

The Usefulness of Consumer Sentiment: Assessing Construct and Measurement

Paul M. Kellstedt*
Texas A&M University

Suzanna Linn
Penn State University

Lee Hannah
Penn State University

Abstract

Given the scholarly and popular prominence of the concept of consumer confidence, it is striking that there are no examinations of the quality of the most commonly used measure of the concept—the University of Michigan’s Index of Consumer Sentiment. This study fills that gap, evaluating both the reliability and validity of the measure. Our analyses lead us to question the reliability of the ICS, as well as the validity of the Index itself (but not its components) as a measure of the concept of consumer confidence.

*The authors can be reached at kellstedt@pols.tamu.edu, slinn@la.psu.edu, and lee.hannah@psu.edu, respectively.

Americans' subjective evaluations of the economy's recent trajectory, and their expectations about its likely future prospects—a concept referred to as *consumer sentiment* or *consumer confidence*—have become a key ingredient in the ever-evolving business of predicting the future of the economy, and also of the politicians entrusted with managing it. Much like we eagerly await the release of monthly reports on jobs or inflation, the monthly readings of consumer confidence are widely reported in the news and discussed extensively by the commentariat class. As often as not, shifts in consumer confidence are described as “surprising” or “unexpected.” The information contained in the monthly sentiment figures is viewed by some as valuable enough that certain traders are willing to pay \$4,500 a year to have early access to a leading measure of consumer confidence—the University of Michigan's Index of Consumer Sentiment (Antilla, 1991). Further, scholarly evidence continues to mount that consumer sentiment influences both evaluations of politicians, public liberalism, trust in government, and election outcomes (Durr, 1993; Erikson, MacKuen and Stimson, 2002; Keele and Kelly, 2006; MacKuen, Erikson and Stimson, 1992). The implication, of course, is that surveys of consumer confidence contain meaningful information for politics and for the economy itself.

And yet the utility of survey measures of consumer confidence as either reflective or predictive measures of consumer behavior in the marketplace is unclear. Although the survey results are widely anticipated and reported as harbingers of future spending, many argue that “the surveys act more like mirrors than crystal balls” (Cohen, 1995). Economist Alan Garner, for example, argues that the index is mainly reflective of what is going on in consumers' lives at the moment but has no predictive power (Garner, 1991). Further, Morin (1992) found that different component measures of sentiment—for example, evaluations of personal versus national economic conditions—can sometimes “move in opposite directions, sending exactly the opposite message.” When the U.S. Department of Commerce announced that it was revising its Index of Leading Economic Indicators to include consumer evaluations, the addition was contested by many economists (Morin, 1992). Scholarly research is also divided on the question of whether sentiment contains information independent of economic conditions useful for predicting consumer behavior,

such as spending.¹

Regardless of the competing views of the sentiment survey measures, it is still reported on with faith that it is telling us *something* about the economy and may be a causal determinant of spending, saving, and investment. In August of 2010, for example, a consumer confidence report gave investors “the good news they were looking for” as stock indexes climbed (*Major Indexes Reverse Opening Losses, Moving Higher on Positive Consumer Sentiment*, 2010). The following month, a drop in the stock market was widely attributed to a discouraging consumer sentiment report “[that] tempered recent signs of economic growth and upbeat earnings” (Bernard, 2010). In fact, many economists continue to interpret these measures as having consequences for the economy (Trumbull, 2010).

In what follows, we assess the usefulness of consumer sentiment as a construct and a measure (or measures) in order to identify the best way to consider the role of sentiment in consumer behavior. This brings us to a consideration of fundamental questions about the reliability and validity of consumer sentiment measures. Our purpose is to provide evidence on this score. We proceed as follows. First, we describe the historical origins of the concept and measures of consumer sentiment, and then we proceed to describe how the concept has been measured. Next, we examine two forms of reliability of the University of Michigan’s Index of Consumer Sentiment (ICS), evaluating first how the ICS and its component measures are related (internal reliability), and second how the Conference Board’s Consumer Confidence Index (CCI) and ICS are related as time unfolds (parallel form reliability). We then turn to an evaluation of the validity of the ICS, with a particular focus on the comparative evidence for the ICS as a whole relative to that for the individual components taken together. We ask which measures best forecast consumer spending (content/predictive validity). We conclude with prescriptive advice for practitioners who use measures of consumer confidence in their models.

¹See, for example: (Adrangi and Macri, 2011; Blanchard, 1993; Carroll, Fuhrer and Wilcox, 1994; Desroches and Gosselin, 2002; Gelper, Lemmens and Croux, 2007; Hall, 1993).

The Origins of Consumer Sentiment as a Concept and a Measure

Consumer sentiment is a concept born out of the field of behavioral economics at its modern-day (post-World War II) inception. George Katona (for example, see Katona 1975), along with other pioneers in the field the likes of Nobel Laureates Herbert Simon (1951; 1955) and Daniel Kahneman (Kahneman and Tversky, 1973), challenged the prevailing view in economics that humans are rational beings exhibiting optimizing behavior like automatons (Hosseini, 2003; Katona, 1975). Broadly, they argued that economic processes result from human behavior—that the motives, attitudes, and expectations of human beings influence their behavior.² In 1944, in an effort to demonstrate the importance of consumers’ attitudes for spending and saving decisions, Katona convinced the Board of Governors at the Federal Reserve to add attitudinal questions to their surveys of consumers.³ Katona directed the survey, which was prepared and fielded by the University of Michigan’s Survey Research Center, and used these data to demonstrate a “clear relationship” between attitudes and consumer behavior (Likert, 1972). This early survey has become the Michigan Survey of Consumers, the data from which the widely anticipated Michigan Index of Consumer Sentiment (ICS) is calculated.⁴

Katona made a simple argument: Consumers make decisions about whether to spend (now or later) on discretionary goods based both on their ability to purchase the goods and their “willingness” to do so. Income and asset assessments determine ability, and assessments of future income determine willingness. The two together, then, should determine spending, in particular on durable goods.⁵ Additionally, because assessments of the economic future are uncertain, the consumer faces some risk associated with a poor estimate in the wrong direction: too little savings, too much debt, mortgage foreclosure, an inability to purchase necessities, and so on.

²Simon’s development of the concept of *bounded rationality* was an important advance in behavioral economics. Kahneman’s role (with Tversky) in the development of psychological economics, in particular prospect theory, was also central to the development of the behavioral economics field.

³Katona argued that survey respondents were much more likely to reveal information about their incomes and spending habits if they were first asked questions about their evaluations of their own and national economic conditions.

⁴Richard Curtin now directs the Survey of Consumers.

⁵Non-discretionary purchases cannot be easily postponed and therefore should be inelastic to assessments of future income. Purchases of durable goods can often be delayed until consumers feel less risk is associated with reducing savings or borrowing to make purchases.

