FORECASTING CONSUMPTION

JIRI SLACALEK

German Institute for Economic Research
October 13, 2004
E-mail: jslacalek@diw.de

Abstract. Several recent papers have demonstrated that the consumer sentiment has predictive power for consumption growth beyond what could be expected from the usual models of consumption dynamics. This paper examines how much forecasting power sentiment indexes have. I simulate real-time out-of-sample forecasting and evaluate the accuracy of several forecasting models using various statistical procedures. They confirm that sentiment indexes can significantly improve the forecasts of various consumption series.

Keywords: consumption, sentiment, forecasting

1. Introduction

Consumer sentiment indexes and their role in forecasting various economic variables have over the last few years received attention in both academic and popular publications.[1] It has been noticed that sentiment indexes are positively correlated with major cyclical indicators, such as GDP, stock prices, and consumption (Acemoglu and Scott, 1994, Carroll *et al.*, 1994). Many papers report that the classical permanent income hypothesis and its generalization due to Campbell and Mankiw

Jiri Slacalek, German Institute for Economic Research, DIW Berlin, Department of Macroeconomics and Forecasting, Königin-Luise-Straße 5, 14195 Berlin, Germany. This research is a part of my Ph. D. dissertation in the Department of Economics, Johns Hopkins University. I would like to thank Laurence Ball, Daniel Leigh, Michael McCracken, Serena Ng, Martin Sommer, Jonathan Wright, participants of the Johns Hopkins macro lunch, referee and especially Christopher Carroll for helpful comments and suggestions. Replication files are available on http://www.econ.jhu.edu/people/slacalek/research/sentiment.zip.

[1] The number of articles in press mentioning consumer sentiment is overwhelming. For example the news search for "consumer sentiment" on news.google.com on April 2, 2003 returned about 650 links to various articles mentioning the phrase in the preceding week only. Academic references are discussed below.

(1989) fail to capture important aspects of aggregate consumption dynamics. This paper, in contrast, suggests how to make use of the relationship between consumption and sentiment to improve *real-time forecasts of consumption growth.* In addition, I find that consumer sentiment indexes provide additional information that is not contained in the variables typically included in consumption regressions.

I construct three alternative forecasting models. The first is a random walk, the second is based on all relevant economic variables except the sentiment index, and the last is based on the same economic variables plus the sentiment index. I produce the forecasts of these models in a simulated real-time framework of rolling regressions and evaluate the forecasts using various statistical procedures. The results indicate that including sentiment in the forecasting regression improves the performance of the model over the time range that I consider. The sentiment-based forecasts have significantly lower mean squared errors than the other forecasts and contain relevant additional information as revealed by the forecast combination regressions. Also the performance of sentiment-based regressions relative to the other models has been particularly strong over the 1994–2002 period. I confirm this finding for various forecasting horizons, sentiment and consumption series.

The plan of the paper is as follows. Section 2 describes the estimation and testing procedures. Section 3 discusses the main results of the statistical evaluation of the various forecasting models. Section 4 reports the robustness of my findings to using alternative forecasting horizons, consumption series and lag selection schemes. Section 5 concludes.

2. Estimation

There is a growing literature on the role of sentiment indexes in explaining aggregate consumption dynamics. Carroll *et al.* (1994) find that the University of Michigan Index of Consumer Sentiment is an important variable for explaining consumption growth. Their regressions produce significant coefficients on the sentiment index and have $\bar{R}^2$ values of about 0.15. Acemoglu and Scott (1994) report similar findings for consumer sentiment in the United Kingdom.[2] Howrey (2001) finds that the Michigan index improves predictions of the probability of a recession. He also finds that sentiment is significant in consumption regressions but claims that it does not substantially reduce the standard error of the regression.

All these papers focus on a different issue from mine: their major goal is typically to test the standard theories of consumption (such as the Campbell–Mankiw model). In contrast, the goal of this paper is to investigate the real-time out-of-sample forecasting performance of sentiment-based models and to compare them to those that do not contain sentiment.

Bram and Ludvigson's (1998) article is probably the most closely related to this one. These authors provide some results on the out-of-sample forecasting performance of sentiment indexes. However, because they only analyze quarterly data, they do not obtain enough observations to investigate how the mean squared errors have evolved over time and they are often unable to find a significant difference in the mean squared errors of various forecasts.

