Sticky Information and Model Uncertainty in Survey Data on Inflation Expectations*

William A. Branch†
University of California, Irvine

July 7, 2005

Abstract

This paper compares models of heterogeneity in survey inflation expectations. On the one hand, we specify two models of forecasting inflation based on limited information flows of the type developed in Mankiw and Reis (2002). We present maximum likelihood results that suggests a sticky information model with a time-varying distribution structure is consistent with the Michigan survey of inflation expectations. We also compare these 'sticky information' models to the endogenous model uncertainty approach in Branch (2004). Non-parametric evidence suggests that model uncertainty is a more robust element of the data.

JEL Classifications: C53; C82; E31; D83; D84

Key Words: Heterogeneous expectations, adaptive learning, model uncertainty, survey expectations.

1 Introduction

Despite the prominence of rational expectations in macroeconomics there is considerable interest in its limitations. As an alternative some researchers propose modeling agents as econometricians (Evans and Honkapohja (2001)). This adaptive learning approach typically assumes agents have a correctly specified model with unknown

*This paper has benefited enormously from comments and suggestions by Bruce McGough and Ken Small.
†Email: wbranch@uci.edu; Department of Economics, 3151 Social Science Plaza, Irvine, CA 92697; Phone: (949) 824-4221; Fax: (949) 824-2182.

parameters. In many models agents' optimal decision rules are functions of these beliefs.

Other approaches impose bounded rationality at the primitive level; see, for example, Mankiw and Reis (2002), Ball, Mankiw, and Reis (2003), Branch, Carlson, Evans, and McGough (2004) and Sims (2003). Of these the sticky-information model of Mankiw and Reis (2002) yields important (and tractable) implications for macroeconomic policy. Mankiw and Reis (2002) replace the staggered pricing model of Calvo (1983), which is employed extensively in Woodford (2003), with a model of staggered information flows. Each period, each firm, with a constant probability, updates its information set when optimally setting prices. The remaining firms are free to set prices also, but do not update their information from the previous period. Importantly, this leads to a phillips curve with inflation as a function of past expectations of current inflation rather than current expectations of future inflation as in Woodford (2003). Mankiw and Reis (2002) and Ball, Mankiw, and Reis (2003) show that this implies greater persistence in response to monetary shocks.

In an innovative paper, Mankiw, Reis, and Wolfers (2003) seek evidence of sticky information in survey data on inflation expectations. They examine surveys of professional forecasters and construct a data set based on the Michigan Survey of Consumers. Their results show that these survey data are inconsistent with either rational or adaptive expectations and may be consistent with a sticky-information model.

There is considerable interest in empirically inferring the methods with which agents form expectations. In particular, there is compelling evidence that survey expectations are heterogeneous and not rational. For example, Bryan and Venkatu (2001 a,b) document striking differences in survey expectations across demographic groups. Carroll (2003) provides evidence that the median response in the Survey of Consumers is a distributed lag of the median response from the Survey of Professional Forecasters. Branch (2004), adapting Brock and Hommes (1997), develops a methodology for assessing the forecasting models agents use in forming expectations. In that paper, evidence suggests survey responses are distributed heterogeneously across univariate and multivariate forecasting models. Brock and Durlauf (2004) argue that if agents are uncertain about the prevailing inflation regime then this uncertainty may manifest itself in agents switching between myopic and forward-looking predictors; hence, model uncertainty is a key aspect of expectation formation.[1]

This paper has three objectives: first to characterize sticky information in survey data in the sense that a proportion of agents do not update information each period; second, to test whether these proportions are static or dynamic; third, to provide evidence whether model uncertainty or sticky information is a more robust element of the survey data. Carroll (2003) and Mankiw, Reis, and Wolfers (2003) provide indirect evidence of limited information flows in expectation formation. This paper

[1]Other papers which show heterogeneity across forecasting models include Baak (1999), Chavas (2000), and Aadland (2003). Experimental evidence is provided by Heemeijer, Hommes, Sonnemans, and Tuinstra (2004) and Hommes, Sonnemans, Tuinstra, and Velden (2005).

elaborates on the nature of these flows in survey data. We also bridge the sticky information and heterogeneous expectations literature by presenting evidence of both model heterogeneity and limited information flows.

This paper extends Branch (2004) by focusing on predictors which differ temporally rather than spatially. Our approach, like Mankiw, Reis, and Wolfers (2003), tests for sticky information flows in agents' survey expectations. We also extend Mankiw, Reis, and Wolfers (2003) by proposing two formulations of sticky-information. The first is the Mankiw-Reis approach which we refer to as the *static sticky information* model. The other approach assumes that expectations are formed by a discrete choice between forecasting functions which differ by the frequency with which they are recursively updated. Using data from the Survey of Consumers at the University of Michigan, we provide evidence of sticky information by testing the sticky information models against the full-information alternative. Maximum likelihood evidence shows that: sticky information in survey data is dynamic in the sense that the distribution of agents across predictors is time-varying; the distribution of agents is not geometric so that, on average, the highest proportion of agents update information somewhat infrequently. This last result is in contrast to an implication of the Mankiw-Reis model which has the highest proportion of agents updating each period.

Our final objective is to determine whether model uncertainty is a more robust element of the survey data than sticky information. We address this issue by comparing the Rationally Heterogeneous Expectations (RHE) model of Branch (2004) with the sticky information models presented in this paper. In Branch (2004) agents are uncertain about the correct model for the economy and so each period they make a discrete choice between alternatives.[2] We (non-parametrically) estimate the density functions implied by these models and compare the fit to the histogram of the actual survey data. We find that neither the sticky information or the model uncertainty approaches are statistically identical to the distribution of the survey data. However, on average, the model uncertainty approach provides a better fit than the sticky information models. As a corollary to these non-parametric results, we show that a sticky information model which lets the distribution of information across agents vary over time provides a better fit than the static version of Mankiw and Reis (2002).

In this last empirical exercise, we use non-parametric techniques to assess whether model uncertainty or sticky information provide a better fit to the entire distribution of survey expectations. Attempting to fit the entire distribution is a challenging hurdle indeed. Because of the demographic characteristics identified by Bryan and Venkatu (2001a,b), and other unobserved idiosyncrasies, it is not reasonable to expect that the simple empirical models, motivated by theory, will exactly match the entire period by period distribution of survey data. Instead, we examine how closely they can track the evolution across time of the central tendencies and dispersion of survey expectations. Our results suggest that the model uncertainty and sticky information theories capture the time-variation reasonably well. Moreover, the model uncertainty

[2] Pesaran and Timmermann (1995) also find evidence of model uncertainty.

approach seems to provide a better fit than sticky information. This paper's aim is to, at a first pass, assess to what degree the theoretical models can explain the primary *dynamics* of survey expectations.

These results are new and significant. There is considerable interest by the monetary policy literature in whether agents have limited information or uncertainty about the true economic model. Our evidence suggests that model uncertainty plays a more important part of survey data but that sticky information is a feature as well. Based on the results of this paper, a high priority of future research should intertwine both sticky information and model uncertainty. We present evidence which suggests that during periods of high volatility agents' uncertainty about the economic environment is a key factor in expectation formation. During periods of low volatility, model uncertainty is less critical and agents may be inattentive. One methodological novelty of this paper is that it provides a measure of fit in terms of a model's ability to fit the evolution of the full distribution of survey expectations through time.

A few qualifications are in order. We acknowledge, at the outset, that our results do not address whether there is heterogeneity across multivariate and univariate models which also differ by updating frequency. We leave this study, as well as a dynamic version of Mankiw-Reis' approach in Branch, Carlson, Evans, and McGough (2004), to future research. Our interest is in whether model uncertainty or sticky information provide the closest fit to the data relative to alternative models of expectation formation. It is important to note, then, that our results are specific to classes of expectation formation models. These models are those most frequently used in macroeconomics. Our approach prevents us from making more general statements about *all* classes of models. Although the theoretical models of expectation formation do not provide a perfect fit, we argue that they capture important characteristics of the survey data and, therefore, the results here provide the first empirical comparison of heterogeneous expectations models. Section 3.2 discusses these issues and, in particular, emphasizes that we have endeavored to keep our models and empirical assumptions grounded in first principles and in line with the adaptive learning literature.

This paper proceeds as follows. In Section 2 we present the three expectation formation models. Section 3 discusses the maximum likelihood results. Section 4 compares the fit of the heterogeneous expectations models to the distribution of survey responses. Finally, Section 5 concludes.

2 Two Models of Sticky Information

Limited information flows as a component of expectation formation have been developed in Mankiw and Reis (2002), Sims (2003), and Branch, Carlson, Evans, and McGough (2004). In each of these models underlying expectations formation is a (costly) information gathering process. Because of the time and effort involved some agents may update their information sets infrequently. Unlike the adaptive learning literature (e.g. Evans and Honkapohja (2001)), agents in these settings have rational expectations but they do not condition on complete information. In Mankiw and Reis (2002) information arrives stochastically and so at any moment in time agents have heterogeneous expectations.

Supporting evidence for Mankiw-Reis' approach is found in Carroll (2003) and Mankiw, Reis, and Wolfers (2003). The evidence in favor of sticky information, though, is indirect. This paper extends Branch (2004) and Carroll (2003) by providing an analysis of the nature and robustness of sticky information in survey data. Our methodology compares three alternative models of expectation formation to survey data on inflation expectations.[3] The first two are models of sticky information: first, the Mankiw and Reis (2002) model of sticky information; second, a discrete choice model of sticky information inspired by Brock and Hommes (1997). The third approach is the model uncertainty case of Branch (2004).

2.1 Survey Expectations

This paper characterizes the twelve-month ahead inflation expectations in the Michigan survey. The data come from a monthly survey of approximately 500 households conducted by the Survey Research Center (SRC) at the University of Michigan. The results are published as the Survey of Consumer Attitudes and Behavior and in recent years as the Survey of Consumers. This paper uses the data in Branch (2004) which covers the period 1977.11-1993.12.

We focus on the period 1977.11-1993.12 because it covers a diverse spectrum of inflation volatility and to keep the comparison sharp with our earlier work.[4] This period is well-suited to our purposes because it includes changes in the level of inflation and a decrease in inflation volatility around 1984 (the Great Moderation). Although data is available through 1996 via the Interuniversity Consortium for Political and Social Research (ICPSR), the additional years will not change our results as the sample already consists of a long period of low volatility. The results presented below show that consumers are more likely to update their information sets less often during periods of economic stability. Additionally, our estimation strategy remains robust to structural change over the period.

This paper characterizes expectations of future inflation, and so the two relevant questions are:

[3] Three is a sufficient number of approaches because one of the models was shown by Branch (2004) to fit survey data better than other alternatives including Rational Expectations. Thus, the approaches here encompass the classes of forecasting models employed most frequently in the literature.

[4] There are some missing months in 1977, and so the sample is restricted only to continuous periods.

1. During the next 12 months, do you think that prices in general will go up, down, or stay where they are now?[5]

2. By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

The sample consists of 93142 observations covering 187 time periods. The mean response was 6.9550 with a standard deviation of 12.7010. The large standard deviation is accounted for by a few outliers that expect inflation to be greater than 40 percent. Excluding these responses does not change the qualitative results. Mankiw, Reis, and Wolfers (2003) extend the sample to prior periods by inferring a distribution from the survey sample. We note, though, that the key comparison periods in their sample, such as the Great Disinflation, are also covered in our sample. There are other surveys of inflation expectations. We focus on the Michigan survey because of its size and because it is more likely that a survey of consumers will exhibit sticky information than a survey of professionals. The Michigan survey is also ideally suited to a study of inflation expectations because it consists of households who all make decisions. Since many of these decisions are forward looking, an account of these expectations is an important issue.

In the Michigan survey each agent i at each date t reports their twelve-month ahead forecast $\tilde{\pi}^e_{i,t}$. As is standard in the adaptive learning literature, we assume that agents estimate a VAR of the form,

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t$$

which has the VAR(1) form,

$$z_t = A z_{t-1} + \tilde{\varepsilon}_t \tag{1}$$

where $z_t = (y_t, y_{t-1}, ..., y_{t-p+1})'$ and $\tilde{\varepsilon}_t$ is iid zero mean. If y_t consists of n variables then z_t is $(np \times 1)$ and A is $(np \times np)$. The specification and estimation of the parameters of the model are discussed below. We now turn to specifying sticky information within the context of this VAR forecasting model.

One point is worth stating here: because these expectations are phrased by the surveyor as twelve-month ahead forecasts, we assume agents look at past monthly inflation to forecast twelve-month ahead inflation. In other words, the VAR in (1) consists of monthly data. This assumption is also made in Mankiw, Reis, and Wolfers (2004).[6]

[5] Telephone operators are instructed to ask a clarifying question if respondents answer they expect prices to stay where they are now. That question is: Do you mean that prices will go up at the same rate as now, or that in general they will not go up during the next 12 months?

[6] The Labor Department in its press release reports first the monthly inflation figure. Thus, in a model of costly information gathering this should be the input to agents' forecasting model.

2.2 Static Sticky Information

In a series of papers, Mankiw and Reis (2002), Ball, Mankiw, and Reis (2003), and Mankiw, Reis, and Wolfers (2003) introduce a novel information structure to expectation formation. Unlike much of the bounded rationality literature they assume that agents have the cognitive ability to form conditional expectations (i.e. rational expectations). However, each agent faces an exogenous probability λ that they will update their information set each period. The information structure arises because costly information gathering leads agents to assemble information stochastically. It is important to stress that in the Mankiw-Reis approach expectations are not static; those agents who do not update their information sets still update their expectations.

Define I_{t-j} as the information set of an agent who last updated j periods ago; the set I_{t-j} consists of all explanatory variables dated $t-j$ or earlier. Using the VAR forecasting model (1), an agent who last updated j periods ago will form, in time t, a twelve-step ahead forecast of monthly inflation using all information available through $t-j$. Under these timing assumptions $j=0$ is equivalent to full-information rational expectations.[7] In order to form this forecast the agent must generate a series of i-step ahead forecasts of monthly inflation,

$$\pi^e_{j,t+i} = E\left(\pi_{t+i}|I_{t-j}\right) = \left(A^{j+i}z_{t-j}\right)_\pi$$

where $\left(A^{j+i}z\right)_\pi$ is the inflation component of the projection, with π denoting monthly inflation. The agent then continues out-of-sample forecasting in order to generate the sequence $\pi^e_{j,t+1}, \pi^e_{j,t+2}, ..., \pi^e_{j,t+12}$. The twelve month ahead forecast of annual inflation is generated according to,

$$\hat{\pi}_{j,t+12} = \sum_{i=1}^{12} \pi^e_{j,t+i}$$

It is worth emphasizing the forecasting problem facing agents. Agents estimate a VAR that consists of monthly data. Information arrives stochastically, and so given the most recent information agents forecast annual inflation by iterating their forecasting model ahead $j+12$ periods.