Uncertainty about the future should then inhibit spending by raising its risk. Each of these mechanisms of influence suggests an important role for *both* economic *and* noneconomic information in determining spending decisions. Precisely what those noneconomic factors are has received little attention (but see De Boef and Kellstedt 2004), but they might include the decisions and (in)actions of political actors—legislative stalemate, government shutdowns, and uncertain election outcomes, as well as widespread views of malfeasance of those at the top of the political economy; when those we hold responsible for the performance of the economy cannot produce economic growth and are mired in politics, there is little reason for consumers to be optimistic. Consumer attitudes are also likely to be influenced by unexpected events whose economic effects are unpredictable and are not quickly reflected in economic indicators: wars, terrorism, natural disasters (De Boef and Kellstedt, 2004).⁶

Measurement Details: The Index of Consumer Sentiment and the Consumer Confidence Index

In our analysis we focus primarily on the University of Michigan’s Index of Consumer Sentiment (ICS) because it is by far the most prominent measure used in the literature, though we also consider, for comparative purposes, the Conference Board’s Consumer Confidence Index (CCI).⁷ The ICS is calculated from respondents’ assessments of their personal finances (retrospective and prospective), national economic performance (one and five years out), and (current) buying conditions and is designed to capture “optimism and certainty” or “pessimism and uncertainty” felt by consumers (Curtin, 2008).⁸ The second measure of overall sentiment is the Consumer Confidence Index (CCI), distributed by The Conference Board each month. It is based on answers to five questions asking respondents for evaluations of job prospects and business conditions, both now and in the future, as well as family income.

⁶Published analyses of consumer sentiment itself provide little evidence on this score, but instead focus on how economic indicators move sentiment. These analyses suggest the existence of an endogenous relationship among consumer attitudes and behavior.

⁷A more detailed comparison of the two indexes can be found in Ludvigson (2004).

⁸The survey also contains questions about expectations for future interest rates, unemployment, news heard by respondents about the economy, prices, as well as investment decisions and evaluations of government handling of the economy, though these questions are not a part of the Index.

Specifically, the Index of Consumer Sentiment (ICS) is based on responses to the 5 questions given in column 2 of Table 1, which are asked in a nationally representative telephone survey of 500 respondents who are re-interviewed 6 months later by the University of Michigan Survey Research Center. Roughly 60% of the sample in any month is interviewed for the first time, with the balance representing re-interviewed individuals who had completed the surveys in previous months. The marginals are released on the second Friday every month, and are revised and re-released the fourth Friday of the same month. The Survey began annually in the 1940s, became quarterly in 1958, and has been administered monthly since 1978. For each of the five indicators, a series is created by summing the percentage of positive responses, subtracting the percentage of negative responses, and adding 100. The ICS, then, averages these individual “diffusion” values, weighting each indicator equally. Finally, the results are converted to the base period and 2 is added to correct for sample design changes.⁹

[Table 1 About Here]

We present the Index and its components in Figure 1. The ICS is given in the top left-hand corner. The series is seldom in territory above the 100 “neutral” point, averaging about 87 (standard deviation 12.4), with exceptions in the latter half of the 1990s and early 2000s, before taking its noted and lengthy slide into some of its lowest values—but low values that are not *exceptionally* low by historical standards. There is an upswing at the end of the sample period, but it is too soon to tell what this movement signifies. Moving right is the graph of personal retrospective evaluations, which vacillates around the 100 mark, averaging just above 107 (standard deviation 15.1), but like the ICS, after early 2000 it begins an almost continuous slide downward, reaching 60 before turning upward and reaching 91 in early 2012. Prospective personal evaluations (first column, row two) exhibit the least variance of the components, by more than half (standard deviation 10.1), and tends to be relatively positive with a mean of 122. In contrast, to its right, one-year-out business expectations move about the most, with a standard deviation

⁹The ICS also releases two sub-indices: The Index of Current Conditions (ICC) uses the first 2 questions from Table 1. The Index of Consumer Expectations (ICE) is based on the last 3 questions from the same table. See “Report on Index Value Calculations” at <http://www.sca.isr.umich.edu/> for more details and the questionnaire.

over 32 and mean over 109. They share the pattern seen in all the series, with more negative evaluations beginning shortly after the election of George W. Bush in 2000. These evaluations began to improve after the election of Barack Obama in 2008, but began to falter as quickly as the others. In the bottom row on the left are longer-term forward-looking business evaluations (five years out). These are less positive—Americans’ optimism does not extend far into the future. Mean evaluations drop below the 100 “neutral” mark to 92.7, with a standard deviation of approximately 20. The now-common pattern of decline in evaluations after 2000 is again apparent, but with a more sawtooth-like pattern. Finally, on the bottom right, are assessments of whether now is a good time to make a purchase of major appliances. These are durable goods, a point not lost on Katona, the very things we can hold off buying when times are bad. It appears, though, that consumers did not decide that times were disproportionately poor for making these purchases; this index did not drop below 100 often in its history, and not until the very end of our sample period in the recent crisis. In fact this component series has the highest mean, 140, and a relatively high standard deviation at 19 points.

[Figure 1 About Here]

The Consumer Confidence Survey has been conducted since 1967 for the Conference Board.¹⁰ Questionnaires are mailed each month to a nationwide representative sample of 5,000 households, with a typical response of 3,500 completed surveys.¹¹ The Consumer Confidence Index (CCI) is based on the 5 questions in column 3 of Table 1. For each question, the number of positive responses is divided by the sum of the number of positive and negative responses and multiplied by 100. Each diffusion index is converted to a base-year index and then averaged, weighting each equally.¹² The Conference Board releases index values on the last Tuesday of each month and revises them at the end of the following month. Data exist bi-monthly from

¹⁰TNS conducted the survey until September, 2011. The Nielsen Company replaced TNS in February, 2011. Analyses of data from the overlapping period suggest that the transition had no effect on the continuity of the data.

¹¹The questionnaire is available with membership in The Conference Board. See <http://www.conference-board.org/data/consumerconfidence.cfm> and select the “Technical Notes.”

¹²Like the University of Michigan, the Conference Board also releases two sub-indices: The Present Situation Index is created in the same method as the CCI using the first two questions; the Index of Consumer Expectations from the final three.

1966-1977, after which it becomes available monthly.¹³

The Usefulness of the ICS as a Measure

While inferences about the role of consumer sentiment are debated, the wider success of behavioral economics, Katona's premise, as well as the proliferation of measures and attention of the press, suggest that it is worth asking a set of questions that has attracted little attention. How reliable are different measures of sentiment, and how valid is any given measure of sentiment?

The usefulness of a construct stems from its ability to tell us something new—that is, information that could not be gleaned from other sources. As conceived by Katona, consumer sentiment captures the human element in consumption and savings decisions, which he deemed necessary for understanding these decisions. Therefore, it should both provide information independent of economic conditions and predict the behavior of consumers in the marketplace. These are issues of both construct and measure. Therefore, in order to assess the usefulness of consumer sentiment, we need to return to the very basic questions of the reliability and validity of measures of consumer sentiment.