2.1. **Forecasting Regressions.**

The forecasting regression equations have the following form:

$$\Delta_h c_t = \beta_0 + \sum_{i=0}^{p} \beta_{i+1} S_{t-i+1} + \sum_{i=0}^{p} \boldsymbol{\phi}_{i+1}^{\top} \mathbf{Z}_{t-i+1} + \varepsilon_t, \qquad (1)$$

where c denotes the log of consumption series, S denotes the sentiment index series and $\mathbf{Z}$ represents other variables.[3] Finally, the symbol "Δ_h" denotes the difference between a variable at time $t+h$ and t, i.e. $\Delta_h x_t = x_{t+h} - x_t$. The size of β_{i+1} indicates the predictive power of

[2]In this paper I will consider The University of Michigan and the Conference Board sentiment indexes. They are based on representative samples of 500 and 5,000 respondents respectively each month. The respondents are asked three present conditions questions and two expectations questions as a part of a larger survey. The questions are concerned with the present financial situation and income, with business conditions, and with the availability of jobs at present and in the future. For the precise statement of the questions see e.g. Bram and Ludvigson (1998).

[3]These variables include lagged consumption growth (defined as $\Delta c_{t-i}, i = 1, \ldots, p$), the long–short interest rate spread, inflation, unemployment and the Fed funds rate.

sentiment. I use *monthly* data and set h equal to three, i.e. I consider one-quarter-ahead consumption growth forecasts.

Since the consumption data are available with a two-month lag, I include the *one-period ahead future value* $(t+1)$ of sentiment S_{t+1} and the available variables in $\mathbf{Z}$ (Fed funds rate and interest rate spread).[4] In contrast, the variables in $\mathbf{Z}$ that are not available quickly (such as inflation and unemployment), are entered only with lags dated t and lower. For example, at the end of February, the January consumption data are not yet available while the February sentiment numbers are. Finally, I use two lags of explanatory variables, i.e. $p = 2$. Choosing a different lag structure using information criteria or predictive least squares yields comparable results.

2.2. **Assessing the Quality of Forecasts.** The quality of competing forecasts $\hat{y}$ of a variable y is typically measured by their mean squared errors. Forecast A outperforms forecast B if it has a lower MSE. The Diebold and Mariano (1995) test indicates whether the MSE of A is significantly different from the MSE of B. Forecast $\hat{y}$ of variable y has the MSE defined as

$$\mathrm{MSE}(\hat{y}) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t - y_t\right)^2.$$

Denote the difference in squared errors of two forecasts, $\hat{y}_t^1$ and $\hat{y}_t^2$ as $d_t = \left(\hat{y}_t^1 - y\right)^2 - \left(\hat{y}_t^2 - y\right)^2$ and the spectral density of d_t at frequency zero as $f_d(0)$. Denote

$$\tilde{d} = \frac{1}{T}\sum_{t=1}^{T}\left(\left(\hat{y}_t^1 - y_t\right)^2 - \left(\hat{y}_t^2 - y_t\right)^2\right)\Big/\sqrt{2\pi \times \hat{f}_d(0)}. \tag{2}$$

Under suitable assumptions and under the null of the forecasts having the same MSEs, the normalized statistic $\sqrt{T} \times \tilde{d}$ converges in distribution to the standard normal distribution $\mathcal{N}(0,1)$. Below I estimate the asymptotic covariance matrix $2\pi f_d(0)$ by computing a weighted sum of sample autocovariances of d, and using the Newey–West estimator. I

[4]For example that final January consumption data are available at the end of March. Estimates of January consumption are released at the beginning of March.

assume that the there is no autocorrelation in the periods beyond $h-1$ and set the window equal to $h-1$ lags.

Unfortunately, the standard normal asymptotic distribution of the statistic $\bar{d} = \sqrt{T}\tilde{d}$ is not valid for the nested forecasting models, that I consider.[5] Clark and McCracken (2002) show that for nested models the $\bar{d}$ statistic has a non-standard distribution. Consequently, the correct critical values for $\bar{d}$ statistic have to be simulated by drawing random numbers from the appropriate Clark–McCracken distribution.

Another forecast evaluation procedure designed to test the quality of various forecasts is the forecast combination regression (see Chan *et al.*, 1999, and Stock and Watson, 1999). Suppose I have two different forecasts $\hat{y}^1$ and $\hat{y}^2$ of variable y. I first estimate the regression

$$y_t = \lambda \hat{y}^1_t + (1-\lambda)\hat{y}^2_t + \varepsilon_t, \tag{3}$$

where the sum of coefficients on the right-hand side is restricted to be one. If the coefficient λ is one, forecast 1 encompasses forecast 2, because including forecast 2 in the regression of y on $\hat{y}^1$ does not improve the fit of the regression. Thus, if λ is large, forecast 2 is inferior to forecast 1. In the case of non-nested models the the hypothesis $\lambda = 1$ can be tested by a standard t test. However, when the models are nested, the t statistic has a nonstandard distribution derived by Clark and McCracken (2002).