Given the exogenous probability of updating the information sets, a proportion λ of agents update in time t. Thus, at each t there are λ agents with $\hat{\pi}_{0,t+12}$, $\lambda\left(1-\lambda\right)$ with $\hat{\pi}_{1,t+12}$, $\lambda\left(1-\lambda\right)^2$ with $\hat{\pi}_{2,t+12}$, and so on. The mean forecast is,

$$\bar{\pi}^e_{t+12}(\lambda) = \lambda \sum_{j=0}^{\infty} \left(1-\lambda\right)^j \hat{\pi}_{j,t+12}$$

Below, in our empirical approach, we will sample from this sticky information distribution to generate a predicted survey sample. Carroll (2003) studies whether this

[7] Of course, because we have not posed a model for the economy these expectations may not be rational. Instead, these are the optimal linear forecasts given beliefs in (1). In the terminology of Evans and Honkapohja (2001) these are Restricted Perceptions.

mean forecast is consistent with the mean response of the Michigan survey when $\hat{\pi}$ comes from the Survey of Professional Forecasters (SPF). The approach of Mankiw, Reis, and Wolfers (2003) is analogous if professionals use a VAR to generate their forecasts. Below we will assess this model's ability to explain the entire distribution of Michigan survey responses.[8]

2.3 Rationally Heterogeneous Sticky Information

The Mankiw-Reis approach in the previous subsection is a heterogeneous expectations model. Agents engage in information gathering which leads to a fixed and symmetric probability of updating information sets. The result is a geometric distribution of expectations similar to the Calvo-style pricing structure emphasized in Woodford (2003). There are other heterogeneous expectations models. For instance, the seminal approach of Brock and Hommes (1997) can be applied to ascertain whether agents are distributed across predictors which differ in dimension of the z_t in (1).

This subsection presents a Rationally Heterogeneous Expectations (RHE) extension of Branch (2004) to limited information flows. We assume agents are confronted with a list of forecasting models distinct in the frequency of recursive updating. In each period agents choose their expectations from this list. This is an alternative to Mankiw-Reis in the sense that the choice of updating probabilities is purposefully chosen and (possibly) time-varying. There is a burgeoning literature on dynamic predictor selection. A prime example is the Adaptively Rational Equilibrium Dynamics (A.R.E.D.) of Brock and Hommes (1997). In the A.R.E.D. the probability an agent chooses a certain predictor is determined by a discrete choice model. There is an extensive literature which models individual decision making as a discrete choice in a random utility setting (e.g. Manski and McFadden (1981)). The proportion of agents using a given predictor is increasing in its relative net benefit.

Let $\mathcal{H}_t = \{\hat{\pi}_{j,t+12}\}_{j=0}^{\infty}$ denote the collection of predictors with information sets updated j periods ago. The static information alternative in the previous subsection generates mean responses by placing a geometric structure on the components of $\mathcal{H}_t$. In the alternative approach we assume that there are a finite number of elements in $\mathcal{H}_t$.[9] Moreover, unlike in the previous subsection, we assume that each predictor $\hat{\pi}_{j,t+12}$ is recursive and updated each $(j+1)$ periods.[10] It should be noted that the predictors are updated each $j+1$ periods since $j=0$ was designated above

[8] In Branch and Evans (2005), a recursively estimated VAR is used to forecast inflation and GDP growth and then these forecasts are compared to the SPF. A VAR forecast is found to fit well. We conjecture that replacing the VAR forecasts with the SPF, as in Carroll (2003), will not alter the qualitative results below.

[9] Brock, Hommes, and Wagener (2001) introduce the idea of a Large-type Limit (LTL) model of discrete predictor choice. In the LTL model there are an infinite number of predictors. We note that their approach is beyond the scope of this paper.

[10] In the static case, $\hat{\pi}_{j,t+12}$ is a $(j+12)$ step ahead out of sample forecast. As an alternative we allow updating in $\hat{\pi}_{j,t+12}$.

as full-information.[11] This differs from Mankiw-Reis in that the RHE approach no longer assumes expectations are rational; it imposes that agents ignore information that agents in Mankiw-Reis' approach would not. Although we defend this assumption below, we leave which approach is a better model of bounded rationality as an empirical question.

Let $U_{j,t}$ denote the relative net benefit of a predictor last updated $j+1$ periods ago in time t. We define $U_{j,t}$ in terms of mean square forecast error. The probability an agent will choose predictor j is given by the multinomial logit (MNL) map

$$n_{j,t} = \frac{\exp[\beta U_{j,t}]}{\sum_k \exp[\beta U_{k,t}]} \tag{2}$$

The parameter β is called the 'intensity of choice'. It governs how strongly agents react to relative net benefits. The neoclassical case has $\beta = +\infty$ and $n_{j,t} \in \{0,1\}$. Our hypothesis is that $\beta > 0$. Implicitly, Mankiw, Reis, and Wolfers (2003) impose the restriction $\beta U_{j,t} = \bar{U}_j, \forall t$. Our approach allows us to test this restriction. It is worth emphasizing that (2) is a (testable) theory of expectation formation. It formalizes the intuitively appealing assumption that the proportion of agents using a predictor is increasing in its accuracy.[12]

It is standard in the adaptive learning literature to assume that past forecast error is the appropriate measure of predictor fitness. The motivation is to treat the expectation formation decision as a statistical problem. In such settings mean-square forecast error is a natural candidate for measuring predictor success. Moreover, so long as predictor choice reinforces forecasting success, then alternative measures of fitness will not change the qualitative results.

The Rationally Heterogeneous Expectations (RHE) approach has the advantage over the Mankiw-Reis model that it does not *a priori* impose the structure of heterogeneity. Rather than a stochastic information gathering process, here the process is purposeful.[13] In this approach agents may switch between models with full information to models with dated information which implies agents may forget what they learned in the past. This structure is justified, though, since in the RHE approach each predictor is recursively re-estimated every j periods given recent data. If information gathering is costly then an agent may only incorporate the most recent data point when going from an infrequently updated to a frequently updated predictor. When going from a frequently updated to a less frequently updated predictor, though, it appears the theory imposes that agents dispose of useful information. However, this is logically consistent if one thinks of consumers forming expectations via 'market consensus' forecasts published in the newspaper – agents have acquired the forecast but not the forecasting method, thereby, not disposing of important information themselves. A fully specified model would also allow agents to choose, each

[11] Below we will change this (unfortunate) notation so that j is descriptive and represents the frequency with which the predictor is updated.

[12] The theory assumes that the forecast benefits are identical across individuals. A relaxation of this assumption is beyond the scope of this paper and is the subject of future research.

[13] Up to the noise in the random utility function.

period, how much 'memory' they have. This is intractable in the current framework and we leave this issue to future research.

2.4 A Model Uncertainty Alternative

Rather than examining heterogeneity in information updating, Branch (2004) examines heterogeneity in VAR dimensions. Specifically, suppose that agents choose from a set which consists of a VAR predictor, a univariate adaptive expectations predictor, and a univariate naive predictor. Model uncertainty leads agents to select from a set $\mathcal{H}_t = \{VAR_t, AE_t, NE_t\}$ where VAR_t is identical to the full-information forecast in the static sticky information model, AE_t is an adaptive forecast of the form

$$AE_t = (1-\gamma)AE_{t-1} + \gamma\pi_{t-1}$$

where $\gamma = .216$, and $NE_t = \pi_{t-1}$ is the naive forecast.

This theory restricts the set of predictors to VAR_t, AE_t, NE_t, which are representative of the most commonly used models of expectation formation. This set of predictors is meant to represent the classes of multivariate and univariate forecasting methods. If expectation calculation is costly, and agents are uncertain of the underlying macroeconomic model, they make a discrete choice from the set each period. Branch (2004) shows how heterogeneity across these models fits the data well in comparison to alternatives. One goal of this paper is to extend the set of alternatives to include sticky information predictors. We note that our results are robust to alternative multivariate and univariate predictors.

Our definition of model uncertainty may seem non-standard. Model uncertainty is typically expressed as ambiguity over the correct structural model for the economy, the Fed's interest rate rule, etc. Here model uncertainty is expressed as a forecasting problem: given costly estimation, what is the ideal forecasting model. This choice is not trivial in practice: the forecast advantage of a VAR or adaptive predictor relative to naive depends on the time period. The typical definition and ours are congruent, however. Brock and Durlauf (2004) argue that if agents are uncertain about their inflation regime – where the uncertainty stems from not knowing the Fed's monetary policy stance, for instance – then periods of stability may lead them to choose a naive or myopic predictor, but to switch to a VAR as past forecast errors accumulate. Switching between forecast models is meant to proxy for an underlying, deeper sense of model uncertainty. Then 'disagreement' in the profession over the correct economic model may manifest itself in consumers' expectations.

Agents are distributed across these predictors according to the MNL (2). Branch (2004) found that there is considerable variation across time in the $n_{j,t}$. Moreover, the model with agents split heterogeneously across these three models fits better than alternative approaches such as rational expectations and adaptive expectations. This paper extends the previous work by making a comparison to a class of sticky information models which are not special cases of the model uncertainty case.

It has been suggested by Williams (2003) that model uncertainty is the most plausible explanation for observed heterogeneity in survey data. The disagreement among macroeconomists over the appropriate structure of the U.S. economy may be a likely source of heterogeneity in survey expectations. A prime objective of this paper is to determine whether model uncertainty or sticky information can best explain heterogeneity in survey data. Thus, we compare the RHE-model uncertainty approach to the two sticky information alternatives.

3 Empirical Results

This section presents results of an empirical analysis of the three alternative models of expectation formation in comparison to the survey data on inflation expectations. The approach taken in this paper is in many ways similar to Branch (2004). The RHE-model uncertainty approach is identical, though Section 4 examines its performance in fitting the entire distribution. The specification of the predictors in the RHE-sticky information model is distinct from our earlier work, but the discrete choice mechanism is similar. The sticky information models extend the earlier paper by considering predictors that differ in the frequency with which they are updated. As opposed to the RHE-model uncertainty approach we emphasize a recursive forecasting strategy.

3.1 Predictor Estimation

The first step in the empirical comparison of subjective forecasting models is a careful construction of the predictor functions. The information structure is complicated and there are subtle but important differences across approaches. We begin this section with a description of how we constructed the predictor functions. The construction follows the steps: first, specification of the VAR and vector of explanatory variables z_t; second, a description of predictor functions in the two sticky information alternatives; third, since we are interested in how agents actually forecast, we describe a process of recursive forecasting.

First, we describe the VAR which is the basis for forecasting.[14] We follow Branch (2004) and Mankiw, Reis, and Wolfers (2003) in assuming that the VAR(1) consists of monthly inflation at an annual rate, unemployment, and 3-month t-bill rates.[15] A VAR with this set of variables is parsimonious and forecasts inflation well. Our metric for forecast comparison is squared deviations from actual annual inflation. We focus on monthly inflation as a predictor of annual inflation to remain close to the manner

[14]We focus on VAR forecasting because it is an approximation to rational expectations and is a frequently employed forecasting strategy.

[15]The monthly annual inflation rate is the inflation rate from one month to the next annualized. We also note that because we are not testing for rationality, overlapping forecasts is not a concern.

in which the Labor Department releases the data to the public and the wording of the survey. We choose a lag length of twelve in order to minimize the Akaike Information Criterion. This VAR is used to generate a twelve-month ahead forecast of inflation.

There are several issues to pin down. A model of sticky information is an assumption on the information sets, which evolve stochastically. Each sticky information alternative makes specific assumptions on how agents' information sets evolve. The purpose of this paper is to see if either (or both) are consistent with survey data. However, the parameters of the VAR model may be time-varying and unknown by agents. As a result, we assume that when agents update their information they also update their estimates of the model's parameters. Because these issues are distinct from our earlier paper, this subsection discusses predictor estimation at length.

In the Mankiw-Reis model we assume that λ agents have forecasts based on the most recently available data, $\lambda(1-\lambda)$ have two step ahead out of sample forecasts based on the most recently available data from two periods ago, and so on. To construct the Mankiw-Reis forecasts we recursively generate a vector of out of sample forecasts. We then weight and sum these forecasts as described in the previous section. To test this model against the survey data we generate random draws from the distribution implied by this information structure and then compare the densities of these draws to the histogram of the actual survey data. Below we discuss how parameter estimates are updated in this context. We follow Mankiw, Reis, and Wolfers (2003) by fixing $\lambda = .1$. To ensure robustness of our results, we also let $.05 \leq \lambda \leq .25$. All qualitative results are robust to values of λ in this range.

In order to formulate a tractable empirical model of dynamic predictor selection we must impose bounds on $\mathcal{H}_t$. At this point we make a notational change which will ease exposition. In the previous section, $j = 0$ denoted full-information. To stress that in the RHE setting full-information is equivalent to updating every period we now denote $\hat{\pi}_1$ as the predictor updated each period. We assume that

$$\mathcal{H}_t = \left\{\hat{\pi}_{1,t+12}, \hat{\pi}_{3,t+12}, \hat{\pi}_{6,t+12}, \hat{\pi}^e_{9,t+12}\right\}$$

That is, the available predictors update information every period, every third period, every sixth period, and once very nine periods. We make these restrictions in order to maximize the number of identifiable predictors. We omit VAR's estimated every other period, every fourth period, and so on, because they produce forecasts too closely related to the predictors in $\mathcal{H}_t$.

We also omit VAR's estimated less frequently than every nine months because it seems unlikely agents will update less than once every twelve months. Moreover VAR's estimated every 12 (or 24) months will produce forecasts similar to the 9 month predictor and our estimation strategy will not be able to separately identify agents with a 9 or 12 month predictor.[16] It is important to note that this bound does not affect the qualitative results. An advantage to the empirical procedure below is it

[16]This follows because to construct forecasts on 9 or 12 month predictors we are iterating a VAR whose parameter matrix has eigenvalues inside the unit circle.

will identify those agents that update less than once a year as $\hat{\pi}_9$ which still implies existence of sticky information.