The twin concerns of a measure's reliability and validity—that a construct is captured accurately—are of paramount importance. Without some assurance on these counts, the measure itself cannot be useful. Reliability—being free from unsystematic types of error (Didow and Franke, 1984), being stable, dependable, trustworthy, and consistent (Worthen, Borg and White, 1993)—is not typically examined in time series analysis. Validity—capturing the concept of interest—is also infrequently considered. In part this is because, in time series analysis, we seldom have multiple measures of a concept. But we can tackle these issues, in the first case because we have multiple measures of sentiment, and in the second because it is conceptually well-defined so that we know what it should predict and how.

Our assessments of reliability and validity begin with the assumption that consumer

¹³See <http://www.conference-board.org/>.

sentiment is a unidimensional construct—“willingness to consume”—and that the Index of Consumer Sentiment (and, separately, the Consumer Confidence Index) measures that construct. We see how far this takes us before we examine the unique behavior of ICS components and their relationship to the Index of Consumer Sentiment. Such a procedure can be justified by appealing to the almost exclusive use of the overall Indices in a) published research, b) media attention, and c) Katona’s own work building the Index of Consumer Sentiment.¹⁴

Reliability

The reliability of a measure, at its core, revolves around how *repeatable* a measure is. In what follows, we examine two distinct conceptions of reliability: the more familiar internal reliability, and parallel-form reliability.

Internal Reliability

As a time series and as an index built of 5 component time series, one important aspect of the reliability of the ICS deals with its repeatability in terms of the relationship between the indicators and the Index over time. That is, the more reliable the ICS is, the more that it will be a function of the same components, *to the same degrees*, across time. In a perfect world, each measure of sentiment would repeatedly, over time, capture the construct in the same way. If, by contrast, the components of the ICS contribute to the Index to highly varying degrees—that is, with high correlations to the Index at some points in time, and lower correlations to the Index at other points in time—then to that extent, the Index is less reliable.

Formally, this type of reliability is termed internal consistency, the extent to which the individual series that compose an index, here the ICS, are related to each other. An index is internally consistent if each indicator measures the underlying construct and is therefore strongly correlated with the ICS. The usual course of action, in the cross-sectional world, is to look at a

¹⁴It is true that only the prospective components of the ICS are included in the Leading Economic Indicators index, but the purpose of the LEI is forward looking while that of the ICS is to assess willingness to consume more generally.

correlation matrix or compute an industry-standard measure of reliability (like Cronbach’s alpha) to assess internal consistency. However, these assume away the question at hand by assuming constant correlations across time. Thus we assess internal consistency using a model that estimates dynamic (or conditional) correlations, the Dynamic Conditional Correlation Generalized Autoregression Conditional Heteroscedasticity (DCC-GARCH, hereafter DCC) model. The DCC model estimates the conditional correlation between the two time series as a function of both the previous variability of each series and the previous correlation between them (Engle and Sheppard, 2003; Lebo and Box-Steffensmeier, 2008).¹⁵

We are interested in two sets of information from the DCC models. The first is the estimates of the conditional correlations themselves. These are extracted from the model given the estimated parameters and reveal the path the correlations follow over the time period for which the Index is available. They provide visual evidence about the extent to which the Index is a function of each indicator and the degree to which that relationship changes over time. The second set of information is the two DCC parameters. The first of these, β , gives us a sense of the persistence in the correlations—how much the relationship between the Index and the indicator varies over time. The closer to 1.0, the more persistent the conditional correlations, and the more consistent the relationship between the Index and the component series over time. The second DCC parameter, α , tells us how much “news” impacts the path of the correlations. Smaller news parameters are consistent with a more reliable Index. If $\beta = \alpha = 0$, the conditional correlations equal the unconditional correlations, and the correlations are constant.¹⁶

The simple question of whether the Index is a function of the components to the same degrees across time does not have a simple answer. The evidence is reported in Table 2. All five of the Index components show evidence of dynamic correlations, and thus have “higher

¹⁵Briefly, the estimation process proceeds in two steps. In the first step, univariate mean and variance models are fit for the ICS and each component to remove the effects of the past on the current observation and to produce estimates of the (conditional) volatility within the series. Conditional correlations are estimated using the standardized residuals from the univariate models as estimates in a time varying correlation matrix. Thus the estimated dynamic correlations are purged of that portion that is due simply to the shared dynamic patterns. The models were fit in R using the rugarch and rmgarch packages (Ghalanos, 2013).

¹⁶Such a finding would lead us to conclude that the Index is highly reliable, but these are not the only empirical results that could lead to such an inference.

correlations at some times and not others,” suggesting concerns about the reliability of the Index. All of the Index components exhibit high degrees of persistence in their correlations with the ICS, as evidenced by the $\hat{\beta}$ coefficients of 0.80 or higher. The correlations evolve over time, but they do so quite slowly. The impact of news is unsurprisingly (and of necessity) insignificant substantively and also statistically in most cases. The correlations between the ICS and its components do not appear to respond much to new information but rather to evolve slowly. We have, then, evidence that the conditional correlations are just that: conditional.

[Table 2 About Here]

But graphing the correlations—see Figure 2—provides additional information. It allows us to see which ICS–component pairs have the strongest correlations, when correlations in the pairs are higher and lower, and to get a sense of the significance of the over-time dynamics. Expectations about future business conditions—both one- and five-year ahead—consistently exhibit the highest conditional correlations with the ICS, but they do not move in lock step.

Statistical tests tell us that we can reject the null hypothesis that the correlations between these ICS–component pairs are constant,¹⁷ but there exists no formula for computing the standard error of the conditional correlations. So to get a sense of the significance of the over-time variation in the correlations, we place a confidence interval around the estimate of the *constant* correlation between the series pairs.¹⁸ Doing so quickly reveals that the range over which the conditional correlations move is constrained most of the time to be within a 95% confidence interval of the estimated constant correlation, giving us good reason to feel reassured about the internal reliability of the Index. Still, there are times when four of the indicators—personal retrospective and prospective evaluations, one- and five-year ahead business expectations—exhibit a unique relationship with the Index. In all four cases the biggest changes occur in the earlier time period. Notably, in the most recent period, when the conditional

¹⁷For all 5 pairings the null hypothesis of constant correlations can be rejected at $p < .05$ for a variety of lag lengths.

¹⁸The constant correlation is estimated from the filtered time series using the mean and variance models that served as the basis for the DCC GARCH model.

variance in the indicators is lower, the conditional correlations have grown.

[Figure 2 About Here]

Note also that, in matters of degree, the estimated constant (average) correlations reported in Figure 2 are in fact very high, perhaps surprisingly so given that these correlations are estimated from data that are filtered so that they reflect the relationship between innovations in the series pairs. The average of the estimated dynamic correlations with the Index range from a low of $r^2 \approx 0.66$ for personal retrospective evaluations to $r^2 \approx 0.91$ for one-year-out business expectations.

The tests performed above to assess internal reliability are quite stringent and therefore, while we cannot assign a reliability score to the ICS, we find the evidence for internal reliability to be quite persuasive.