3. The Results

This section describes the data and reports the results of my econometric experiments.

3.1. Preliminary Results.

I will monthly data obtained from the DRI database that cover the time period between February 1978 and December 2002.[6] I use four consumption series: personal consumption expenditures (total consumption, GMCQ in the DRI mnemonics), durable consumption (GMCDQ), nondurable consumption (GMCNQ),

[5] Two models are nested if the RHS of one model is a subset of the RHS of the other model, i.e. the regressors of the two models can be written as x_1 and $x_2 = (x_1, x_{22})$.

[6] I cannot extend my sample to before 1978 because the monthly sentiment data are not available before that date.

and services consumption (GMCSQ). I examine the four sentiment series available from the DRI: the University of Michigan index of consumer sentiment (HHSNTR), the University of Michigan index of consumer expectations (HHSNTN), the Conference Board consumer sentiment index (HHCNF), and the Conference Board index of consumer expectations (HHCNXP). After running some preliminary regressions I decided to include the following potentially relevant variables in my models: the long–short interest rate spread (the difference between the yields on the three-month Treasury bill and ten-year Treasury bond, FYGT10−FYGM3), inflation (PUNEW), unemployment (LHUR) and the Fed funds rate (FYFF).[7]

Based on these regressions I decided to test the performance of the following three models:

(1) A model that includes the following plausibly relevant variables: sentiment, Fed funds rate, unemployment, inflation, long–short spread and lagged consumption (also called the big model below).
(2) A model that is the same as model 1 above except that it does not include sentiment (non-sentiment).
(3) A model with a constant as the only independent variable (random walk model, after Hall (1978).

Some statistics illustrating the performance of these models are reported in Table 1 which is an approximate replication of the findings of previous authors. This table suggests that sentiment indexes generally remain significant variables for consumption growth even when other variables are present on the right-hand side.[8] This finding is confirmed both by the $\bar{R}^2$s and the t statistics. The differences in $\bar{R}^2$s between

[7] I do not report the results here. I ran regressions of one quarter, one year and one month ahead consumption growth on other variables and picked the variables which in at least one regression had a significant t statistic on at least one lag (on 5% confidence level). Other variables that I considered but were not significant are personal disposable income (GMYD), return on the S&P 500 stock index (FSPCOM), the composite index of eleven leading indicators (DLEAD) and the composite index of four coincident indicators (DCOINC).

[8] In Table 1 as in Tables 2–4 below I only report results for the one-quarter-ahead forecast horizon. As illustrated by Table 5 below, similar results obtain for the one-year or one-month horizon, unless otherwise noted.

the big and non-sentiment models in columns two and three typically range between 10–12.5%. Also the t statistics on the first sentiment lag are highly significant and the standard errors of the regressions are approximately 6.5% smaller with the big model than with the model without sentiment.[9]

3.2. **Main Results.**

Simulated Out-of-Sample Forecasting. Simulated out-of-sample forecasting consists of estimating the model (1) on a sample preceding the forecast period. For example, one can first estimate the model for the data between 1978:01 and 1995:01, make a forecast of consumption growth from 1995:02 on, then augment the data with one more observation, 1995:02, and produce another forecast for 1995:03. By repeating this procedure one generates a series of forecasts by a technique that would be available to a forecaster in real time. One can then judge the quality of the forecasts using the statistical procedures outlined above.

One could argue that the framework used here is not really feasible in real time because the data used are revised and are thus not the original data available at the time of the forecast. This objection is not completely justified since all data series that I use, except for consumption, are not being revised (i.e. inflation, unemployment, interest rates). The only series that get revised are the consumption series. Even then it is appropriate to let the dependent variable be the most recently revised value of consumption if the forecaster's ultimate goal is to predict the true value of consumption, not its first unrevised estimate. Strictly speaking, the only series that should be unrevised are the lagged consumption series on the right-hand side. I address this issue below.

I employed the following rolling regression approach. My data cover the 1978:01–2002:12 period. I estimated the regressions for this period

[9] I do not report the F statistics on all sentiment lags. These reject the null hypothesis even more decisively than the t tests. This is despite the fact that the further sentiment lags are individually less significant than the most recent lags. The reason is that the estimates on various sentiment lags are negatively correlated, possibly due to high autocorrelation in sentiment.

starting in 1983:04, i.e. I started with a window of five years. I conducted the analysis with a fixed starting date and an increasing window (not a fixed window). The results for forecasts generated using a fixed window are qualitatively the same.[10]

The mean squared errors of forecasts of total consumption growth are reported in Table 2. All columns display MSEs relative to the MSE of the big model. The MSEs of the big model are by 15% or more lower than the MSEs of the non-sentiment model, which in turn outperforms the random walk. This result implies that including a sentiment index in the forecasting regression increases its precision. Table 2 also displays the values of the Diebold–Mariano statistic (2) and critical values of the Clark–McCracken distribution.