Given a method for updating the models in $\mathcal{H}_t$ and the discrete choice mechanism (2), all that is needed to make the RHE sticky information model well-defined is a predictor fitness measure,

$$U_{j,t} = -\left(\pi_{t-1} - \hat{\pi}_{j,t-1}\right)^2 - C_j \equiv -MSE_{j,t} - C_j \tag{3}$$

where C_j is a constant around which the mean predictor proportions vary. Agents are assumed to base decisions on how a given predictor forecasted the most recent *annual* inflation rate. In Brock and Hommes (1997), C_j plays the role of a cost; predictors with higher computation or information gathering costs will have a larger C_j. However, the theory itself is more flexible and the C_j's may actually pick up predisposition effects. Essentially, the C_j ensure that the empirical estimates of the proportions of agents most closely fits the data. The C_j act as thresholds through which forecast errors must cross to induce switching, by agents, between prediction methods. This role for the constants is consistent with the role of costs in Brock and Hommes (1997) and is discussed in detail in Branch (2004). We note briefly that predictors estimated more frequently produce lower mean square errors, however, in any given period a sticky information predictor could produce a lower forecast error.

A brief justification of the form of (3) is warranted. Mean square error as a fitness measure can be derived from quadratic utility and is consistent with Brock and Hommes (1997), Branch and Evans (2004), and Evans and Ramey (2003). While mean-square error as a metric for forecasting success is motivated by theory, the weighting of past data in the MSE measure is an empirical question. We use only the most recent forecast error not as an *ad hoc* assumption for convenience, but because preliminary explorations indicated it provided the best fit for the sticky information model. It is worth noting that in our earlier paper we assumed a geometric weighting on the past squared errors in the RHE-model uncertainty approach. The qualitative results were robust to the weight placed on past forecast errors and the best fitting weight was close to one. This theory of 'purposeful' sticky information assumes agents look at the relative past success of a sticky predictor and are more likely to choose, in any period, the one with the greatest accuracy.

We now describe how the forecasts are actually computed. We follow the learning literature and deviate from Branch (2004) by assuming agents engage in real-time learning by recursively updating their prior parameter estimates.[17] In VAR's with time-varying parameters and limited samples, recursive estimation is desirable. Our approach is a straightforward extension of Stock and Watson (1996) to a setting where the parameters are updated periodically with the most recent data point. Each forecast function differs in how often it recursively updates its parameter estimates. This approach is motivated by costly information gathering that induces agents to

[17] Some authors advocate forecasting based on real-time data sets (see Croushore and Stark (2002). Such an undertaking is beyond the scope of this paper.

sample recent data periodically and to update their prior parameter estimates at the time of sampling.

The full VAR z_t is estimated according to

$$z_t = A_{t-1} z_{t-1} + \tilde{\varepsilon}_t$$

where

$$\begin{aligned} A_t &= A_{t-1} + t^{-1} R_t^{-1} z_{t-1} \left(z_t' - z_{t-1}' A_{t-1}' \right) \\ R_t &= R_{t-1} + t^{-1} \left(z_{t-1} z_{t-1}' - R_{t-1} \right) \end{aligned}$$

Note that R_t is the sample second moment of z_{t-1}.[18] Since there are 3 variables and 12 lags, the vector z_t is (36×1) and A_t is (36×36). These recursions constitute Recursive Least Squares (RLS). Each predictor is a VAR whose parameter estimates are generated at different frequencies. In the static sticky information alternative only λ agents have these expectations, while $\lambda (1-\lambda)^j$ form projections based on A_{t-j-1}.

It should be noted that the full VAR forecasting approach is standard and has been demonstrated elsewhere to produce good forecasts (e.g. Stock and Watson (1996) and Branch and Evans (2005)). One concern is that because of the Great Moderation – and other structural changes after 1960 as pointed out by Sims and Zha (2005) – inflation is not stationary and our VAR forecasting model may under predict inflation in the early years and over predict in later years. The recursive forecasting approach addresses this concern by allowing for time-varying parameters which remain alert to possible structural change. There is a long history of employing VAR's of this form, with this set of variables, by both professional and academic forecasters.[19]

For the RHE sticky information model, denote $z_{j,t}$ as the VAR updated every j periods. Each restricted VAR $z_{j,t}$ is updated every jth period according to,

$$A_{j,t} = \begin{cases} A_{j,t-1} + t^{-1} R_{j,t}^{-1} z_{j,t-1} \left(z_t' - z_{j,t-1}' A_{j,t-1}' \right) & \text{every } j\text{th } t \\ A_{j,t-1} \text{ otherwise} \end{cases}$$

A similar updating rule exists for $R_{j,t}$ as well. The term t^{-1} defines a decreasing gain sequence since it places lower weight on recent realizations. Some recent models with learning emphasize recursive estimates generated with a constant gain sequence instead. By replacing t^{-1} with a constant, a greater weight is placed on recent realizations than distant ones. We note that our results are robust to RLS parameter estimates generated by a constant gain algorithm. Given an estimate for $A_{j,t}$ a series of twelve one-step ahead forecasts are formed, just as in the static sticky information alternative, to generate a forecast of annual inflation.

In the estimation we set the initial parameter estimates equal to its least squares estimates over the period 1958.11-1976.10. From 1976.11 onwards each predictor

[18]See Evans and Honkapohja (2001) for an overview of real-time learning using recursive least squares.

[19]We refer the reader to Branch and Evans (2005), Stock and Watson (1996), and Mankiw, Reis, and Wolfers (2003).

updates its prior parameter estimates every j periods. This implies that the parameter matrices $A_{1,t}, A_{3,t}, A_{6,t}, A_{9,t}$ will be different across all $t > 1976.11$. This assumption is logically consistent with the underlying model of sticky information. The premise is that agents periodically sample recent data realizations and then form good forecasts based on that data. Implicit in the RHE specification of sticky information is that when agents update information they only acquire the most recent data point; that they do not go back and try to find all new information since the last update is a natural assumption if information gathering is costly. A downside is that as agents switch from models that are updated frequently to those that are updated infrequently agents will be disposing of information acquired in previous periods. This is logically consistent if agents are boundedly rational in the sense that they choose forecasts, $\hat{\pi}_{j,t}$, from a set of alternatives $\mathcal{H}_t$. Consistent with our model is the story that agents pick forecasts but not the forecasting functions themselves. This is as if agents sampled the newspaper infrequently for forecasts of inflation. Again, this is consistent with Carroll (2003) if the newspaper, or professionals' forecast, is derived from a VAR.

3.2 Discussion of Forecasting Models

This subsection discusses subtleties behind several of the modeling choices: the VAR (1) as a forecasting device; the restriction to this particular class of models; the forecasting strategy based on monthly data; recursive updating of parameter estimates in forecasting models; and, finally, the particular discrete choice recursive forecasting model.

Forecasts based on VAR's like (1) are employed extensively in macroeconomics. Our choice of a VAR forecasting approach is motivated by forming the best linear forecasts possible. Ideally we could stay close to the model of Mankiw and Reis (2002) which defines sticky information in terms of expectations conditional on the true probability distribution. Since the true distribution for the U.S. economy is unknown, we treat the VAR model as the best linear projection. In a sense, then, the VAR model approximates for Rational Expectations. Because of its simple structure it is ideally suited for developing forecasts based on limited information. One might also ask why agents do not just adopt professionals' forecast. Our approach is consistent with this alternative if professionals adopt a VAR forecasting approach. *The empirical analysis does not assume or rely on identifying VAR expectations with rational expectations.*

It is important to note that the results in this section are conditional on particular classes of models. We consider three classes: a static sticky information model with a geometric distribution of agents; a discrete choice sticky information model; and multivariate and univariate forecasting models. Branch (2004) showed that the RHE model uncertainty case fits the data better than alternatives such as full-information VAR expectations, rational and adaptive expectations, and other myopic expectation formation models. Thus, our results hold over a wide range of models typically employed by macroeconomists. Our results do not extend to classes of model uncertainty or sticky information not considered here. The restricted set of forecast models is not just for technical convenience, but also is representative of actual forecasting behavior. After all it is unlikely that people update their information more often than the government releases data (monthly).

Stock and Watson (1996) note that the best forecasting model, and estimation procedure, depends on the variables of interest. In Branch and Evans (2005) it is shown that the optimal value of the gain in RLS depends on the out-of-sample forecasting period if the degree of structural change is not constant over time. One might wonder whether allowing the VAR specification and recursive updating procedure to change, in order to always produce the best forecasts, might alter our results. However, one advantage to the empirical approach in this paper is that all that is necessary to identify agents with a predictor is that there is an ordering, in terms of forecast accuracy, on the set of predictors. There may be a VAR which forecasts slightly better and on which agents base their expectations. Our empirical approach, though, will still label agents as VAR forecasters. This paper is designed to identify the method, in the broadest sense, agents use to form expectations. The aim is not to provide a method for forecasting inflation.

One (possible) objection to the theory of RHE-model uncertainty is that agents should use Bayesian model averaging to deal with their uncertainty. We do not allow for a "model-averaged" predictor because the theory assumes that the choice made by agents is how sophisticated of a model they should employ given that expectation calculation is costly. Bayesian model averaging is another step in the sophistication and it should not alter our results. To defend our approach we also appeal to the model uncertainty story of Brock and Durlauf (2004): it is underlying uncertainty about the inflation regime that causes agents to adopt more sophisticated models when the simple predictors have done poorly in the past.

This paper assumes that the VAR model is based on monthly data. Agents are asked to forecast inflation over the next twelve months and are assumed to construct a series of one step ahead monthly inflation forecasts; the sum of these forecasts being their expectation of annual inflation. Monthly inflation as an input is motivated in Branch (2004) by the way the Labor Department releases the CPI data. Since monthly data is the most widely available, agents who face costly information gathering should forecast based on these data. In order to compare the sticky information model with the model uncertainty case we maintain the assumption here.

We stress recursive updating of forecasting models because it is most consistent with a theory of learning and is stressed in forecasting exercises in Stock and Watson (1996). We follow the approach developed in Branch and Evans (2005) for initializing the parameter estimates and ensuring that our results are not sensitive to choice of gain sequence. Although the sample length of the survey data is 1977.11-1993.12, we follow Stock and Watson's out-of-sample forecasting approach and initialize the VAR over 1958.11-1977.10. We do not present detailed results on mean square forecast error as they have been documented extensively elsewhere.[20] Recursive forecasting fits with the theory of sticky information well because forecasting based on parameter estimates over the entire period would make some agents condition, in part, on data not in their information set.

The recursive updating in the RHE sticky information case was chosen to remain close to Mankiw-Reis but with agents choosing their predictor each period. A discrete choice approach requires formulating a tractable and logically consistent model of expectation formation with sticky information. In the approach presented in this paper we assume that sticky information implies that agents periodically sample information. We argue that this is equivalent to using a predictor which is updated every j periods. We extend Mankiw-Reis by allowing agents to switch between these prediction methods. An implication of our approach is that agents may ignore information as they go from a predictor updated every j periods to one updated every $j' > j$ periods. We defend this result as logically consistent by appealing to a story in which agents choose forecasts not forecasting functions: they 'purchase' a forecast from a source who updates predictor functions according to the process described above.

Of course, empirical implementation of these theoretical models requires some trade-offs and choices. We have labored to minimize the *ad hoc* assumptions. For instance, the discrete choice mechanism comes directly from Brock and Hommes (1997), the predictors are identical to those in the learning and dynamic macroeconomic literature, and the fitness measure is standard in statistical learning models. In cases where theory does not provide guidance – such as the weight in the MSE predictor fitness measure or the number of sticky information predictors – we based our choice on empirical fit. Finally, as discussed above, empirical explorations suggest that our qualitative results are robust to the econometric choices.

3.3 Maximum likelihood estimation of RHE Sticky Information Model

This subsection presents results of estimation of the RHE sticky information model. In this subsection our objective is to test for dynamic sticky information and to estimate the distribution of agents across predictors. We achieve the first objective by testing $H_0 : \beta > 0$, the second objective is achieved by estimating the hierarchy of the $C_j, j = 1, 3, 6, 9$. First, a brief discussion of the estimation approach.

Although the predictors are different in the RHE-sticky information and RHE-model uncertainty cases, the specification of the econometric procedure, below, is based on Branch (2004). The theory predicts that survey responses take a discrete number of values; namely, there are only four possible responses. However, the survey itself is continuously valued. We specify a backwards multinomial logit approach.

[20] See, for example, Branch and Evans (2005).

Each survey response is assumed to be reported as

$$\tilde{\pi}^e_{i,t} = \hat{\pi}_{j,t+12} + \nu_{i,t} \tag{4}$$

where $\nu_{i,t}$ is distributed normally and $j \in \{1, 3, 6, 9\}$. A survey response is a twelve-month ahead forecast reported in time t. The theory behind (4) assumes agents choose a forecast based on past performance. After selecting a predictor, agents make an adjustment to the data, $\nu_{i,t}$, and report an expectation that is their perception of future inflation. The stochastic term $\nu_{i,t}$ has the interpretation of individual idiosyncratic shocks and is the standard modeling approach in RHE models. In particular, it is consistent with the findings of Bryan and Venkatu (2001 a,b) and Souleles (2001) who emphasize the difference between an expectation and a perception. Our model assumes that agents form expectations and report perceptions. In particular, $\nu_{i,t}$ can account for many of the idiosyncrasies reported in (Bryan and Venkatu 2001a,b) as well as difference in market baskets, etc.