Parallel-form Reliability

The existence of a second measure of consumer confidence, the Conference Board's Consumer Confidence Index (CCI), affords us a second opportunity to assess the *repeatability* of the ICS. To the extent that the two measures of consumer confidence measure, in fact, consumer confidence, they should exhibit a high degree of correlation and do so consistently over time. This type of reliability is formally referred to as parallel-form reliability. We can assess the reliability of both measures of consumer confidence by looking at the time varying correlations between the ICS and CCI.

The evidence for parallel-form reliability is not strong.¹⁹ The estimates of the DCC parameters are both significant. The dynamic correlation between the ICS and CCI is persistent ($\hat{\beta} = .59$), but much less so than between the ICS and its components. The DCC estimate of the

¹⁹We fixed the estimates of the autoregressive and (G)ARCH coefficients in this shorter time period to match those of the full time period so as to use the more complete information about the underlying dynamics of the ICS available in the longer time series, while allowing a different estimate of the mean. This explains the absence of standard errors for these estimates in the final column of the table.

news parameter, $\hat{\alpha}$, is also more than double that estimated in the previous pairings. New information affects the ICS and CCI differently, so that their correlation changes systematically over time—evidence against reliability. We can see this in the graph of the estimated dynamic correlations as well (see the bottom-right panel of Figure 2). While the average correlation is similar to the others we have estimated, the range is much wider over the shared time period of analysis. The dynamic correlations range from 0.41 to 0.90 in this case, while in all other cases the range was less than half that. The correlations also frequently cross the confidence interval around the estimated unconditional correlation. In other words, the two indexes are not related to each other in the same way over time. This is consistent with other research noting that the CCI relies uniquely and disproportionately on measures affected by unemployment Ludvigson (2004).

Assessing the Validity of the ICS

Judgments about the validity of a measure like the ICS traditionally involve assessments of face validity, content validity, and predictive (or construct) validity. On its face, the ICS is clearly valid. As we noted in the introduction, reporting on the monthly releases of the ICS is a common and routine feature of newscasts, and profit-seeking organizations are willing to pay for early access to the survey results. This suggests that the measure contains useful information about the beliefs of consumers and is, on the face of it, valid.

Our evaluation of the content validity of the ICS is mostly positive. If the underlying construct includes a consumer's beliefs about both his or her personal situation, and the situation of business conditions in the country as a whole, and that the consumer's beliefs about both personal and national economic circumstances have both retrospective and prospective components, then consumer sentiment neatly forms a 2X2 grid. The five items that comprise the Index, displayed in Table 1, show that three of the four cells are covered in the Index: personal/retrospective (item 2), personal/prospective (item 5), and business/prospective (items 3 and 4). The index does not contain the respondent's opinion about how business conditions have improved or deteriorated in the recent past.²⁰ Moreover, the Index includes an item that is not

²⁰This is despite the fact that the Survey of Consumer Attitudes and Behaviors actually includes such a question.

clearly a part of the 2X2 matrix—the question about whether or not now is a good time to buy household items. It is, therefore, perhaps not surprising that the household-items question has the lowest correlation with the underlying index (see Table 3).²¹

[Table 3 About Here]

Of course, the most crucial validity check for any measure is that of construct (or predictive) validity. We make that determination next. Together these analyses will help us to assess the predictive validity of the ICS and the index components. Specifically, it will help us to identify more precisely how (and if) the different economic evaluations provide information about consumer behavior independent of economic conditions.

Construct (or Predictive) Validity

Consumer spending, we know, depends on objective economic conditions. Positive economic signals lead to more spending, weaker signals to less. Consumer sentiment, in theory, captures something distinct or, at the very least, more—the *willingness* to spend. The well-developed theoretical construct embodied in consumer sentiment offers a clear prediction: Consumer sentiment should predict consumer spending, conditional on objective information about the economy. This suggests a straightforward test of construct validity.

Our strategy is to first estimate a model of consumer spending growth that is saturated by economic indicators capturing both current conditions and leading indicators, and to save the residuals from that model—in effect, purging consumer spending of these other economic factors. Next, we model these residuals—the portion of consumer spending that cannot be attributed to other economic forces—as a function of the ICS and of its constituent parts. If consumer sentiment predicts spending above and beyond the effects of other economic forces, then we have

The wording is: “Would you say that at the present time business conditions are better or worse than they were a year ago?”

²¹On the other hand, the dynamic correlations presented in Figure 2 show that the relationship between the household-items question and the Index is the least volatile.

strong evidence of its construct validity. Comparing the performance of the ICS to the Index components allows us to assess whether the Index is more (or less) than the sum of its parts.

We begin this exercise in Table 4, where we present block F-tests for a variety of economic causes of consumer spending that have been shown to affect consumer spending. Specifically, we regress consumer spending growth on durable goods on contemporaneous values and four lags of each of the Conference Board's Index of Lagging Economic Indicators, growth in disposable personal income, the quarterly inflation rate, the 3-month Treasury Bill rate, and growth in unemployment. For the sake of brevity, rather than presenting the individual coefficient estimates, we instead display block F-tests, representing the probability that each variable does not predict consumer spending. There is little surprise in these results.²²

[Table 4 About Here]

The residuals from the saturated model estimated here should be understood as the portion of consumer spending growth on durable goods that cannot be attributed to objective economic factors. In effect, we have controlled for economic conditions, and we have done so in such a way as to maximize the ability of economic conditions to predict durable goods spending growth. As a test of construct validity, the bar is thus set high.

[Table 5 About Here]

Column A of Table 5 presents estimates of the effects of the ICS on consumer spending (net of economic reality), allowing 2 lags of spending growth to enter the model.²³ Two features of this model are noteworthy. First, the estimated effect of the ICS is essentially zero, and is, in any case, statistically insignificant. Second, the r^2 is small, with most of the variance accounted for by the previous two quarters' spending growth.

A different picture emerges, though, when the ICS is decomposed into its component parts. Those results are presented in Column B. In that model, 4 of the 5 Index components are

²²The Lagging Economic Indicator Index contains unemployment, so the absence of a significant block F-test on growth in unemployment is not particularly surprising.

²³These results are robust to the specification of GARCH dynamics. Results presented in the the Appendix.

statistically significantly related to growth in durable goods spending, and in ways that are largely intuitive. In particular, when people feel it is a good time to buy durable goods, they do so. Expectations about the national economy one and five years out also affect spending on durable goods. In the first case, as optimism about the next 12 months increases, people spend more in the present. Controlling for these attitudes, expectations for brighter conditions five years down the road lead people to hold off spending, so that growth in durable goods consumption drops in the current period, perhaps because they believe the better time to buy appliances or computers, for example, is in the future. Expectations about one's personal financial situation in the year ahead are not significantly related to growth in durable goods consumption, and retrospective evaluations about personal finances over the last year are negatively related to current durable spending growth. This last finding is somewhat perplexing.²⁴

Taken individually *and* as a collective, at least with respect to consumer spending on durable goods, the multi-indicator ICS predicts less well than do its components. The model in column B explains more of the variance in spending growth in durable goods, and the RMSE is smaller. It is important to note that the index components do not individually predict spending growth. The effects are conditional. Willingness to consume appears to be a complex construct, but—at least in this analysis—one that is better captured by the inclusion of multiple indicators than by the inclusion of the Index created from those indicators. In this particular sense, the Index of Consumer Sentiment is less than the sum of its component parts.