The Diebold–Mariano statistics in Table 2 test the null hypothesis that the MSE of a given model is the same as the MSE of the big model. Looking at the critical values, the big model outperforms both the non-sentiment and random walk models—at any reasonable significance level, I was able to reject the null of equal MSEs. Furthermore, judging by the relative sizes of MSEs, the non-sentiment and random walk models do not perform well. However, the variance of the squared forecast errors for the random walk model tends to be high, which causes the Diebold–Mariano statistic to be low compared to that of the non-sentiment model.

Evolution of Forecasts over Time. These results suggest that sentiment indexes can improve the forecasts of consumption growth. I will now look into how the performance of the various forecasting models has evolved over time. One way to do this is to look at the plot of (squared) forecast errors. I first smooth the squared errors by filtering out a lot of noise and look only at the general trend.

This trend is displayed in Figure 1, that shows the results for the four available sentiment indexes I use the two-sided simple moving

[10]In general, the MSEs of forecasts generated using a fixed window are likely to be higher than those based on an increasing window. The relative size of the MSEs from these two methods depends on the size of the window. For example, using a five-year window, the MSEs are higher for the fixed window for all models. Using a seven-year window, the MSEs are higher for the fixed window for all the models except for the non-sentiment model.

average filter with a window of three years (36 months), defined as $\bar{e}_t = \sum_{s=-m}^{m} e_{t+s}^2/(2m+1)$ with $m = 36/2 = 18$. The figures reveal that the big model forecasts very well over the most of the time sample considered (with the exception of the mid 1990s). It has done especially well in the early and late 1990s. Overall, Figure 1 suggests that the good performance of forecasts based partly or completely on sentiment indexes is not caused by a few outliers with very low MSEs but results from better performance throughout most of the sample.

One interesting finding shown in Figure 1 is the significant increase in the precision of all consumption forecasts after 1994. This is to a large extent caused by the decline in the variance of consumption growth. The variance of one-quarter ahead consumption growth in 1978–1994 was 12.41 but was only 3.85 after 1994. This change is also reflected in the decrease in the mean squared error of the random walk forecasts.

Forecast Combination. I now turn to the forecast combination regressions (3) described above. The results from this experiment are reported in Table 3. The numbers in the table are the estimates of λ's from equation (3), letting $\hat{y}^2$ be the forecast of the big model and $\hat{y}^1$ that of the non-sentiment or the random walk models. The numbers below them are the t statistics testing the null $\lambda = 1$ based on the HAC robust standard errors. Further below are the p values of the Clark–McCracken asymptotic distribution of $\hat{\lambda}_{OLS}$. The larger λ is, the more useful is a given forecast compared to the forecast of the big model, i.e. the more weight it has in the optimal combined forecast. The λ coefficients are very far from one, in fact, they are often close to zero. This means that the forecast combination confirms the previous results about the quality of the big model. I strongly reject the hypothesis that the big model contains no extra information. In addition, it is often hard to reject the hypothesis that $\lambda = 0$, i.e. that the forecasts of the alternative models are encompassed in the big one.

Comparison of Various Sentiment Indexes. Next, I discuss the relative performance of various sentiment indexes in terms of their MSEs. A relatively robust conclusion from the previous and following experiments reported in Tables 1–3 is that

$$\text{MSE(CBO)} < \text{MSE(ME)} < \text{MSE(MO)} < \text{MSE(CBE)}, \tag{4}$$

where CBO, ME, MO and CBE denote the overall Conference Board, Michigan expectations, overall Michigan and Conference Board expectations, respectively. The overall Conference Board index does a considerably better job than other indexes. The difference between other measures of sentiment is not so high. However, the Michigan expectations index performs better than the overall Michigan. Surprisingly, the Conference Board expectations index does a relatively bad job. This is somewhat surprising as one would have expected the expectations components to produce more precise forecasts. Overall, with the exception of the overall Conference Board index, the differences in the performance of various sentiment indexes are quite small.

3.3. **Real-Time Data.** A possible critique of the above results is that the reported results are not really real-time since the lagged consumption series used as an explanatory variable in the forecasting regression uses final data, not data actually available when the forecast is made. This section shows that the results remain to hold even when the real-time consumption data are used.