Given the process (4), utility function (3), the density of an individual survey response $\tilde{\pi}^e_{i,t}$ is

$$P(\tilde{\pi}^e_{i,t}|MSE^t) = \sum_{l \in \{1,3,6,9\}} n_{l,t} P\left(\tilde{\pi}^e_{i,t}|j = l\right)$$

where

$$P(\tilde{\pi}^e_{i,t}|j = l) = \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left[-\frac{1}{2}\left(\frac{\tilde{\pi}^e_{i,t} - \hat{\pi}_{j,t+12}}{\sigma_\nu}\right)^2\right],$$

and $MSE^t = \{MSE_{1,t-k}, MSE_{3,t-k}, MSE_{6,t-k}, MSE_{9,t-k}\}_{k=0}^t$. The density is composed of two parts: the probability that the predictor is l, given by $n_{l,t}$; the probability of observing the survey response given the agent used predictor l. The log-likelihood function for the sample $\{\tilde{\pi}^e_{i,t}\}_{i,t}$ is

$$\begin{aligned} \mathcal{L} &= \sum_t \sum_i \ln \sum_{j \in \{1,3,6,9\}} \frac{\exp\{\beta[-(MSE_{j,t} + C_j)]\}}{\sum_{k \in \{1,3,6,9\}} \exp\{\beta[-(MSE_{k,t} + C_k)]\}} \\ &\times \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{\pi}^e_{i,t} - \hat{\pi}_{j,t+12}}{\sigma_\nu}\right)^2\right\} \end{aligned} \tag{5}$$

The Appendix provides details on the derivation of the log-likelihood function. The empirical procedure is to choose the parameters $\beta, C_j, j = 1, 3, 6, 9, \sigma_\nu$ which maximize the likelihood function. The maximum value of the log-likelihood function gives the relevant metric of fit.

The theoretical model of heterogeneous expectation formation assumes that the parameters β, C_j are constant throughout the sample. One can imagine scenarios where these deep parameters depend in some way on the economic environment. In such situations the MNL assumption may not be ideally suited and instead a non-parametric approach may be more appropriate in estimating the fractions assigned to each predictor. One goal of this paper is to determine which characteristics of the survey data are accounted for by the models of heterogeneous expectations. Thus, a completely non-structural econometric strategy is orthogonal to the intent of this study and is best left for future research.

This section tests whether the dynamic specification fits the data better than a static version, i.e. that $\beta > 0$, and to provide an estimate of the distribution of predictor proportions. We can also test an implication of the Mankiw-Reis model. Although the static sticky information model is not nested in the RHE approach, it makes two testable implications, that $\beta = 0$ and that $C_1 < C_3 < C_6 < C_9$. To test this implication, we also test the hierarchy of the constant and that the distribution of agents is geometric. We conduct the analysis by obtaining maximum likelihood parameter estimates of (5).

Table 1 presents maximum likelihood parameter estimates for the RHE sticky information model. Identification of the parameters in the model requires normalizing one of the C_j's to zero. For this reason, the parameter results in Table 1 are segmented by normalization. The results are robust across normalization.[21] Estimates of β, the 'intensity of choice,' are on the order of about .14. Although, quantitatively the parameters vary across normalization it is straightforward to test that they yield similar predictor proportion estimates. Calculating the correlation of estimated predictor proportions across normalizations yields correlation coefficients of .99 and above.

INSERT TABLE 1 HERE

Table 1 also presents estimates of the constant or 'cost' parameters. As was mentioned above these parameters ensure that the mean predictor proportions fit the data best. Under each normalization the predictor updated every 6 months carries the lowest cost. Following the 6-month predictor the ordering is $C_3 < C_1 < C_9$. This implies that, on average, the predictors updated every 3 and 6 months are used by a greater proportion of agents than the predictors updated very frequently ($j = 1$) and infrequently ($j = 9$). Although this structure is different than the static sticky information alternative–where the highest proportion of agents update each period–these results are intuitive. Given the low volatility in the annual inflation series, we should expect that if agents are *predisposed* to not updating information every period, then the 3 and 6-month predictors should be the most popular: all else equal, lower cost implies higher proportions which use that predictor.[22]

That we have bound the least updated predictor at 9 months – while the static sticky information model has updating every 12 months on average – does not affect this result. We placed the bound because predictors updated every 9 months or less produce forecasts too similar and create an identification problem. As mentioned above, the empirical strategy will identify agents who update every 12 or more months

[21]For a discussion of identification in these types of models see McFadden (1984).

[22]This intuition implicitly assumes that people are inclined to update more than once a year.

as using the 9 month predictor. If it were possible to identify a distinct 12 (or 24) month predictor the qualitative results would be identical.

The finding that the constants, or 'costs', lead to the nine-month predictor carrying the highest cost is not paradoxical. In the A.R.E.D. the cost acts as a threshold that forecast errors must cross to induce agents to switch forecast methods. We interpret these parameters also as a threshold or predisposition effect. Empirically, they ensure that the estimated proportions fit the data best.

These results shed light on the nature of sticky information in survey data. We proposed two alternative models of limited information flows. The first was a static model with a geometric distribution of agents across models. The second is a dynamic model of RHE sticky information. We tested the hypothesis $H_0 : \beta = 0$ and found a log-likelihood value of -1.4619×10^6, thus, in a likelihood ratio test, we reject the null that $\beta = 0$. Moreover, our estimates of the constants suggest that agents are not distributed geometrically. These results indicate that if sticky information exists in survey data, it takes a dynamic form. We emphasize that these results exist when we restrict ourselves to the class of sticky information models. The next section expands the comparison to a larger class of non-nested models.

There are two distinctions in the RHE-sticky information from the static sticky information model: each predictor is updated at different rates; the proportions across these predictors are time-varying. The restriction that $\beta = 0$ is the case where there is heterogeneous updating but with fixed proportions. Thus, the test of the restriction $H_0 : \beta = 0$ is a test of whether these two distinguishing properties are jointly significant. As will be seen below, it is the time-varying proportions that are important in accounting for the evolution of survey expectations.

These results are interesting in the context of Carroll (2003), who finds that 'inattentiveness' is a distributed lag of professional forecasts in the SPF. Our results provide evidence that the length of the lag may vary across agents and time. Below we provide further evidence which suggests a generalization of Carroll (2003) will provide the best fit to the survey data.

Figure 1 plots the estimated predictor proportions. These proportions are estimated by simulating (2) and (3) with the parameter estimates in Table 1. Figure 1 illustrates the results from Table 1. On average, predictors 6 and 3 are used most frequently, followed by predictors 1 and 9 in that order. The most striking feature of Figure 1 is the volatility around the mean predictor proportions. Although the predictors 3 and 6 are used most often, on average, there are times when few agents are identified with them. In fact, at times many agents have full-information while other periods a scant minority do. Most of the volatility in predictor proportions occurs during the Great Inflation and Disinflation, in line with the hypothesis of Brock and Durlauf (2004). The figure also demonstrates how the constants create a threshold or predisposition effect. It is only when forecast errors rise above this threshold that the proportions of agents decrease from their mean values. We conclude that dynamic sticky information is consistent with the survey data.

INSERT FIGURE 1 HERE

It is worth emphasizing that these maximum likelihood results are conditional on there being sticky information in the survey data. Given that there is sticky information, we find that a dynamic specification fits best. Our results do not directly address the static sticky information model of Mankiw-Reis. However, our results are suggestive of a dynamic specification over a static specification, and a different distribution of agent-types than assumed by the Mankiw-Reis model. Below we conduct a comparison to a non-sticky information model and the static sticky information model.

4 Fitting the Full Distribution

The previous section tested whether dynamic predictor selection is consistent with the survey data. The distribution of agents across predictors was time-varying and distinct from Mankiw-Reis' static sticky information approach. These results are not evidence that the RHE sticky information model fits the survey data better than the Mankiw-Reis model. Because of the timing differences between the two models, it is not possible to nest the Mankiw-Reis approach as a testable restriction of the RHE approach; the RHE approaches have more free parameters. Similarly, it is also not possible to present likelihood evidence in favor or against RHE sticky information *vis a vis* the RHE model uncertainty of Branch (2004). It is because these models are all non-nested that we turn in this Section to a non-parametric approach. We argue though that although a formal test of model uncertainty against sticky information is not possible, an empirical comparison is nonetheless a useful exercise.

Because these three approaches are not nested, the metric for model fit should be its ability to explain the full distribution of survey responses. We have already shown that there is heterogeneity in survey responses. Moreover, this heterogeneity is time-varying. A complete model comparison should study which theoretical model of expectation formation tracks the shifting distribution of survey responses across time.

A theoretical model of heterogeneous expectations suggests two channels for a time-varying distribution of survey expectations. The first is through the distinct response of heterogeneous forecasting models to economic innovations. The second is through a dynamic switching between forecast models. The first channel is implied by both the static sticky information model and the two RHE models. The Mankiw-Reis approach implies a time-varying distribution because agents' beliefs will adjust differently to economic shocks based on how frequently they update their information sets. Mankiw, Reis, and Wolfers (2003) demonstrate that a static sticky information model, during a period such as the Great Disinflation, may produce 'disagreement' in the form of a multi-peaked distribution with skewness which varies over time. The RHE approaches also may yield divergent expectations because each heterogeneous forecasting model may respond distinctly to economic shocks. The RHE models are also consistent with the second channel: agents dynamically select their forecasting model based on past forecast success. Thus, the dynamic predictor selection mechanism, at its very core, is a theory of time-varying distributions. This Section examines to what extent each theory can account for the period-specific distributions of survey responses.

In order to study which model can best account for the time-varying distribution of survey responses, in this subsection, we compare the density functions of the RHE sticky information model, static sticky information model, and the RHE model uncertainty approach to the histogram of the actual survey data. Our interest is not to compare the fit to the entire sample of survey data, but how each model fits the survey data in *each* period. The hypotheses are: the evolution of the distribution of survey responses results from the dynamics of the economy when expectations are formed according to the static sticky information model; the change in the distribution of survey responses is because agents adapt their predictor choice and so the degree of heterogeneity is time-varying. A novelty to our paper is that we are able to study to what extent these hypotheses are confirmed by the data.

4.1 Non-Parametric Estimation of Density Functions

Our methodology is non-parametric estimation of the density functions and the histogram of the survey data. We make sample draws from the estimated distribution functions of all three alternative approaches. Taking the histogram of the survey data as the density function of the true model, we construct non-parametric estimates of the model density functions and compare their fit with the histogram from the actual data set. This allows a test for whether any of the models are the same as the true economic model of expectation formation. We also provide a measure of 'closeness' between these models and the data. We follow White (1994) and conclude that the model which yields the smallest measure between densities is also the model most consistent with the data. This conclusion, though, is more informal than the preceding analysis since it is not possible to put this conclusion to a testable hypothesis.

To estimate the density functions we make 465 draws from the distribution defined by each model in each time period.[23] For the RHE sticky information model (2), (3) and (4) defines a density function given the estimated parameters in Table 1. From this estimated density function we generate a sample of predicted survey responses. Given an assumption on the information flow parameter λ, it is also possible to make random draws from the static sticky information model's distribution. We follow Mankiw, Reis, and Wolfers (2003) in fixing $\lambda = .10$.[24] Along these lines, we draw

[23]A sample size of 465 is approximately the mean size each period of the Michigan survey.

[24]It is straightforward to choose a λ which minimizes the distance, in a measure-theoretic sense, between the density of the actual data with that of the Mankiw-Reis density (which is a function of λ). We instead pick $\lambda = .10$ in order to keep the analysis as close to Mankiw, Reis, and Wolfers (2003) as possible. Though to ensure robustness, we checked the qualitative conclusions when

from the same distribution estimated in Branch (2004).

The first question to address is whether these three models of expectation formation make distinct implications about the economy. The main distinctions between these three models is not how they forecast data – the VAR in each model forecasts well – but whether different choice sets of predictors have distinct implications for survey responses. This is the issue addressed in this Section and the first evidence is presented in Figure 2. Figure 2 plots the mean responses of these draws from the estimated density functions. The figure makes it clear that each model yields quantitatively different predictions about survey responses. For instance, the static sticky information model adjusts slowly to the Great Disinflation as it takes time for information to disseminate through all agents' information sets. The RHE-model uncertainty responds to changes in the economic environment more quickly as agents adjust to past forecast errors. Since each approach is essentially a distribution of agents over heterogeneous autoregressive forecast models, it is the differences in the heterogeneous expectations models which leads to the differences in Figure 2.

INSERT FIGURE 2 HERE

We turn to a test of whether any of these models coincide with the true data generating process. We construct non-parametric estimates of the densities and the histogram of the actual survey data. We then test whether these estimated densities are statistically identical to the density of the actual survey data, construct measures of the closeness of the densities to the actual survey data, and present plots. The Appendix provides details on the construction of these estimators and hypothesis tests.

We first present hypothesis test results. Following the framework in Pagan and Ullah (1999), denote $d(\pi^e), f(\pi^e), g(\pi^e), h(\pi^e)$ as the true densities of the Mankiw-Reis sticky information model, the RHE sticky information model, the RHE model uncertainty case and the survey data, respectively. The Appendix details construction of the estimates $\hat{d}, \hat{f}, \hat{g}, \hat{h}$. We are interested in the following hypotheses,

$$\begin{aligned} H_0 : \hat{d}(\pi^e) &= \hat{h}(\pi^e) \\ H_0 : \hat{f}(\pi^e) &= \hat{h}(\pi^e) \\ H_0 : \hat{g}(\pi^e) &= \hat{h}(\pi^e) \end{aligned}$$

We report two test statistics T, T_1 of these hypotheses which are distributed standard normal.

Table 2 reports the results of the hypothesis tests. The tests are computed monthly between 1979.1-1982.12. We report results for this period because it is emphasized in Mankiw, Reis, and Wolfers (2003). Although, Mankiw, Reis, and Wolfers (2003) conduct analysis over a much longer time period, they hypothesize that sticky information should be most evident during the Great Disinflation. In each case, we

$.05 \leq \lambda \leq .25$.

reject the null hypothesis at the .01 significance level.[25] This suggests that none of the model alternatives are identical to the actual survey data generating process.

INSERT TABLE 2 HERE

That none of our alternative expectation formation models match up statistically with the survey data is not surprising. In each sticky information model and the RHE of Branch (2004), numerous tractability and identification assumptions are made. We first presented this stringent test for completeness. We now turn to other measures of fit besides hypothesis testing. That is, we instead turn to determining which approach provides the closest fit. We address this issue by constructing a measure of closeness between the estimated densities. The measure adopted here is the Kullback-Leibler distance measure of White (1994).[26] If the Kullback-Leibler measure equals a positive number then the area between two density functions is positive. We say that the model with the lowest distance measure is the model most consistent with how survey responses are formed. There is one important *caveat*: we do not have a formal hypothesis that one distance is statistically less than another. Thus, it may be that these distances are not statistically different. To bolster our informal conclusions we also present plots to visually compare these densities.

