tests of the Life-Cycle Consumption Model Using Panel Data." *The Review of Economics and Statistics* 75, 48–56.

McNeil, John (1974). "Federal Programs to Measure Consumer Purchase Expectations, 1946–1973: A Post-Mortem." *Journal of Consumer Research* 1, 1–10.

Nicholson, Sean, and Nicholas S. Souleles (2002a). "Physician Income Prediction Errors: Sources and Implications for Behavior." University of Pennsylvania working paper.

Nicholson, Sean, and Nicholas S. Souleles (2002b). "Physician Income Expectations and Specialty Choice." University of Pennsylvania working paper.

Runkle, David E. (1991). "Liquidity Constraints and the Permanent-income Hypothesis." *Journal of Monetary Economics* 27, 73–98.

Shapiro, Matthew D., and Joel Slemrod (1995). "Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding." *American Economic Review* 85, 274–283.

Souleles, Nicholas S. (1999). "The Response of Household Consumption to Income Tax Refunds." *American Economic Review* 89, 947–958.

Souleles, Nicholas S. (2001). "Household Securities Purchases, Transactions Costs, and Hedging Motives." University of Pennsylvania working paper.

Thomas, Lloyd B. Jr. (1999). “Survey Measures of Expected U.S. Inflation.” *Journal of Economic Perspectives* 13, 125–144.

Throop, Adrien W. (1992). “Consumer Sentiment: Its Causes and Effects.” Federal Reserve Bank of San Francisco. *Economic Review* 0, 35–59.

Zeldes, Stephen P. (1989). “Consumption and Liquidity Constraints: An Empirical Investigation.” *Journal of Political Economy* 97, 305–346.