Due to the lack of monthly real-time consumption data I examine the predictive power of consumer sentiment at quarterly frequency for years 1970–2002.[11] The results—replications of experiments of Table 2—are reported in Table 4.

The results have the same implications as those in Table 2. They indicate that sentiment continues to be a good predictor of consumption growth even if the real-time data are used. The relative MSEs continue to be about the same as in Table 4. Interestingly, the ranking of MSEs (4) changes to

$$\text{MSE(CBE)} < \text{MSE(CBO)} < \text{MSE(MO)} < \text{MSE(ME)}.$$

[11]The real-time consumption data were downloaded from Dean Croushore's real time data set at the Philadelphia Fed, `http://www.phil.frb.org/files/forecast/data/qvqd.zip`, variable RCON.

This means that the Conference Board Expectations index predicts better in the real-time than with final data in comparison with other sentiment indexes. The Diebold–Mariano statistics fall but their remain overwhelmingly significant as indicated by p values. Consequently, the results reported above hold not just for final data, but also for data available in real time.

4. Extensions

In this section of the paper I investigate the robustness of my findings to some alternative assumptions. First, I explore whether consumer sentiment indexes are useful for predicting total personal consumption expenditure at the one-year and one-month ahead horizons. Then I investigate the following other consumption series: durable, nondurable and services consumption.

Alternative Forecasting Horizons. This paper has focused on one quarter ahead forecasts because this seems like the most relevant horizon in the real-time context. In this section I explore the implications of my forecasting models for one year and one month horizons.

In general, the results for alternative horizons, displayed in Table 5, are similar to those obtained in Table 2. The big model outperforms the non-sentiment model, which in turn outperforms the random walk. This finding is particularly significant for longer forecasting horizons, such as one-year-ahead. The difference is not as significant for the one-month ahead forecasts. There are two explanations for this finding. First, monthly consumption growth is much more volatile than quarterly or annual growth, and is thus inherently harder to predict. Second, several authors (e.g. Wilcox, 1992, Sommer, 2002) and even the Bureau of Economic Analysis (see Bureau of Economic Analysis, 2002) which collects the consumption data, argued that there is a large component in consumption data that is estimated, imputed or interpolated.[12] It is likely that this component is relatively large for monthly data and will be averaged out for lower frequency data. In that case the big model, which is admittedly closer to the true model, performs

[12] Sommer (2002) argues that this fraction is more than 30%.

better for longer horizons where the data are less noisy and more predictable, as documented in Table 5.

Alternative Consumption Series. I now turn to other consumption series, such as durable, nondurable and services consumption. Table 6 reports the findings (this table basically replicates the results on MSEs from Table 2 for other consumption series).

The general results are the same as for the forecasts of total consumption. The big model again produces better forecasts than the non-sentiment and random walk models in most cases. The exception is services consumption that is relatively hard to predict, judging by the fact that the random walk typically performs best for this consumption series. Nevertheless, the results reported in this table are in line with the previously reported results, i.e. including sentiment in the forecasting regression decreases the mean squared error by about 5–10%.

5. Conclusion

The goal of this paper is to quantify the predictive power of sentiment indexes for consumption growth. I use the well documented in-sample relationship between sentiment and consumption and generate out-of-sample forecasts of consumption growth that are more precise than those obtained from the considered models without sentiment. The results imply that consumer sentiment indexes are strong predictors of consumption growth. They contain important additional information not contained in the economic variables typically included in the consumption regressions. I find that including a sentiment index in the forecasting regression results in a decline in the mean squared errors of about 5–15%. My results are robust to using various statistical procedures, alternative forecasting horizons, and alternative consumption and sentiment series.