Figure 3 illustrates the Kullback-Leibler distance measure between the densities of the predicted responses for each model and the density of the actual survey responses. Figure 3 gives a clearer sense of the magnitudes of differences between the models and the true data generating process rather than a characterization of statistical significance. Figure 3 demonstrates that which model is the "closest" to the survey data is time-varying and period specific. On average, the RHE model uncertainty case fits best but not in each period. Following the RHE model is the RHE sticky information case and then the static sticky information case. In particular, the RHE model uncertainty case provides the best fit during the period of inflation and disinflation during the late 1970's and early 1980's. This is an intuitive result as the relative volatility of that period should induce agents to update their information frequently. However, because of disagreement over the appropriate model there is heterogeneity in expectations. Over the period 1987-1990 the RHE sticky information fits best. The two RHE cases provide a closer fit than the static sticky information case.

INSERT FIGURE 3 HERE

These findings suggest a story of expectation formation in which agents have a mix of model uncertainty and inattentiveness. During periods of economic volatility, like the 1970's, agents are attentive but uncertain about the economic structure. During the 1980's and 1990's agents tend to be less concerned with model uncertainty and, because of the stability, may be inattentive.

[25]We note that estimates of whether the models are the same as the actual data over the whole sample period are also rejected.

[26]Details are in the Appendix.

(Mankiw, Reis and Wolfers 2003) note that the hallmark of sticky information should be a multiple-peaked density function during disinflationary periods. To compare the shape of the density functions Figures 4-6 plot the estimated densities and survey histogram for particular times during 1979.1-1982.12. Periods 1979.4, 1981.4, 1982.12 correspond to times when RHE sticky information, RHE model uncertainty, and static sticky information, respectively, produce the lowest Kullback-Leibler distance measure. Figure 4 is for the case where RHE Sticky Information dominates, Figure 5 is for the case where the Mankiw-Reis approach provides the best fit, and Figure 6 is one period where the RHE of Branch (2004) is the closest to actual data. In Figure 4 there is a double peaked shape to the static sticky information model as found in Mankiw, Reis, and Wolfers (2003). In Figure 4 the RHE sticky information density has four peaks. The RHE sticky information and Mankiw-Reis sticky information may lead to different shaped density functions because the RHE sticky information allows for the degree of sticky information to change over time. That there are periods where the multi-peaked density fits best suggests that elements of each model may be important in explaining survey data. These figures, though, show that model uncertainty as in Branch (2004) can also account for multiple-peaked histograms in the survey data. This result was suggested by Williams (2003) that agents split across models could account for 'spikes' in the histograms.

INSERT FIGURES 4-6 HERE

These results suggest that the RHE model uncertainty approach provides the closest fit to the survey data when compared to two classes of sticky information models. When restricted to sticky information models a time-varying RHE model provides a better fit, on average, than the static approach. However, there are periods where the sticky information models provide a better explanation than model uncertainty. Based on the evidence in figures 4-6 this conclusion may seem based on weak evidence. Below we present additional evidence to bolster this conclusion. It follows from the results in this paper that a fully specified model which includes model uncertainty and dynamic sticky information will provide the best and most compelling fit of the survey data. Such an examination, though, is beyond the scope of this paper and left to future research.

It is possible to provide greater detail by constructing confidence intervals around the empirical distributions of the theoretical expectation formation models. We turn to an examination of which confidence interval best 'covers' the Michigan survey data.

To undertake this further analysis, we first construct 95% confidence intervals for the empirical distributions of the RHE model uncertainty and static sticky information approaches.[27] These confidence intervals are subsets of $\mathbb{R}^2$. To construct a measure of 'coverage' – that is, what proportion of the survey data lie within these

[27] We focus on these two models to economize space. As suggested by figure 3, and verified by our own explorations, the two RHE models produce similar qualitative results relative to the static sticky information model.

confidence intervals – we examine separately discrete sections of the confidence interval: for any given survey expectation value, and any given period, we calculate the proportion of the survey sample which reports that value. We then check whether this proportion falls in the 95% confidence interval of either the RHE model uncertainty or the static sticky information approach. The figures above suggest that there may be instances where the histogram of the actual survey data lies, at least in part, inside the 95% confidence interval of both approaches. Our desired measure of coverage is the percentage of these proportions out of the total number of cases considered in each period. For example, if in a given month we separate the confidence interval into 26 discrete survey responses coinciding with $-5, ..., 0, ...20$ then there are 26 cases considered and we calculate the proportion of these cases which lie in the various confidence intervals. Figure 7 reports the results.

INSERT FIGURE 7 HERE

Figure 7 plots the histograms for the coverage measure discussed above. On the horizontal axis is the percentage of cases which lie inside a confidence interval. The vertical axis is the number of survey sample periods in which this value was realized. The dark and light bars are for the coverage of the 95% confidence interval for the RHE model uncertainty and static sticky information approaches, respectively. The highest measure of coverage was .45 which was realized fewer than 10 times for both models. Figure 7 also demonstrates that the coverage is greater for the RHE model uncertainty alternative as the histogram is skewed toward higher values.[28] In a given sample period the RHE model uncertainty approach is more likely to have 25% of the survey responses within its 95% confidence interval than the sticky information.

Figure 7 also gives a greater quantitative sense in which these models explain the survey data. The hypothesis tests presented above – that the distributions are statistically identical in a measure theoretic sense – is demanding indeed. The Kullback-Leibler distance measures give a better sense of the fit of the empirical distributions to the actual data. Still, since distance in this section is defined as the area between two density functions it is, in a sense, an average rather than median discrepancy. Figure 7 instead illustrates the percentage of the survey data's histogram, in each period, that lies within the confidence interval of the empirical distributions.

Neither model explains the data perfectly. This is not surprising as survey expectations consist of idiosyncrasies for which simple economic theories set forth in this paper can not account. For instance, Bryan and Venkatu (2001a,b) and Souleles (2001) document demographic effects in the Michigan survey data. We argue, however, that the theories presented in this paper are the underlying basis behind the formation of these expectations. These theories predict that survey expectations should be heterogeneous with the distribution of expectations time-varying. The empirical evidence presented above demonstrates that both theories are able to capture this important characteristic of the data. Moreover, the evolution of the distribu-

[28]The coverage for the RHE sticky information model has similar skewness.

tion of the survey expectations is systematic, moving in much the same way as the theoretical models predict. The demographic characteristics may be able to capture some of the central tendencies, but not the time-varying distribution of the data. A precise characterization of the entire distribution is a daunting challenge for our simple theories of heterogeneous expectations.

Ultimately, this paper argues that the primary evidence in favor of the theoretical models of expectation formation is their ability to capture the evolution across time of the distribution of survey data. Figures 8-9 provide further evidence on how well the theoretical models capture the time-varying dispersion of the survey responses. Figures 8-9 report the Interquartile range (IQR) of the estimated density functions. The IQR is the computed difference between the 75th and the 25th percentiles of the estimated density functions, for each period of the sample. Tracking the IQR overtime gives an estimate of how the dispersion of the data changes over time. It has the advantage over Kullback-Leibler in that it ignores the tails of the distribution. Figure 8 scatters changes in the IQR for the RHE model against changes in the IQR for the survey data, and Figure 9 does similarly for the sticky information model.[29] Each figure also plots the trend line. The slope of the trend in Figures 8 and 9 are .167 and .0099, respectively.

In Figure 8 there is an upward trend suggesting that as the survey data becomes more disperse over time, then the model uncertainty theory of expectation formation predicts a greater dispersion of the survey data as well. In Figure 9 the trend line is increasing slightly, but the slope is not statistically different from zero. Figures 8-9 provide further evidence about which theoretical model fits the data best. The positive correlation between the RHE model's IQR and the Michigan survey's IQR demonstrates that the model uncertainty approach can explain, in part, the time-varying dispersion of the survey data. The main difference in the IQR's for the RHE and sticky information approaches is that the static sticky information approach is more likely to predict little or wide dispersion. This is a feature of the Mankiw-Reis structure of overlapping information updating: in periods of relative economic stability the degree of heterogeneity is small and as the economy switches to a period of instability the dispersion will be higher.

Although, the actual distribution of survey data may result from a more complicated stochastic process than that predicted by the RHE-model uncertainty alternative, it is remarkable that a simple model of heterogeneous expectations which predicts that the distribution of agents across predictors should evolve with past forecast errors, can help explain important properties of the entire distribution of survey responses. For these reasons, the models of heterogeneous expectations presented in this paper appear to be a good approximation of the actual expectation formation process.

[29] Plotting the level of IQR produces similar results as Figures 8-9. Also, the plot for changes in IQR for the RHE sticky information model produces an upward trend that is not statistically significant.

This section presented evidence regarding the fit of the theories of heterogeneous expectations. Both sticky information and model uncertainty can explain some of the evolution across time of the distribution of survey data. We also presented evidence which suggests the RHE approach provides a better fit. There are some limitations to the interpretation of these results. The models are not nested and so we present non-parametric evidence. This precludes a formal test, but the analysis is still instructive regarding general empirical implications of these theoretical models. Because the approach is non-parametric it is also, to a certain extent, subjective evidence. For this reason we emphasize again that both models can explain time-varying dispersion in inflation expectations. The time-varying proportions of agents using each predictor intuitively lead to the RHE model providing a better fit to this feature in the data. Because model uncertainty and sticky information are non-nested, and formally testing one against the other is not possible, it might seem more fruitful to alter one theory to take into account non-normal idiosyncrasies and try to better fit this distribution. Such an exercise may be useful, however, both model uncertainty and sticky information are receiving considerable attention in the theoretical literature. We argue that it is useful to have some basis for empirical comparison – even between non-nested models. While both theoretical approaches can produce non-degenerate distributions of expectations, the results of this section demonstrate qualitative differences.

4.2 Further Discussion

This Section presented evidence regarding the time-varying nature of the distribution of survey responses. Three candidate theoretical models have been advanced, each of which could plausibly account for the evolution of inflation expectations. The static sticky information model of Mankiw and Reis (2002) implies that, when information is dispersed across the population, sudden shocks to the economy can produce heterogeneous expectations; the distribution of expectations will evolve as information disseminates through the economy. We also argued that the distribution of agents across sticky information forecasting models may change over time. According to the RHE-sticky information model, heterogeneity may arise under sticky information and dynamic predictor selection. The third theory is that agents are split across heterogeneous forecasting models, rather than information sets, and that this distribution may change over time. A time-varying distribution of survey responses may arise as the forecast models change *and* the distribution of agents across these models varies over time. The results in the previous subsection suggest that the RHE-model uncertainty approach provides the best fit, on average, to the distribution of the survey data: the evidence also shows that each theory may provide the best fit in a particular period.

As mentioned in the introduction, these three theories do not constitute an exhaustive list. One alternative merits special attention. Fry and Harris (2002) present an interesting alternative theory which they term "digit preferencing." Fry and Harris demonstrate that survey responses, drawn from an Australian version of the Survey of Consumers, exhibit higher frequencies at round values such as 5%, 10%, 15%, and so on. They posit a theory of expectation formation where agents have a preference for certain digits. This preference depends on demographic characteristics as suggested by Bryan and Venkatu (2001a,2001b). An examination of the histogram of the Michigan data over the entire sample, for example, also shows multiple peaks at integer values $3\%, 5\%, 10\%$. If there is 'digit preferencing' then a multi-modal distribution is a possible outcome.

A priori, the digit preferencing theory or the heterogeneous expectations theories are plausible explanations of the survey data. Our interest in this paper is to assess whether the sticky information and model uncertainty approaches often applied in the dynamic macroeconomics literature are consistent with the data. Considering all alternative theories is beyond the scope of this paper. However, an empirical concern of this paper is whether the finding that the model uncertainty approach provides the best fit to the survey data results from assuming agents form expectations by selecting from just a few predictors. The finding that the RHE-model uncertainty fits the data best might be spurious if agents have 'digit preferencing': what we identify as rationally heterogeneous expectations would really be a model of "digit preferencing."

This subsection provides greater detail of the density functions to serve two purposes: first, for greater insight into the dynamic properties of the estimated density functions; second, to provide support that our findings in the previous subsection were not spurious. We argue that "digit preferencing" – that is, a multi-modal distribution – is also consistent with theories of heterogeneous expectations. As demonstrated by Fry and Harris (2002), digit preferencing manifests itself in the data as multiple peaks in the histogram of survey responses. The previous subsection mentioned that multiple peaks might appear in the density function from heterogeneous expectations and/or sticky information. Moreover, the theories advocated here suggest that the number and size of peaks in the histogram may be time-varying.

If "digit preferencing" is consistent with the theories in this paper, then the peaks in the density functions should vary over time. To bolster this argument we first note that the percent of the sample at values of either $3\%, 5\%$, or 10% varies across time. See, for instance, Table 3. This suggests that survey responses clustering around 3% is not because of an innate digit preference for 3% but because the economy is more likely to produce inflation of about 3%.[30] Over the sample period there are peaks in the distribution of actual annual inflation at values near $2\% - 5\%, 10\%, 11\%, 14\%$.

INSERT TABLE 3 HERE

Figure 10 summarizes the estimated number of peaks in the density functions for the survey data, RHE-model uncertainty, and static sticky information models. The

[30] A (potential) econometric concern is whether survey respondents report integer values because of "digit preferencing" (Fry and Harris (2002)) or because agents round off their true expectation. Our argument is that the theories of sticky information and heterogeneous expectations is consistent with time-varying distributions of survey responses. The central tendencies of the estimated densities tend to accord well with the data and adjusting the likelihood function to incorporate rounding is unlikely to alter the qualitative results of this paper.

left column plots the frequency (histogram) of peaks over the period 1979.1-1982.12 while the right column is for 1979.1-1993.12. It is evident in Figure 10 that the peaks and shape of the density functions varies over time. The top histogram shows that "digit preferencing" by agents, while apparent in the aggregate, is period specific. The bottom two histograms also show that each theory is consistent with multi-peaked distributions. In particular, that the static sticky information model can produce multiple peaks is supportive of Mankiw, Reis, and Wolfers (2003) and also implies that it is not the assumption of a finite number of RHE predictors which accounts for the finding that RHE-model uncertainty is a more robust element of the data than the sticky information. We base this conclusion on the finding in Figure 10 that the static sticky information, which is a weighted average of many out-of-sample forecast models, also produces multiple peaks. We can expect that increasing the number of predictors in the RHE-model uncertainty alternatives set will not alter our finding of multiple peaks in the density functions. Figure 10, along with the measures of fit in the previous subsection, provide striking evidence in favor of our theories of heterogeneous expectations.