The magnitude of these effects is not trivial. A standard-deviation change in expectations regarding the state of the national economy in the next year (approximately 32 points) is associated with a shift in durable goods spending *after accounting for economic conditions* by approximately three quarters of a percent (annualized) from the previous period. In our view, this is a substantial effect. The remaining indicators exhibit smaller standard deviations and ranges, and, combined with their effects in Table 5, somewhat smaller impacts on spending. But it is an open (and subjective) question as to how much variation measures of consumer sentiment *should* explain. We know that we have set the bar high, but how far over it a measure should rise

²⁴The colinearity of the Index components offers a potential explanation.

in order to be a valid indicator of the construct requires a subjective response.

Conclusions

On March 1, 2013, a widely printed headline featured the Index of Consumer Sentiment: “U.S. Michigan Consumer Sentiment Rises Above Estimate.” The article went on to state that “Confidence among American households rose more than projected in February in a gain that could boost consumer buying, the largest part of the economy.” Stories like these that link sentiment to spending (and economic conditions more broadly) have been a regular feature of media reports on the state of the economy for 50 years. Political fortunes, too, are linked to consumer sentiment by the media. Just prior to the election, NBCNews.com reported that “U.S. consumer sentiment unexpectedly rose to its highest in five years in October in the latest in a string of encouraging signs from the economy that may boost President Barack Obama’s re-election hopes next month.” Scholarly research, too, has found evidence that consumer sentiment influences the economy and the fortunes of political incumbents.

Consumer sentiment is clearly a meaningful concept, particularly as a dynamic concept. There are times when the public is optimistic about the economy, and times when it is pessimistic. Absent any information on objective economic indicators such as inflation, unemployment, income growth, and the like, consumer sentiment predicts spending. Consumer sentiment ebbs and flows with our knowledge of economic history, most recently by dipping to some of its lowest historic values during the recession that ushered in President Obama.

But how useful is the most prominent measure of consumer sentiment to the scholar or investor trying to predict the future or to politicians trying to turn around a sagging economy? We have broached this question by examining the Index of Consumer Sentiment and its component series, considering theoretical foundations, measurement, and in particular the evidence for the reliability and validity of the construct and measures.

As in most exercises of this sort, the evidence is qualified. First and foremost, the ICS and

its component measures do seem to capture the same things in the same way over most of their history, as indicated by the dynamic conditional correlations. If the ICS captures “willingness to consume,” it has done so consistently since its inception. This has been true during periods of recession and expansion, during political and economic turmoil, and during elections. This is really quite remarkable. The two indicators measuring expectations about the future of the national economy are more strongly related to the Index in an absolute sense than are the remaining three series, but all are highly correlated with the Index (and each other), even after filtering the time-series dynamics from each series. It’s also the case that the dynamic correlations have become stronger over time, although we cannot explain why.

Yet, the components of the ICS do not, in the strict sense, relate to the Index in the same way over time. Those relationships are all dynamic. Parsing out systematic causes of these dynamics—and possibly using that information to generate a new index—is an exercise that awaits future analysis. Equally troubling are the results of the parallel-form reliability analysis. The dynamic relationship between the ICS and the next-most-prominent measure of consumer sentiment—the CCI—is particularly striking for its substantial volatility.

Whether the ICS captures “willingness to consume” is less clear from our analysis. We have created a difficult test, to be sure, giving economic conditions every chance to determine spending before assessing the predictive utility of consumer sentiment. But we have done so with spending on durable goods—precisely the type of spending where we expect sentiment to matter most. Economic conditions anchor spending decisions, and when their effects are removed we find that knowing current sentiment as measured by the ICS would add nothing to our crystal ball. By contrast, the component series taken collectively contain useful information about spending behavior. In short, the Index is less useful, statistically, than are the component pieces. But so, too, is any given indicator less than the five considered as a set. This suggests the possibility that a better sentiment index could be created, assuming our goal is to predict durable goods spending. This is assuredly only one way a sentiment index can be useful.

In the end, we advise practitioners against the uncritical use of the ICS as a composite

measure in their analyses, and prescribe instead that analysts consider using some subset of the component indicators, depending on the theoretical question at hand. In general, when speaking theoretically about how “consumer confidence” might influence some other concept, analysts might be able to be more specific, and refer instead to how “optimism about the economic future” or “positive recollections about the direction of the economy” might be a more specific substitute. In at least this one analysis—and, we speculate, perhaps in others—the whole of the ICS is something less than the sum of its component parts.

References

- Adrangi, Bahram and Joseph Macri. 2011. "Consumer confidence and aggregate consumption expenditures in the United States." *Review of Economics and Finance* 2011(1):1–18.
- Antilla, Susan. 1991. "Consumer statistics hot on Wall Street." *USA Today* p. 8B.
- Bernard, Stephen. 2010. "Stocks fluctuate after drop in consumer sentiment." *Associated Press Online* p. Business News.
- Blanchard, Olivier. 1993. "Consumption and the recession of 1990-1991." *American Economic Review* 83(2):270–274.
- Carroll, Christopher D., J.C. Fuhrer and D.W. Wilcox. 1994. "Does consumer sentiment forecast household spending - If so, why?" *American Economic Review* 84(5):1397–1408.
- Cohen, Sarah. 1995. "Consumer surveys don't always give expert opinions." *St. Petersburg Times* p. 1E.
- Curtin, Richard. 2008. *Economic Discontent: Causes and Consequences*. Technical report University of Michigan.
- De Boef, Suzanna L. and Paul M. Kellstedt. 2004. "The Political (and Economic) Origins of Consumer Confidence." *American Journal of Political Science* 48(4):633–633.
- Desroches, Brigitte and Marc-Andre Gosselin. 2002. "The usefulness of consumer confidence indexes in the United States." Working Paper.
- Didow, Nicholas M. and George Franke. 1984. "Measurement Issues in Time-Series Research: Reliability and Validity Assessment in Modeling the Macroeconomic Effects of Advertising." *Journal of Marketing Research* 21(1):12–19.
- Durr, Robert H. 1993. "What Moves Policy Sentiment?" *American Political Science Review* 87:158–70.