References

1. Acemoglu, Daron and Andrew Scott (1994), "Consumer Confidence and Rational Expectations: Are Agents' Beliefs Consistent with the Theory?," *The Economic Journal* **104**: 1–19.
2. Bureau of Economic Analysis (2002), "Updated Summary NIPA Methodologies," *Survey of Current Business,* October.
3. Bram, Jason and Sidney Ludvigson (1998), "Does Consumer Confidence Forecast Household Expenditure? A Sentiment Index Horse Race," *FRBNY Economic Policy Review* **4(2)**: 59–77.
4. Campbell, John Y., and N. Gregory Mankiw (1989), "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989,* (eds); O. Blanchard and S. Fischer., MIT Press, Cambridge, MA.
5. Carroll, Christopher, Jeffrey Fuhrer, and David Wilcox (1994), "Does Consumer Sentiment Forecast Household Spending? If So, Why?" *American Economic Review,* **84(5)**: 1397–1408.
6. Chan, Y. L., James H. Stock, and Mark W. Watson (1999), "A Dynamic Factor Model for Forecast Combination," *Spanish Economic Review* **1(2)**: 91–122.
7. Clark, Todd E., and Michael W. McCracken (2002), "Evaluating Long-Horizon Forecasts," Federal Reserve Bank of Kansas City working paper.
8. Diebold Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics,* **13(3)**: 253–263.
9. Hall (1978), Robert E., "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis," *Journal of Political Economy* **86(6)**: 971–987.
10. Howrey, E. Phillip (2001), "The Predictive Power of the Index of Consumer Sentiment," *Brookings Papers on Economic Activity* **2001(1)**: 175–216 (with discussion).
11. Sommer, Martin (2002), "Habits, Sentiment and Predictable Income in the Dynamics of Aggregate Consumption," working paper, Johns Hopkins University.

12. Stock James H., and Mark W. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics* **44**: 293–335.
13. Wilcox, David W. (1992), "The Construction of the U.S. Consumption Data: Some Facts and Their Implications for Empirical Research," *American Economic Review* **82(4)**: 922–941.

TABLE 1. $\bar{R}^2$'s and t statistics on Sentiment

Forecasting Model	Big	Non-sentiment
Michigan Overall: $\bar{R}^2$	0.30	0.21
t statistic	3.81	–
Std. error of reg.	2.46	2.63
Michigan Expectations: $\bar{R}^2$	0.30	0.21
t statistic	4.19	–
Std. error of reg.	2.46	2.63
Conf. Board Overall: $\bar{R}^2$	0.36	0.21
t statistic	3.70	–
Std. error of reg.	2.36	2.63
Conf. Board Expectations: $\bar{R}^2$	0.30	0.21
t statistic	3.21	–
Std. error of reg.	2.47	2.63

Notes: Time frame: February 1978–December 2002; Total Consumption, One Quarter Ahead Forecasts. The "t statistic" numbers are t statistics on the most recent sentiment used. They are derived using the Newey–West HAC robust standard errors with 2 lags (i.e. one lag less than the forecasting horizon). The "std. error of regression" numbers are the standard deviations of the residuals, calculated as $\sqrt{1/T \sum e_t^2}$. The regressions are calculated using 2 lags of variables on the right-hand side. All forecasts are for total personal consumption expenditures.

TABLE 2. Relative Mean Squared Errors and Diebold–Mariano Statistics

Forecasting Model	Big	Non-sentiment	Random Walk
Michigan Overall	1.00	1.15	1.24
DM statistic	0.00	2.60	1.93
p Value	–	0.0000	0.0000
Michigan Expectations	1.00	1.16	1.24
DM statistic	0.00	2.42	1.84
p Value	–	0.0000	0.0000
Conf. Board Overall	1.00	1.29	1.38
DM statistic	0.00	2.23	2.70
p Value	–	0.0000	0.0000
Conf. Board Expectations	1.00	1.14	1.22
DM statistic	0.00	1.76	1.71
p Value	–	0.0004	0.0000

Notes: One quarter ahead total personal consumption expenditure forecasts. All mean squared errors are calculated with respect to the big forecast. The “p value” row denotes p values from the Diebold–Mariano test of forecast accuracy, the null being that the MSE of a given forecast is the same as the MSE of the big model forecast in the first column. The p values were generated by simulating 10,000 independent draws from the Clark–McCracken asymptotic distribution. The regressions are calculated using 2 lags of variables on the right-hand side. The Diebold–Mariano statistics are based on HAC standard errors with Newey–West window and 2 lags.

TABLE 3. Combination Regressions

Forecasting Model	Non-sentiment	Random Walk
Michigan Overall λ	-0.12	0.27
t Statistic	4.65	6.29
Clark–McCracken p	0.0000	0.0000
Michigan Expectations λ	-0.04	0.29
t Statistic	4.83	6.44
Clark–McCracken p	0.0000	0.0000
Conf. Board Overall λ	0.08	0.14
t Statistic	5.02	7.10
Clark–McCracken p	0.0000	0.0000
Conf. Board Expectations λ	0.11	0.29
t Statistic	4.31	6.04
Clark–McCracken p	0.0000	0.0000

Notes: The estimated regressions are: $y_t = \lambda \hat{y}_t^i + (1-\lambda)\hat{y}_t^{Big} + \varepsilon_t, \ i \in \{\text{Non-sentiment,RW}\}, \mathsf{H}_0 : \lambda = 1$. The t statistics are calculated using HAC standard errors based on the Newey–West window with 2 lags. The "Clark–McCracken p" numbers are the p values of the distribution of the t statistic described in Clark and McCracken (2002). They were generated by simulating 10,000 independent draws from the asymptotic distribution. The t statistic tests the null hypothesis: $\lambda = 1$. All forecasts are for total personal consumption expenditures.