INSERT FIGURE 10 HERE

While "digit preferencing" may be a plausible alternative to the theories advanced here, this paper studies whether a reasonable theory of heterogeneous expectations, motivated by micro-foundations, can account for time-variation in the distribution of survey responses? The novelty of the theories in this paper is that they provide a theoretical explanation for the evolution of the distribution of survey responses. While the results show that the heterogeneous expectations theories can not perfectly explain the distribution, they do a fine job at fitting the central tendencies (e.g., time-varying mean, variance, and skewness).[31] The appeal of the theories set forth in this paper is that the moments of the distribution are determined by the distribution of agents across forecasting model and/or the distribution of information across agents.

5 Conclusion

This paper examined sticky information and model uncertainty in survey data on inflation expectations. This paper achieved two objectives: first, a characterization of sticky information in survey data; second, a check of whether sticky information or model uncertainty is a more robust element of the data. We compared two models of sticky information against the Rationally Heterogeneous Expectations model uncertainty approach of Branch (2004). Our first model of sticky information was an application of the novel approach in Mankiw and Reis (2002). Our second model, was an extension of Mankiw and Reis (2002) to the framework of Brock and Hommes

[31] As a robustness check we also investigated which model best fits the variance and skewness of the period specific distributions of survey responses. The results of this less demanding test are analogous to those in this Section.

(1997) where we assume agents make a discrete choice between recursive forecasting functions which differ by the frequency with which they are updated.

We first characterized limited information flows in the survey data by restricting agents to a class of sticky information models. We show that a sticky information model with a time-varying distribution structure provides a better fit than the static approach of Mankiw and Reis (2002). We provide maximum likelihood evidence that, on average, the highest proportion of agents in the Michigan survey update their information sets every 3 to 6 months. A lower proportion of agents update their expectations every period and few agents update their expectations at periods of 9 months or more. We also provide evidence, like Branch (2004), that these proportions vary over time.

We also presented a test of whether any of the three expectation formation models imply a density function identical to the density of the true model. We reject the hypothesis that any of these models are identical to the data generating process. Instead we provide non-parametric evidence suggestive of which model lies closest to the data. Non-parametric evidence suggests that model uncertainty is a more robust element of the data than sticky information. We construct estimates of the density functions for each model and compare them to the histogram of the actual survey data. The model uncertainty case is closer, in a measure-theoretic sense, to the actual data than the sticky information models. However, this result holds, on average, and there are periods, particularly during the late 1980's and early 1990's, in which sticky information provides the best fit.

These results suggest that agents' expectations are a mix of uncertainty about the true economic model and inattentiveness to new data. During the 1970's and 1980's the high inflation volatility led agents to update each period but they switched between candidate models. During the late 1980's the relatively low inflation volatility made it possible for agents to be inattentive and not switch forecasting models.

These results are significant. There is considerable interest in the applied literature on the effects of model uncertainty and sticky information. This paper suggests that both are elements of the data. The models presented here, though, do not allow for sticky information across competing models of the economy. Future research should address this issue as it may present the best fit of the data.

Appendix

Log-likelihood Function Derivation

Recall, that the actual observed survey response is given by

$$\tilde{\pi}^e_{i,t} = \hat{\pi}_{j,t+12} + \nu_{i,t} \tag{6}$$

where $\hat{\pi}_{j,t+12} \in \{\hat{\pi}_{1,t+12}, \hat{\pi}_{3,t+12}, \hat{\pi}_{6,t+12}, \hat{\pi}_{9,t+12}\}$. The probability of using the jth predictor was given by the theoretical model as a MNL,

$$Pr(j|U_{j,t}) = n_{j,t} = \frac{\exp\{\beta\left[-(MSE_{j,t} + C_j)\right]\}}{\sum_{k\in\{1,3,6,9\}} \exp\{\beta\left[-(MSE_{k,t} + C_k)\right]\}}$$

Since $v_{i,t}$ is distributed normally, the density of $\tilde{\pi}^e_{i,t}$ is

$$P(\tilde{\pi}^e_{i,t}|MSE^t) = \sum_{k\in\{1,3,6,9\}} n_{k,t} P(\tilde{\pi}^e_{i,t}|j=k)$$

where $MSE^t = \{MSE_{j,t}\}_{j\in\{1,3,6,9\}}$ and

$$P(\tilde{\pi}^e_{i,t}|j=k) = \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{\pi}^e_{i,t} - \hat{\pi}_{j,t+12}}{\sigma_v}\right)^2\right\}.$$

Since the sample changes each period, the probability of observing the sample is given by the following density function:

$$\begin{aligned} &P(\tilde{\pi}^e_{i,t}, i=1,...,N, t=1,...,T|MSE^t, \mathcal{H}_t(\pi^t), t=1,...,T) \\ &= \textstyle\prod_t \prod_i P(\tilde{\pi}^e_{i,t}|MSE^t) \\ &= \textstyle\prod_t \prod_i \left\{\sum_{k\in\{1,3,6,9\}} n_{k,t} P(\tilde{\pi}^e_{i,t}|j=k)\right\}. \end{aligned}$$

Taking logs leads to the form (5).

Non-parametric Density Estimation

We discuss the details of the non-parametric density estimation in Section 3.3. Our approach uses the Rosenblatt-Parzen Kernel Estimator as detailed in Pagan and Ullah (1999). This approach computes an empirical density function. Essentially it replaces a histogram, which computes the number of observations in a given window-width, with a probability density function which assigns a probability mass to a given window-width. Thus, the kernel estimator is

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where $\hat{f}$ is the estimator of f, the true density, $\{x_i\}_{i=1}^n$ is the sequence of observed values, and h is the window-width. The function K is the kernel function which is usually chosen to be a well-known probability density function. Following Pagan and Ullah (1999) we choose K to be the pdf of a standard normal. The remaining issue is the selection of h. Clearly the estimator is sensitive to the choice of h. We note that Figures 4-6 are illustrative and not a test of the validity of the estimate density functions. The density hypothesis testing are robust to choices of h. Pagan and Ullah note that a popular choice of h is one that minimizes the Integrated Mean Squared Error, which is essentially a measure of both the bias and variance of the estimates. The recommendation of this approach is to set $h = n^{-.2}$ where n is the number of observations in the sample.

Non-parametric Density Hypothesis Tests

The text considers the test of whether two non-parametrically estimated densities are identical. That is, for two pdfs $f(x), g(x)$, the null hypothesis is $H_0 : f(x) = g(x)$.

In the text we consider three different hypothesis. Denote $d(x), f(x), g(x), h(x)$ as the densities of the Mankiw-Reis, RHE sticky information, RHE model uncertainty, and actual data respectively. We test the following three hypotheses:

$$\begin{aligned} H_0 : d(x) &= h(x) \\ H_0 : f(x) &= h(x) \\ H_0 : g(x) &= h(x) \end{aligned}$$

Pagan and Ullah (1999) detail test statistics for these null-hypotheses when the true density is unknown. These tests are based on Kernel estimates of the density functions. Pagan and Ullah show that the appropriate test statistics are:

$$\begin{aligned} T &= nh^{.5}\frac{\left(\tilde{I} - c_2(n)\right)}{\hat{\sigma}} \\ T_1 &= nh^{.5}\frac{\tilde{I}_1}{\hat{\sigma}} \end{aligned}$$

where

$$\begin{aligned} \tilde{I} &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) - 2K\left(\frac{y_i - x_j}{h}\right)\right] \\ \tilde{I}_1 &= \frac{1}{n^2 h}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) - K\left(\frac{y_i - x_j}{h}\right) - K\left(\frac{x_i - y_j}{h}\right)\right] \\ \hat{\sigma} &= \frac{2}{n^3 h}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[K\left(\frac{x_i - x_j}{h}\right) + K\left(\frac{y_i - y_j}{h}\right) + 2K\left(\frac{x_i - y_j}{h}\right)\right]\int K^2(\omega)d\omega \end{aligned}$$

Because K is Gaussian $\int K^2(\omega)d\omega$ can be estimated as $n^{-1}\sum_i K(\omega_i)$ where $\omega_i = \left(\frac{x_j - x_i}{h}\right)\left(\frac{y_j - y_i}{h}\right)$.[32]

Pagan and Ullah note that T, T_1 are distributed $N(0,1)$ if $h \to 0$ and $hn \to \infty$. In our estimation of T, T_1 we set $h = n^{-.2}$, but we note that our results are robust to values of $h \in (0,1)$.

In the text we reject each H_0, however, we are also interested in which density is closest to our estimate of the sample density. The text reports the Kullback-Leibler Information Measure as a measure of the distance between estimated densities. White (1994) calculates this measure as

$$I^* = \int_x f(x) log\left\{\frac{f(x)}{h(x)}\right\} dx$$

The measure I^* approximates distance in the sense that if $f(x) = h(x)$ in the appropriate sense, then $I^* = 0$. Thus, $I^* \neq 0$ iff $f(x) \neq h(x)$ for some x. We use our non-parametrically estimated density functions to compute this measure.

[32]See Pagan and Ullah (1999) for a justification.

References

[1] Aadland, David, 2003, "Cattle Cycles, Heterogeneous Expectations, and the Age Distribution of Capital," forthcoming *Journal of Economic Dynamics and Control.*

[2] Baak, Saan Joon, 1999, "Tests for Bounded Rationality with a Linear Dynamic Model Distorted by Heterogeneous Expectations," *Journal of Economic Dynamics and Control*, 23, 1517-1543.

[3] Ball, Laurence, N. Gregory Mankiw, and Ricardo Reis, 2003, "Monetary Policy for Inattentive Economies," forthcoming *Journal of Monetary Economics.*

[4] Branch, William A., 2004, "The Theory of Rationally Heterogeneous Expectations: Evidence from Survey Data on Inflation Expectations," *Economic Journal*, 114, 497.

[5] Branch, William A., John Carlson, George W. Evans, and Bruce McGough, 2004, "Monetary Policy, Endogenous Inattention, and the Inflation-Output Variance Tradeoff," mimeo.

[6] Branch, William A., and George W. Evans, 2004, "Intrinsic Heterogeneity in Expectation Formation," forthcoming *Journal of Economic Theory.*

[7] Branch, William A., and George W. Evans, 2005, "A Simple Recursive Forecasting Model," mimeo.

[8] Brock, William A., and Steven N. Durlauf, 2004, "Macroeconomics and Model Uncertainty," mimeo.

[9] Brock, William A., and Cars H. Hommes, 1997, "A Rational Route to Randomness," *Econometrica*, 65, 1059-1160.

[10] Brock, William A., Cars H. Hommes, and Florian O.O. Wagener, 2001, "Evolutionary Dynamics in Markets with Many Trader Types," forthcoming *Journal of Mathematical Economics.*

[11] Bryan, Michael F., and Guhan Venkatu, 2001a, "The Demographics of Inflation Opinion Surveys," *Economic Commentary*, Federal Reserve Bank of Cleveland.

[12] Bryan, Michael F., and Guhan Venkatu, 2001b, "The Curiously Different Inflation Perspectives of Men and Women," *Economic Commentary*, Federal Reserve Bank of Cleveland.

[13] Calvo, Guillermo, 1983, "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12, 393-398.

[14] Carroll, Christopher D., 2003, "The Epidemiology of Macroeconomic Expectations," *Quarterly Journal of Economics*, 118,1.

[15] Chavas, Jean-Paul, 2000, "On Information and Market Dynamics: The Case of the U.S. Beef Market," *Journal of Economic Dynamics and Control*, 24, 833-853.

[16] Croushore, Dean, and Tom Stark, 2002, "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" forthcoming *Review of Economics and Statistics*.

[17] Evans, George W., and Seppo Honkapohja, 2001, *Learning and Expectations in Macroeconomics*, Princeton University Press.

[18] Evans, George W., and Garey Ramey, 2003, "Adaptive Expectations, Underparameterization, and the Lucas Critique," forthcoming *Journal of Monetary Economics*.

[19] Fry, Tim R.L., and Mark N. Harris, 2002, "The DOGEV Model," Monash University Working paper.

[20] Heemeijer, P., C.H. Hommes, J. Sonnemans, and J. Tuinstra, 2004, "Forming Price Expectations in Positive and Negative Feedback Systems," mimeo.

[21] Hommes, C.H., J. Sonnemans, J. Tuinstra, and H. van de Velden, 2005, "A Strategy Experiment in Dynamic Asset Pricing," forthcoming *Review of Financial Studies*.

[22] Mankiw, N. Gregory, and Ricardo Reis, 2002, "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 4, 1295-1328.

[23] Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers, "Disagreement about Inflation Expectations," forthcoming *NBER Macroeconomics Annual 2003*, ed. Mark Gertler and Kenneth Rogoff.

[24] Manski, Charles F., and Daniel McFadden, 1981, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA. trilligator, Elsevier.

[25] McFadden, Daniel, 1984, "Econometric Analysis of Qualitative Response Models," *Handbook of Econometrics, Vol. II*, ed. Z. Griliches and M.D. Intrilligator, Elsevier.

[26] Pagan, Adrian, and Aman Ullah, 1999, *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

[27] Pesaran, M. Hashem, and Allan Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *The Journal of Finance*, 50, 4, 1201-1228.

[28] Sims, Christopher A., 2003, "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 3, 665-690.

[29] Sims, Christopher A., and Tao Zha, 2005, "Were there Regime Switches in U.S. Monetary Policy?" forthcoming *American Economic Review.*

[30] Souleles, Nicholas S., 2001, "Consumer Sentiment: Its Rationality and Usefulness in Forecasting Expenditure – Evidence from the Michigan Micro Data," forthcoming *Journal of Money, Credit, and Banking.*

[31] Stock, James H., and Mark W. Watson, 1996, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics,* 14, 1, 11-30.

[32] White, Halbert, 1994, *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

[33] Williams, John C., 2003, "Discussion of 'Disagreement about Inflation Expectations'," forthcoming *NBER Macroeconomics Annual 2003*, ed. Mark Gertler and Kenneth Rogoff.

[34] Woodford, Michael, 2003, *Interest and Prices: Foundations of a Theory of Monetary Policy,* Princeton University Press.