- Engle, Robert F. and Kevin Sheppard. 2003. Theoretical Properties of Dynamic Conditional Correlation Multivariate GARCH. Technical report New York University and Oxford University.
- Erikson, Robert S., Michael B. MacKuen and James A. Stimson. 2002. *The Macro Polity*. New York: Cambridge University Press.
- Garner, C. A. 1991. "Forecasting Consumer Spending: Should Economists Pay Attention to Consumer Confidence Surveys?" *Economic Review - Federal Reserve Bank of Kansas City* 76(3):57.
- Gelper, S., A. Lemmens and C. Croux. 2007. "Consumer sentiment and consumer spending: Decomposing the Granger causal relationship in the time domain." *Applied Economics* 39:1–11.
- Ghalanos, Alexios. 2013. *rmgarch: Multivariate GARCH models*. R package version 1.2-0.
- Hall, R.E. 1993. "Macro theory and the recession of 1990-1991." *The American Economic Review* 83(2):275–279.
- Hosseini, Hamid. 2003. "The arrival of behavioral economics: from Michigan, or the Carnegie School in the 1950s and the early 1960s?" *Journal of Socio-Economics* 32(4):391–409.
- Kahneman, Daniel and Amos Tversky. 1973. "On the Psychology of Prediction." *Psychological Review* 80(4):237–51.
- Katona, George. 1975. "BUSINESS TRENDS AHEAD: Why Consumer Attitudes Matter More Than Economics." *Nation's Business (pre-1986)* 63(9):34.
- Keele, Luke J. and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* Forthcoming.
- Lebo, Matthew J. and Janet M. Box-Steffensmeier. 2008. "Dynamic Conditional Correlations in Political Science." *American Journal of Political Science* 52(3):688–704.

- Likert, Rensis. 1972. Courageous Pioneer: Creating A New Field of Knowledge. In *In Human Behavior in Economic Affairs: Essays in Honor of George Katona*, ed. B. Strumpel, J. Morgan and E. Zahn.
- Ludvigson, Sydney C. 2004. "Consumer confidence and consumer spending." *The Journal of Economic Perspectives* 18(2):29–50.
- MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *American Political Science Review* 86:597–611.
- Major Indexes Reverse Opening Losses, Moving Higher on Positive Consumer Sentiment.* 2010. *Midnight Trader* .
- Morin, Richard. 1992. "A matter of consumer confidence, but does it really matter?" *The Washington Post* p. A19.
- Simon, Herbert A. 1951. *Models of Man*. New York: Wiley.
- Simon, Herbert A. 1955. "A behavioral model of rational choice." *The Quarterly Journal of Economics* 69(1):99–118.
- Trumbull, Mark. 2010. "Consumer confidence: Americans pessimistic, economists aren't; Economists say sluggish growth and sagging consumer confidence aren't signs of another recession but indicate that the nation's financial recovery will be slow going." *The Christian Science Monitor* .
- Worthen, Blain R., Walter R. Borg and Karl White. 1993. *Measurement and evaluation in the schools*. New York: Longman.

Table 1: Consumer Sentiment Survey Data

University of Michigan Survey of Consumer Attitudes and Behavior	Conference Board Consumer Confidence Survey
Retrospective Evaluations	
1. Do you think now is a good or bad time for people to buy major household items? [good time to buy;uncertain, depends;bad time to buy]	How would you rate present general business conditions in your area? [good;normal;bad]
2. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago? [better;same;worse]	What would you say about available jobs in your area right now? [plentiful;not so many;hard to get]
Prospective Evaluations	
3. Now turning to business conditions in the country as a whole. Do you think that during the next twelve months, we'll have good times financially or bad times or what? [good times; uncertain;bad times]	Six months from now, do you think business conditions in your area will be [better;same;worse]?
4. Looking ahead, which would you say is more likely? That in the country as a whole we'll have continuous good times during the next five years or so, or that we'll have periods of widespread unemployment or depression, or what? [good times;uncertain;bad times]	Six months from now, do you think there will be [more;same;fewer] jobs available in your area?
5. Now looking ahead do you think that a year from now, you (and your family living there) will be better off financially, or worse off, or just about the same as now? [better;same;worse]	How would you guess your total family income to be six months from now? [higher;same;lower]

Table 2: DCC GARCH Estimates for the Index of Consumer Sentiment with its Components (Quarterly, 1960-2012Q3); Index of Consumer Sentiment and Consumer Confidence Index (Quarterly, 1977-2012Q3)

	Personal Retropsections	Personal Expectations	One Year National	Five Year National	Durable Purchasing	Consumer Confidence
ICS						
Constant	87.52** (4.82)	87.52** (4.78)	87.52** (5.05)	87.52** (4.84)	87.52** (4.75)	86.26 ** (4.34)
AR(1)	0.92** (0.03)	0.92** (0.03)	0.92** (0.03)	0.92** (0.03)	0.92** (0.03)	0.92 -
a	0.07 (0.05)	0.07 (0.05)	0.07 (0.05)	0.07 (0.05)	0.07 (0.05)	0.07 -
b	0.74** (0.10)	0.74** (0.10)	0.74** (0.10)	0.74** (0.10)	0.74** (0.10)	0.74 -
Component						
Constant	111.22** (5.46)	126.52** (3.85)	118.23** (14.13)	95.40** (7.47)	144.77** (6.16)	104.89** (2.58)
AR(1)	0.71** (0.08)	0.61** (0.07)	0.93** (0.03)	0.91** (0.03)	0.71** (0.07)	0.64** (0.16)
AR(2)	0.22** (0.08)	0.31** (0.07)			0.21** (0.07)	0.29+ (0.15)
MA(1)						0.58** (0.13)
a	0.20+ (0.11)	0.15+ (0.09)	0.11+ (0.06)	0.05 (0.05)	0.01 (0.03)	0.40** (0.13)
a_2					0.23 (0.17)	
b	0.49** (0.14)	0.69** (0.20)	0.73** (0.07)	0.84** (0.06)		0.41** (0.09)
DCC Parameters						
α	0.05 (0.04)	0.02 (0.01)	0.06** (0.02)	0.04 (0.06)	0.01 (0.02)	0.15* (0.09)
β	0.80** (0.18)	0.98 ** (0.01)	0.90** (0.03)	0.93** (0.19)	0.96** (0.07)	0.59+ (0.31)
Shape Parameter						
	8.27** (2.76)	9.03** (3.07)	10.18** (3.35)	10.91** (4.03)	9.57** (3.36)	
Constant Correlation						
r^2	0.66	0.76	0.91	0.83	0.70	0.76

Model is fit in R using the rmgarch and rugrarch packages (Ghalanos, 2013).

$h_t = c_i + a_i \epsilon_{t-1}^2 + b_i h_{t-1}$. $R_t = (1 - \alpha - \beta)\bar{R} + \alpha \epsilon_{t-1} \epsilon_{t-1}' + \beta R_{t-1}$. R_t is $k \times k$ time varying correlation matrix. \bar{R} is the unconditional covariance matrix. h_t contains the time varying standardized deviations. An ** indicates the coefficient is statistically significant at the .01 level, * at the .05 level, and + at the .10 level.