Table 4. Real Time Data, Relative Mean Squared Errors and Diebold–Mariano Statistics

Forecasting Model	Big	Non-sentiment	Random Walk
Michigan Overall	1.00	1.18	1.18
DM statistic	0.00	1.35	0.71
p Value	–	0.0022	0.0049
Michigan Expectations	1.00	1.15	1.15
DM statistic	0.00	1.26	0.66
p Value	–	0.0028	0.0055
Conf. Board Overall	1.00	1.38	1.38
DM statistic	0.00	1.93	1.12
p Value	–	0.0000	0.0013
Conf. Board Expectations	1.00	1.49	1.48
DM statistic	0.00	2.08	1.36
p Value	–	0.0000	0.0005

Notes: One Quarter Ahead Total Consumption Forecasts. The table reports the MSEs of forecast of final consumption growth using real-time data. All mean squared errors are calculated with respect to the big forecast. The "p value" row denotes p values from the Diebold–Mariano test of forecast accuracy, the null being that the MSE of a given forecast is the same as the MSE of the big model forecast in the first column. The p values were generated by simulating 10,000 independent draws from the Clark–McCracken asymptotic distribution. The regressions are calculated using 2 lags of variables on the right-hand side. The Diebold–Mariano statistics are based on HAC standard errors with Newey–West window and 2 lags. All forecasts are for total personal consumption expenditures.

Table 5. Relative Mean Squared Errors for Alternative Forecasting Horizons

Forecasting Model	Big	Non-sentiment	Random Walk
One Quarter Ahead			
Overall Michigan	1.00	1.15	1.24
Michigan Expectations	1.00	1.16	1.24
Conf. Board Overall	1.00	1.29	1.38
Conf. Board Expectations	1.00	1.14	1.22
One Year Ahead			
Overall Michigan	1.00	1.31	1.60
Michigan Expectations	1.00	1.36	1.66
Conf. Board Overall	1.00	1.46	1.79
Conf. Board Expectations	1.00	1.18	1.44
One Month Ahead			
Overall Michigan	1.00	1.04	1.16
Michigan Expectations	1.00	1.03	1.15
Conf. Board Overall	1.00	1.11	1.23
Conf. Board Expectations	1.00	1.06	1.18

Notes: All MSEs are measured relative to the MSE of the big model. All forecasts are for total personal consumption expenditures.

Table 6. Relative Mean Squared Errors for Alternative Consumption Series

Forecasting Model	Big	Non-sentiment	Random Walk
Durable Consumption			
Overall Michigan	1.00	1.06	1.28
Michigan Expectations	1.00	1.07	1.30
Conf. Board Overall	1.00	1.12	1.35
Conf. Board Expectations	1.00	1.07	1.29
Nondurable Consumption			
Overall Michigan	1.00	1.04	1.14
Michigan Expectations	1.00	1.04	1.14
Conf. Board Overall	1.00	1.05	1.16
Conf. Board Expectations	1.00	1.02	1.12
Services Consumption			
Overall Michigan	1.00	1.06	0.93
Michigan Expectations	1.00	1.06	0.93
Conf. Board Overall	1.00	1.17	1.02
Conf. Board Expectations	1.00	1.06	0.93

Notes: All MSEs are measured relative to the MSE of the big model; all forecasts are one quarter ahead.