	Parameter Estimates					
Normalization	β	C_1	C_3	C_6	C_9	σ_ν
Predictor 1	0.1439 (0.3620$e-$007)	0	-0.6718 (0.2236$e-$007)	-1.344 (0.4486$e-$007)	0.1309 (0.1660$e-$007)	6.0044 (0.2897$e-$005)
Predictor 3	0.1363 (0.7483$e-$005)	0.3598 (0.1048$e-$005)	0	-0.8648 (0.2577$e-$005)	0.5651 (0.1947$e-$005)	6.0037 (0.2727$e-$005)
Predictor 6	0.1418 (0.0016$e-$006)	1.3132 (0.1029$e-$006)	0.6509 (0.0741$e-$006)	0	1.4327 (0.6530$e-$006)	6.005 (0.0145$e-$006)
Predictor 9	0.1442 (0.0331$e-$006)	-0.1182 (0.8447$e-$006)	-0.809 (0.4736$e-$006)	-1.4703 (0.0028$e-$006)	0	6.0048 (0.0968$e-$006)

Table 1. Maximum Likelihood Estimation Results.

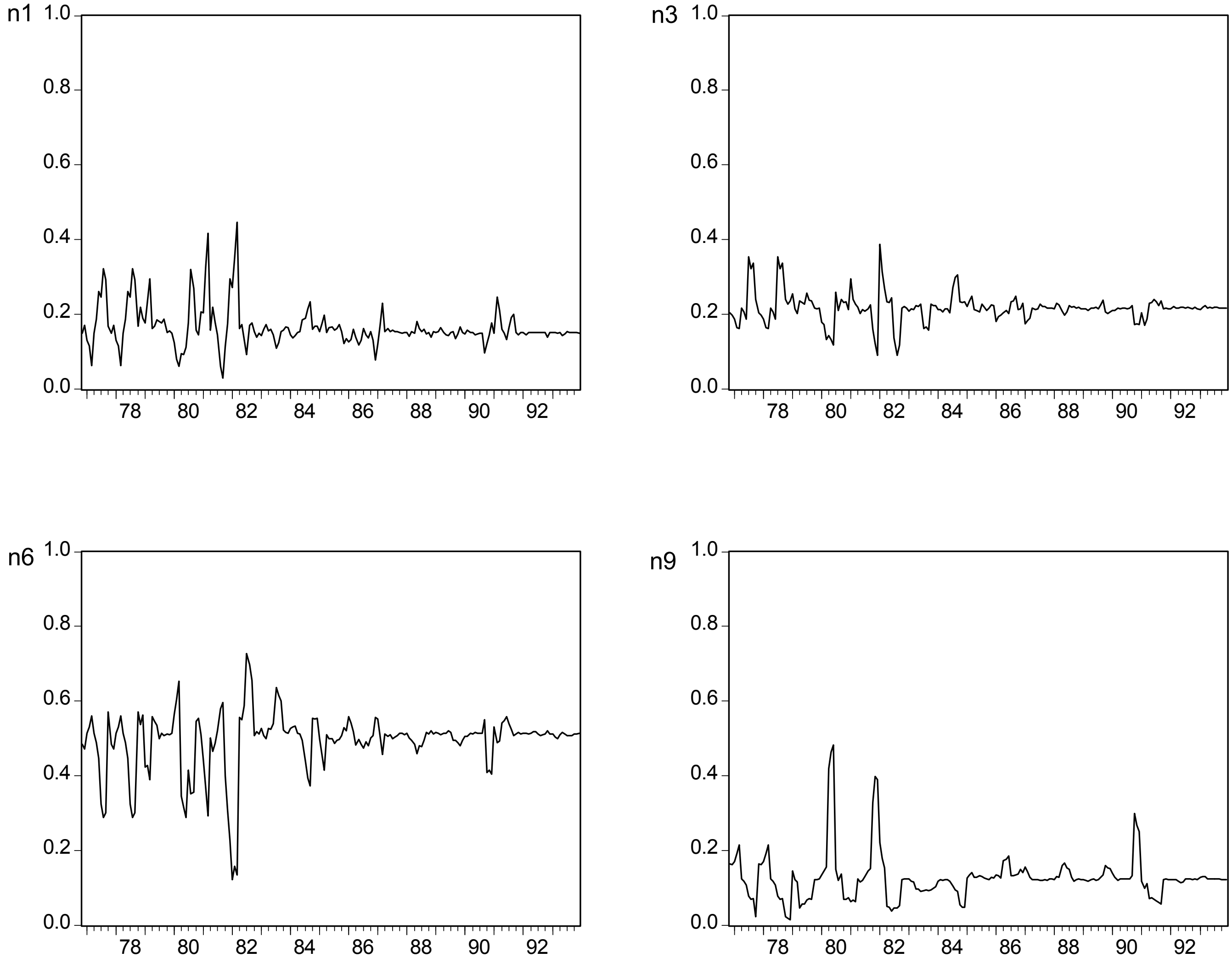

Figure 1. Estimated proportions.

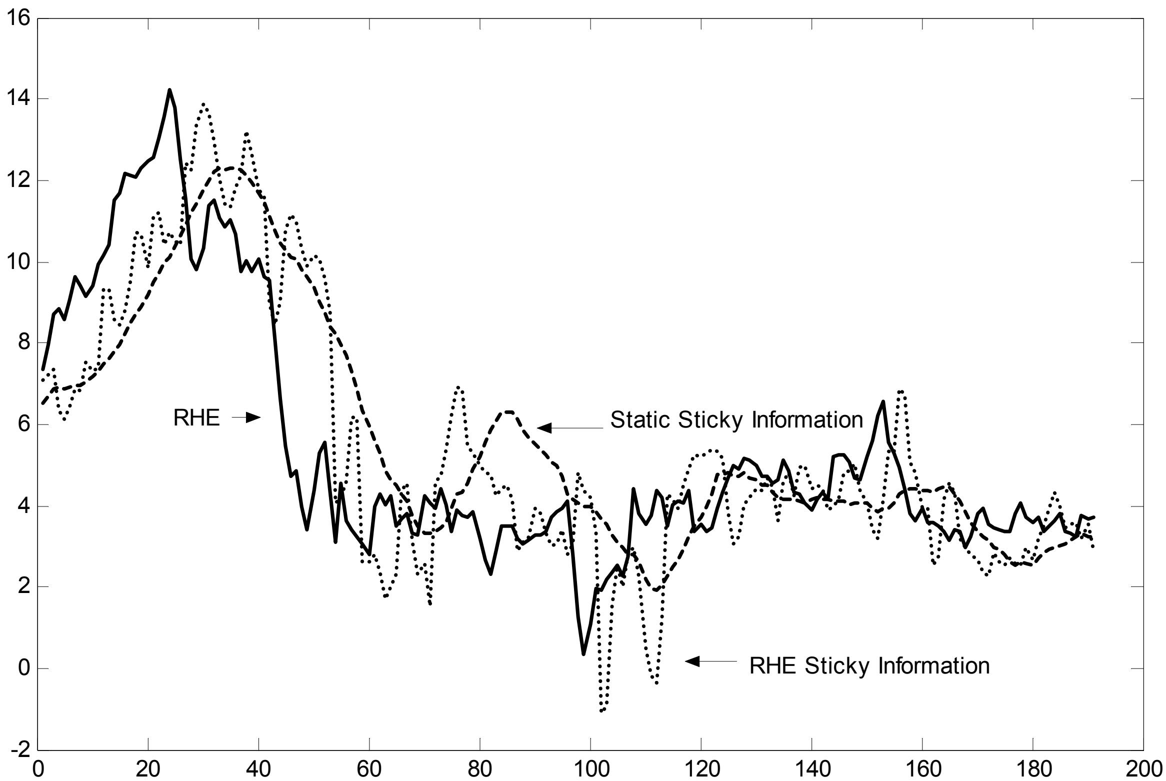

Figure 2. Plot of mean forecasts from various models.

Table 2. Comparison Tests between Survey Data and Models

Period	T_{RHE}	$T1_{RHE}$	T_{static}	$T1_{static}$	T_{sticky}	$T1_{sticky}$
1979.1	17.0305	17.1699	8.2598	8.3666	4.7907	4.8968
1979.2	19.4906	19.6064	7.4275	7.5461	4.6109	4.7121
1979.3	12.106	12.2169	9.9078	10.0397	6.9138	7.0124
1979.4	15.0421	15.167	9.6291	9.7192	8.4984	8.6003
1979.5	15.1087	15.216	12.3315	12.4343	10.9529	11.0591
1979.6	14.534	14.671	11.8848	12.0857	9.8689	9.9672
1979.7	15.8832	16.031	11.2654	11.5359	11.0209	11.1419
1979.8	23.3168	23.4432	11.5422	11.7371	11.0776	11.1746
1979.9	26.8332	26.9904	13.0444	13.212	10.1568	10.2525
1979.1	27.5622	27.7356	16.1276	16.3149	7.8478	7.9541
1979.11	30.4326	30.5609	14.5018	14.7027	8.0026	8.0925
1979.12	12.0512	12.2179	17.6508	17.8293	8.9829	9.0923
1980.1	9.0179	9.1377	15.2256	15.3569	13.0823	13.1854
1980.2	10.4772	10.6593	23.5388	23.8476	17.5971	17.7349
1980.3	12.0747	12.2942	24.058	24.3829	26.4844	26.6943
1980.4	10.7613	11.0399	20.7186	21.1666	27.7169	27.8984
1980.5	17.6441	17.8195	27.0396	27.1927	35.3574	35.4716
1980.6	19.1935	19.3529	26.2669	26.5073	28.4363	28.5876
1980.7	13.3359	13.443	25.862	26.1176	17.018	17.1196
1980.8	11.5546	11.7071	22.5579	22.8008	14.7498	14.917
1980.9	15.5485	15.7154	22.4435	22.7178	14.7423	14.8665
1980.1	10.18	10.3499	24.626	24.806	18.37	18.4938
1980.11	8.8232	9.0102	25.1987	25.3424	19.9147	20.0608
1980.12	5.0894	5.2278	19.6853	19.8205	21.2911	21.3926
1981.1	11.2183	11.3764	24.3005	24.4214	26.3471	26.4863
1981.2	11.5222	11.6307	17.1982	17.3171	19.1534	19.2502
1981.3	9.7333	9.8731	21.4443	21.5371	18.9642	19.0872
1981.4	4.6875	4.9372	17.5319	17.8779	7.1303	7.3141
1981.5	5.873	6.0199	9.1211	9.2433	7.2256	7.3269
1981.6	3.2464	3.4046	14.0716	14.2045	9.7455	9.8666
1981.7	6.0579	6.1829	16.2935	16.3735	19.4017	19.5202
1981.8	13.1149	13.277	13.8964	14.0245	17.5818	17.7203
1981.9	12.4837	12.6439	21.2932	21.4349	17.7757	17.9011
1981.1	13.707	13.8725	14.4468	14.6089	13.2229	13.3309
1981.11	16.5333	16.7322	22.2799	22.4252	16.7655	16.9049
1981.12	10.7386	10.874	19.0246	19.2311	21.017	21.1136
1982.1	10.8331	11.0285	14.9733	15.1661	16.8875	17.0367
1982.2	5.8288	6.0222	23.6378	23.9743	25.4094	25.5467
1982.3	8.2214	8.4171	12.5815	12.8615	18.617	18.7369
1982.4	9.1854	9.4258	21.7817	22.028	7.0939	7.2453
1982.5	6.1456	6.2432	7.4413	7.7102	8.7123	8.8075
1982.6	7.6277	7.7351	6.8249	7.0874	4.6196	4.7001
1982.7	7.1692	7.3103	9.0756	9.3997	7.3943	7.4976
1982.8	9.7981	9.9453	5.3009	5.716	5.4967	5.5892
1982.9	14.9797	15.1494	12.2449	12.5382	15.3088	15.415
1982.1	11.4699	11.6176	6.7563	6.9289	14.8782	14.9871
1982.11	6.4667	6.6604	5.6603	6.1018	13.0441	13.2119
1982.12	3.9079	4.0115	3.1044	3.2894	7.1316	7.2407

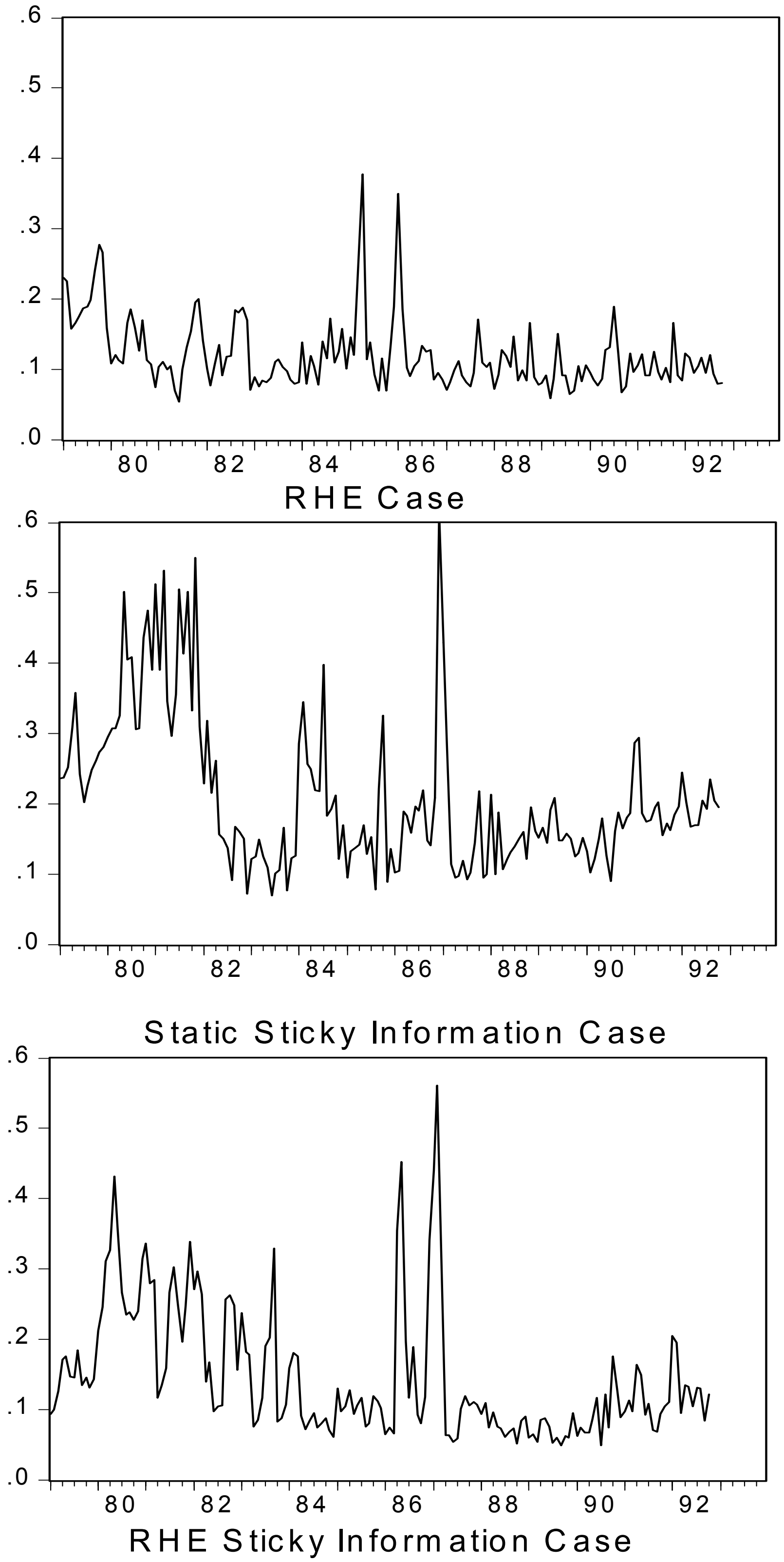

Figure 3. Kullback-Leibler Distance Measures for various models.