Table 3: Correlation Matrix for Index of Consumer Sentiment and its Components, Quarterly 1960Q1-2012Q3

	ICS	Personal Retrospections	Personal Prospections	One Year Business	Five Year Business	Durable Purchases
ICS	1.0000					
Personal Retrospections	0.9098	1.0000				
Personal Prospections	0.8943	0.8063	1.0000			
One Year Business	0.9246	0.7958	0.8176	1.0000		
Five Year Business	0.9148	0.7477	0.8533	0.9078	1.0000	
Durable Purchases	0.8094	0.7695	0.6463	0.5720	0.5937	1.0000

See Table 1 for the question wording for each component series.

Table 4: The *Economic* Causes of Consumer Spending

Block of coefficients	<i>p</i> -Value for Block F-tests
Index of Lagging Indicators	0.00
Growth in Disposable Personal Income	0.00
Quarterly Inflation Rate	0.00
3-month Treasury Bill Rate	0.84
Growth in Unemployment	0.60
Adj. R^2	0.75

Note: The dependent variable in this equation is growth in consumer spending on durable goods, measured quarterly from the first quarter of 1960 through the third quarter of 2012. Cell entries represent block F-tests for the joint significance of excluding five lags from the unrestricted equation. For example, for the Index of Lagging Indicators cell, lags 0 through 4 are included. The *p*-value represents the probability that the variable does *not* help explain growth in consumer spending.

Table 5: The Subjective Causes of Consumer Spending

Variable	(A)	(B)
	Coeff. (s.e.)	Coeff. (s.e.)
Lagged dependent variable	0.464** (0.070)	0.451** (0.070)
Second lagged dependent variable	-0.154* (0.070)	-0.202* (0.071)
Index of Consumer Sentiment	-0.001 (0.008)	
Personal Expectations		-0.011 (0.022)
Personal Retrospections		-0.032* (0.015)
12-Month Business Expectations		0.024* (0.008)
5-Year Business Expectations		-0.027+ (0.014)
Durable Goods Sentiment		0.018* (0.009)
Adj. R^2	0.169	0.198
RMSE	1.451	1.426
Portmanteau Q	7.872	8.237
χ^2	(0.641)	(0.606)
N	204	204

Note: The dependent variable in these analyses is the residuals of the growth in consumer spending after being regressed on the variables in Table 4, measured quarterly from the first quarter of 1960 through the third quarter of 2012. An ** indicates the coefficient is statistically significant at the .01 level, * at the .05 level, and + at the .10 level.

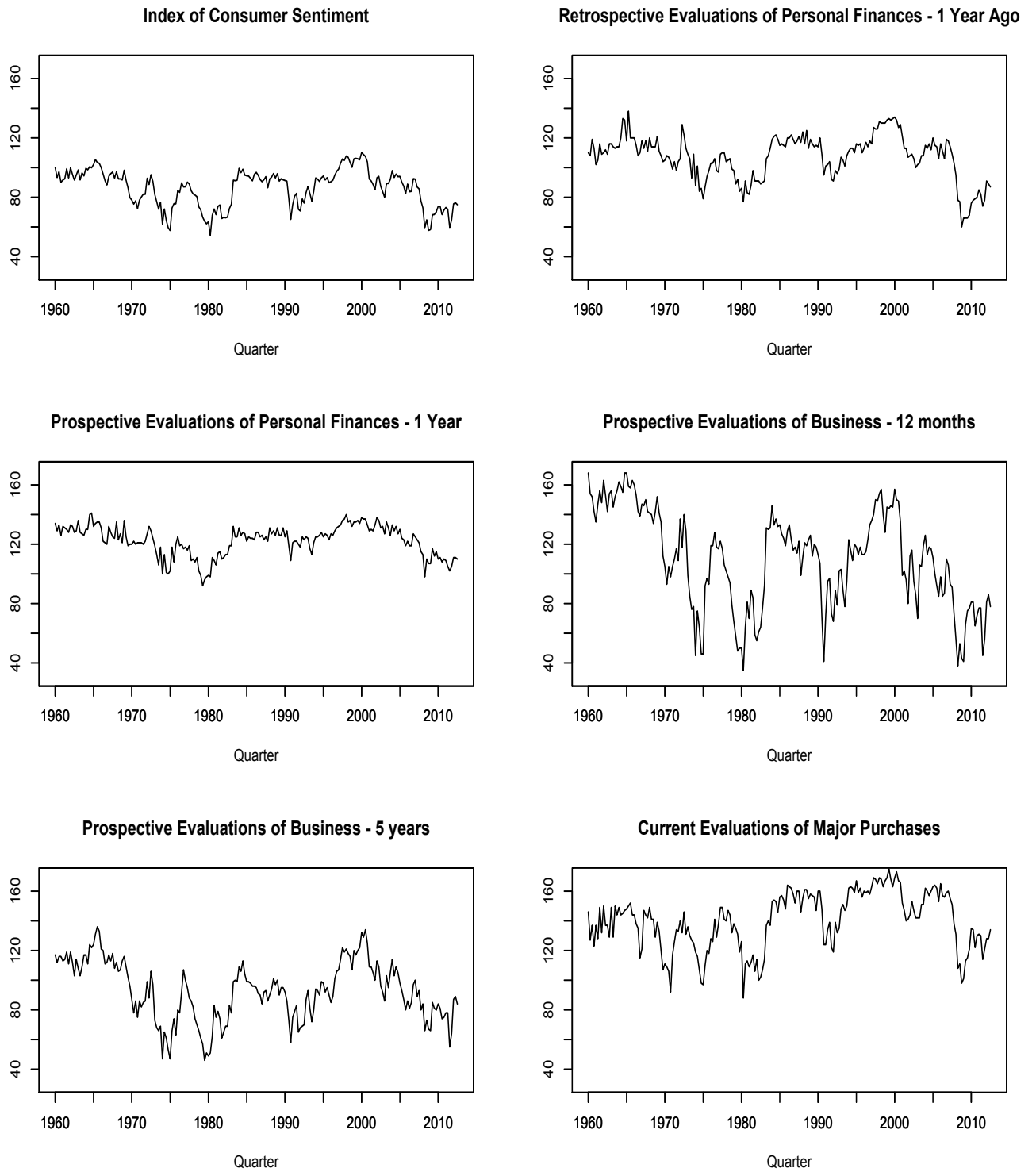


Figure 1: University of Michigan Index of Consumer Sentiment and Its Component Series, 1960:Q1-2012:Q3.