FIGURE 1. Smoothed Mean Squared Errors

(a) Overall Michigan

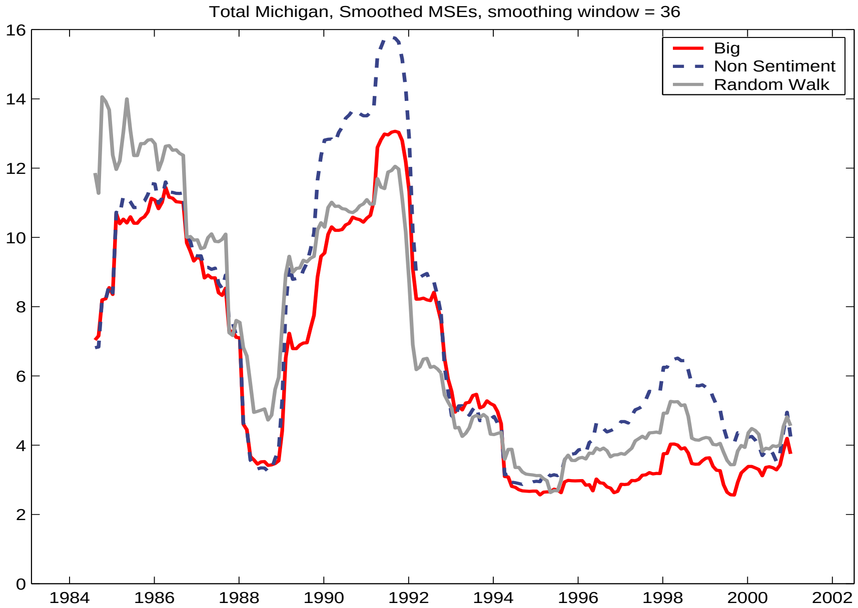

Figure 1. Smoothed Mean Squared Errors

(b) Michigan Expectations

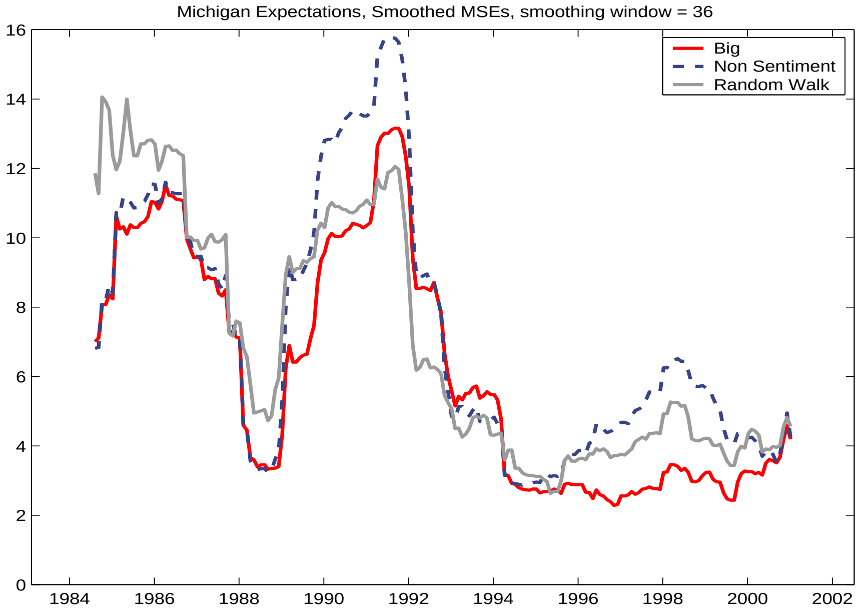

FIGURE 1. Smoothed Mean Squared Errors

(c) Overall Conference Board

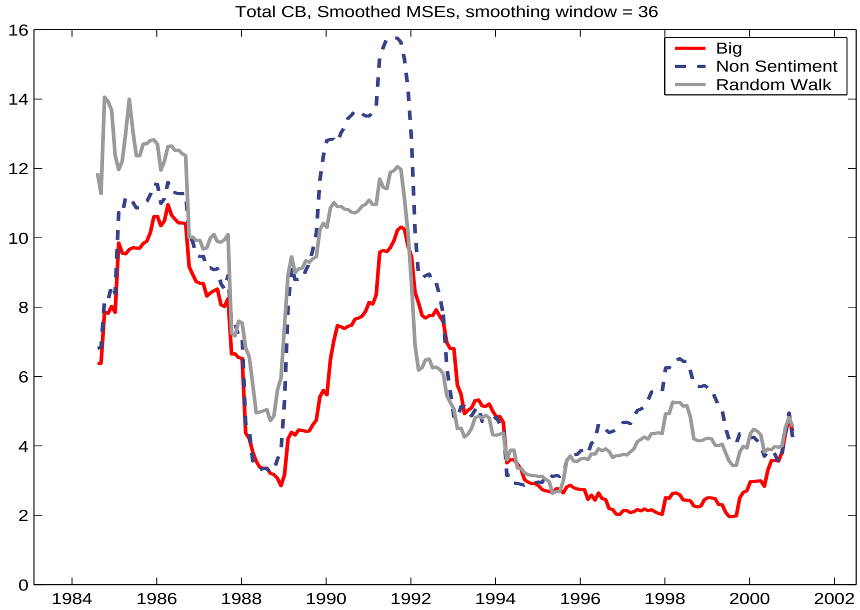

FIGURE 1. Smoothed Mean Squared Errors

(d) Conference Board Expectations