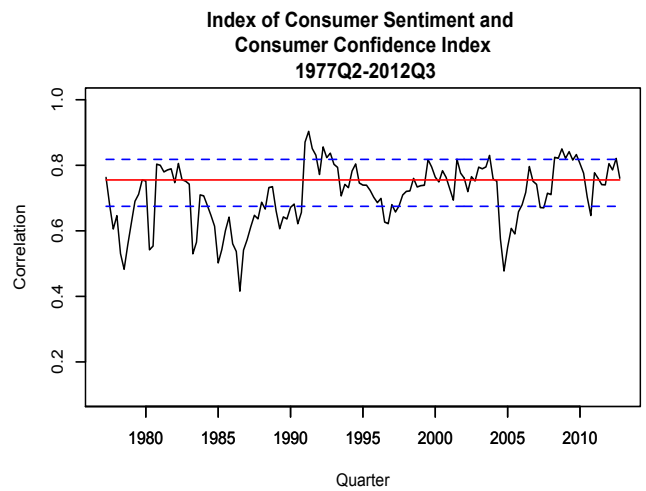
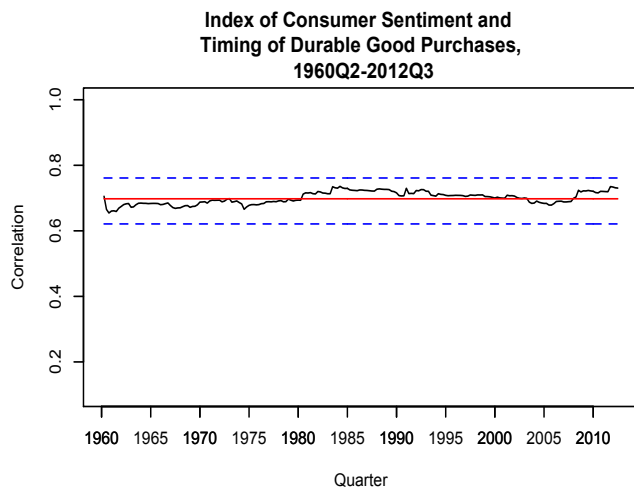
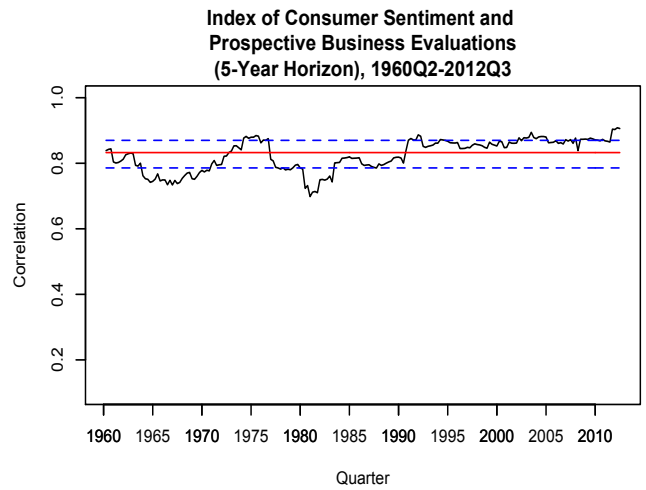
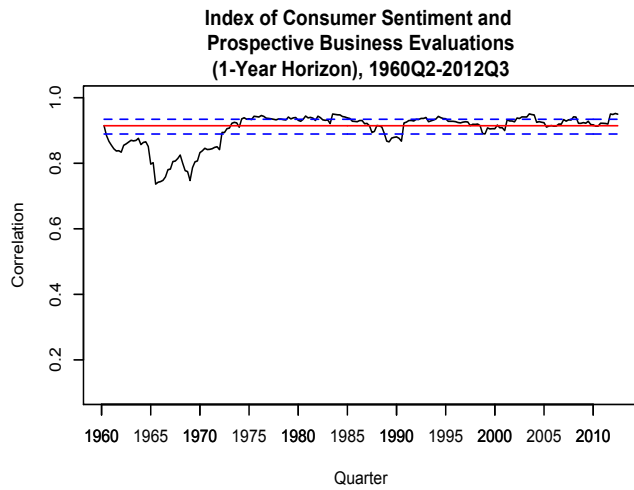
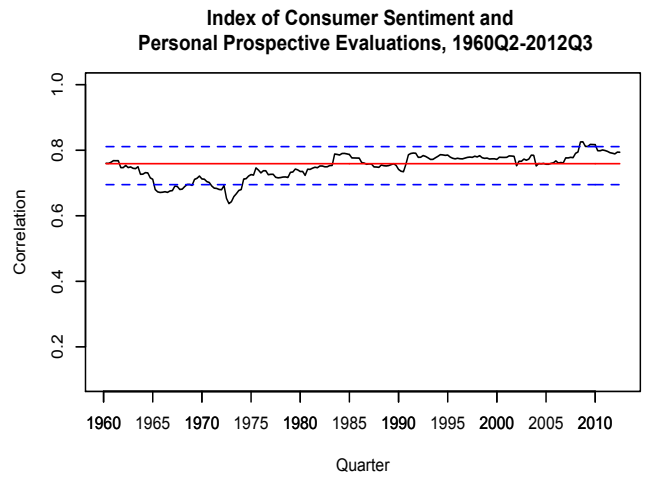
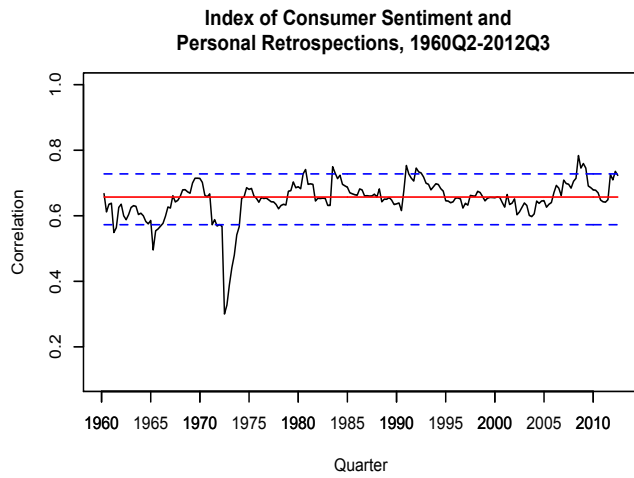


Figure 2: Dynamic Conditional Correlations

Appendix

Table 1: The Subjective Causes of Consumer Spending: AR(2) GARCH(1,1) Models

Variable	(A) Coeff. (s.e.)	(B) Coeff. (s.e.)
Index of Consumer Sentiment	-0.002 (0.011)	
Personal Expectations		-0.018 (0.025)
Personal Retrospections		-0.039** (0.015)
12-Month Business Expectations		0.025* (0.011)
5-Year Business Expectations		-0.023* (0.009)
Durable Goods Sentiment		0.023* (0.009)
AR(1)	0.475** (0.073)	0.461** (0.077)
AR(2)	-0.107 (0.069)	-0.137+ (0.071)
ARCH	0.064 (0.040)	0.075 (0.047)
GARCH	0.867** (0.085)	0.855** (0.085)
Wald χ^2	43.92	61.36
Log likelihood	-362.111	-355.827
Portmanteau Q	9.351	10.393
χ^2	(0.499)	(0.407)
N	206	206

Note: The dependent variable in these analyses is the residuals of the growth in consumer spending after being regressed on the variables in Table 4, measured quarterly from the first quarter of 1960 through the third quarter of 2012. . An ** indicates the coefficient is statistically significant at the .01 level, * at the .05 level, and + at the .10 level.