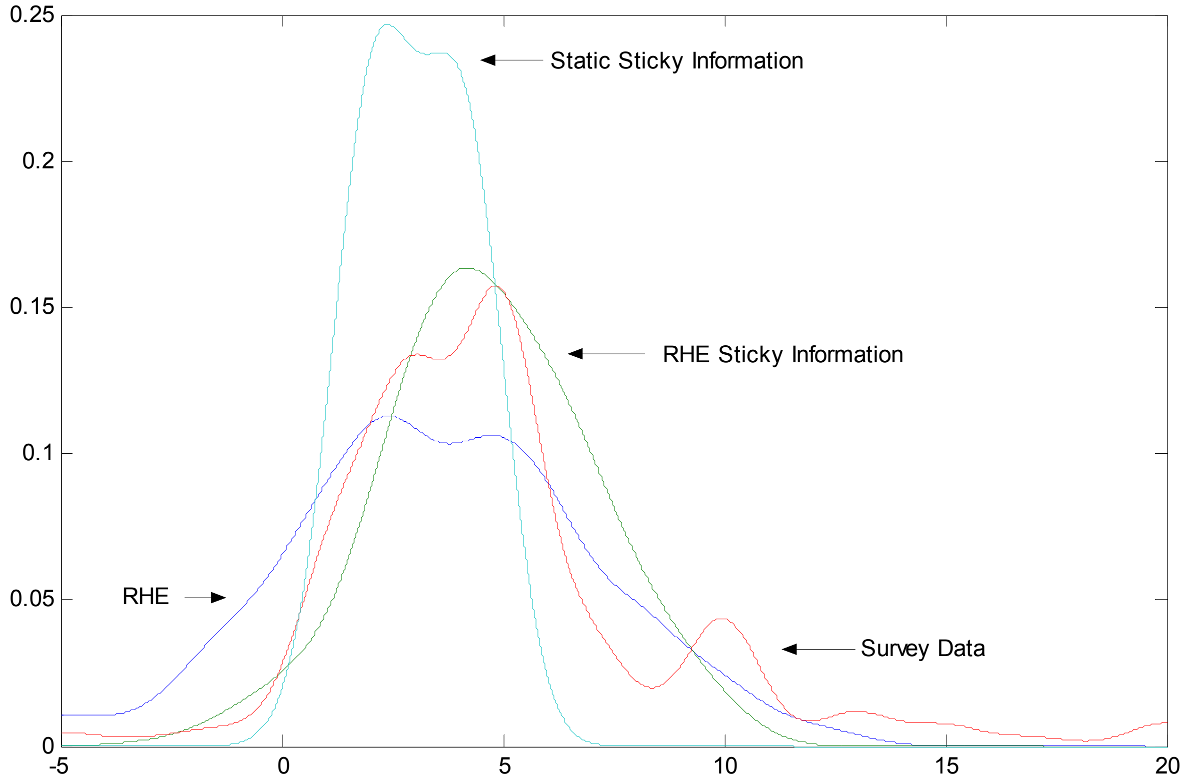

Figure 4. Comparison of Histograms for Period 1980.5, where RHE Sticky Information model provides best fit.

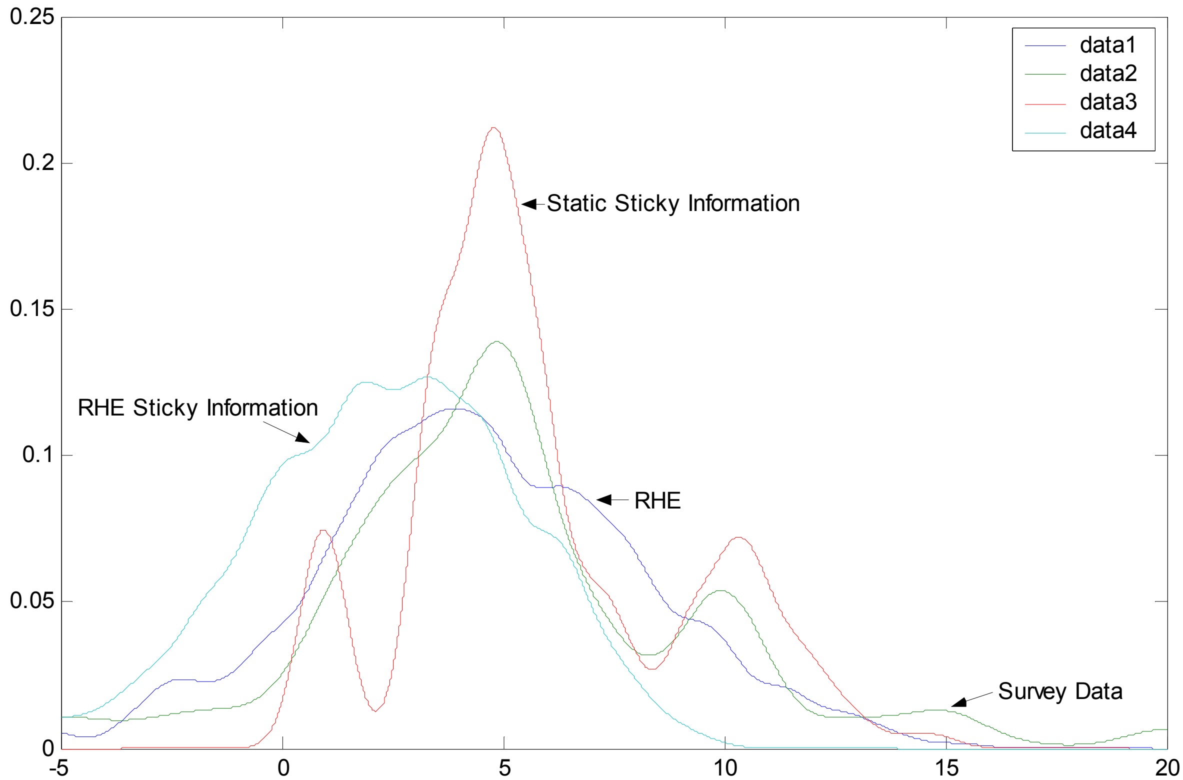

Figure 5. Comparison of Histograms for Period 1982.12, where Static Sticky Information model provides best fit.

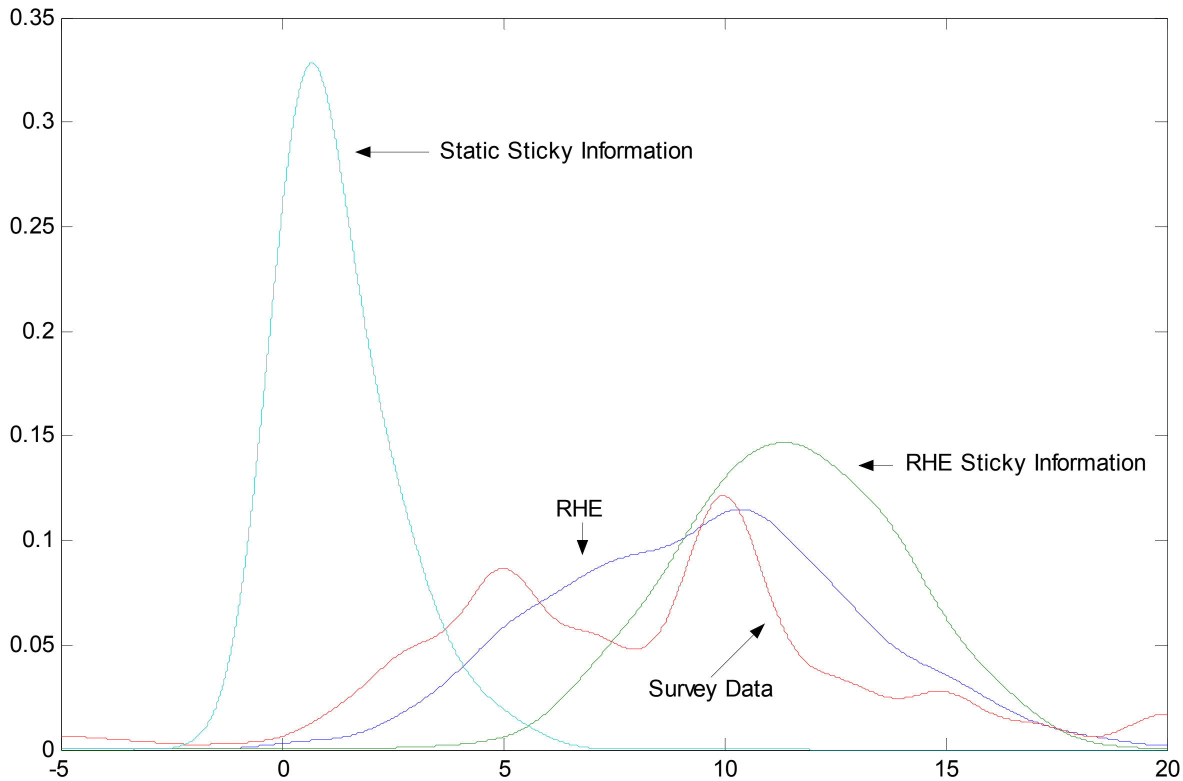

Figure 6. Comparison of Histograms for Period 1981.4, where RHE model provides best fit.

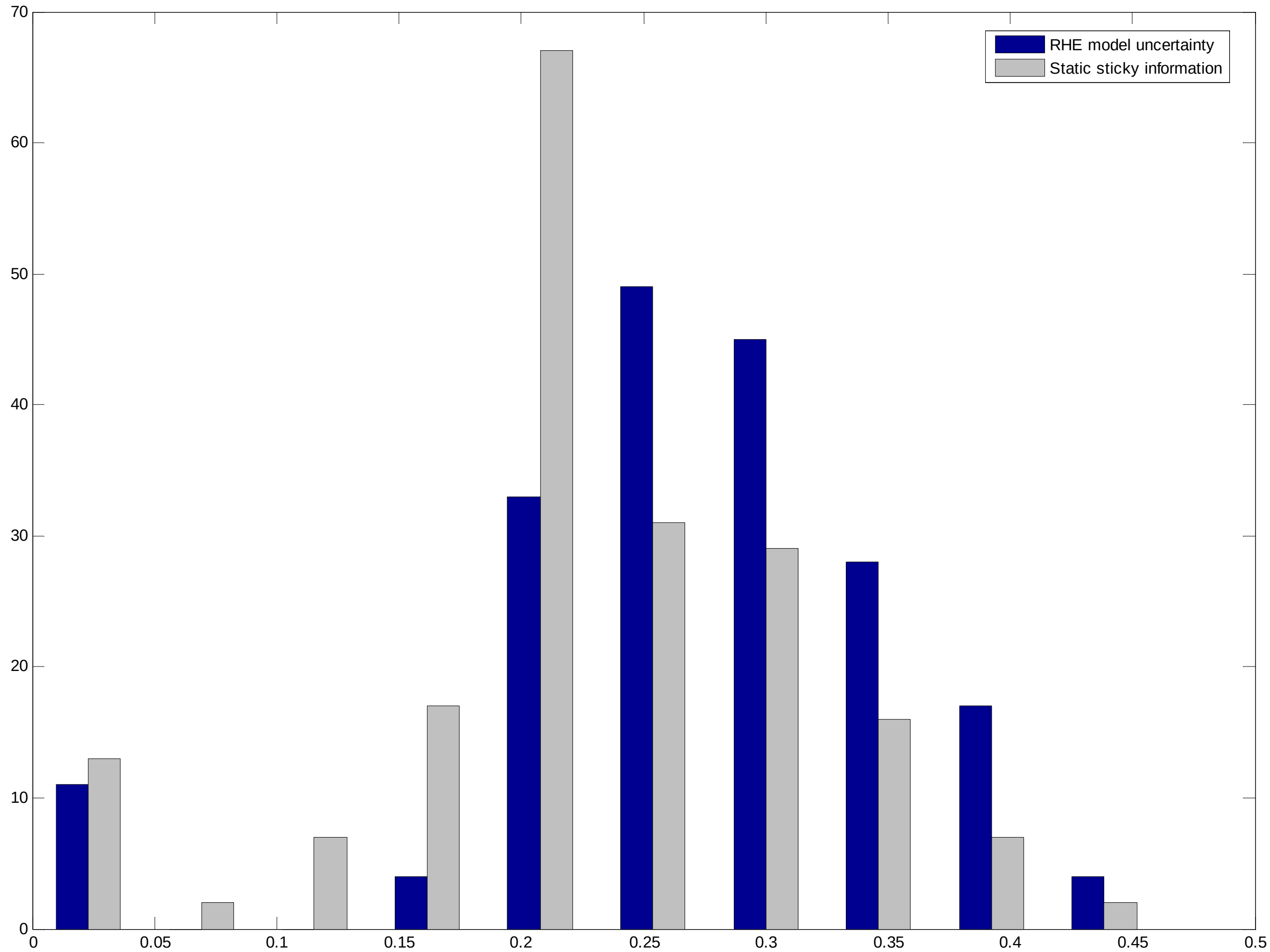

Figure 7. Comparison of Coverage: percent of price expectations lying in the 95% confidence intervals of RHE and static sticky information models.

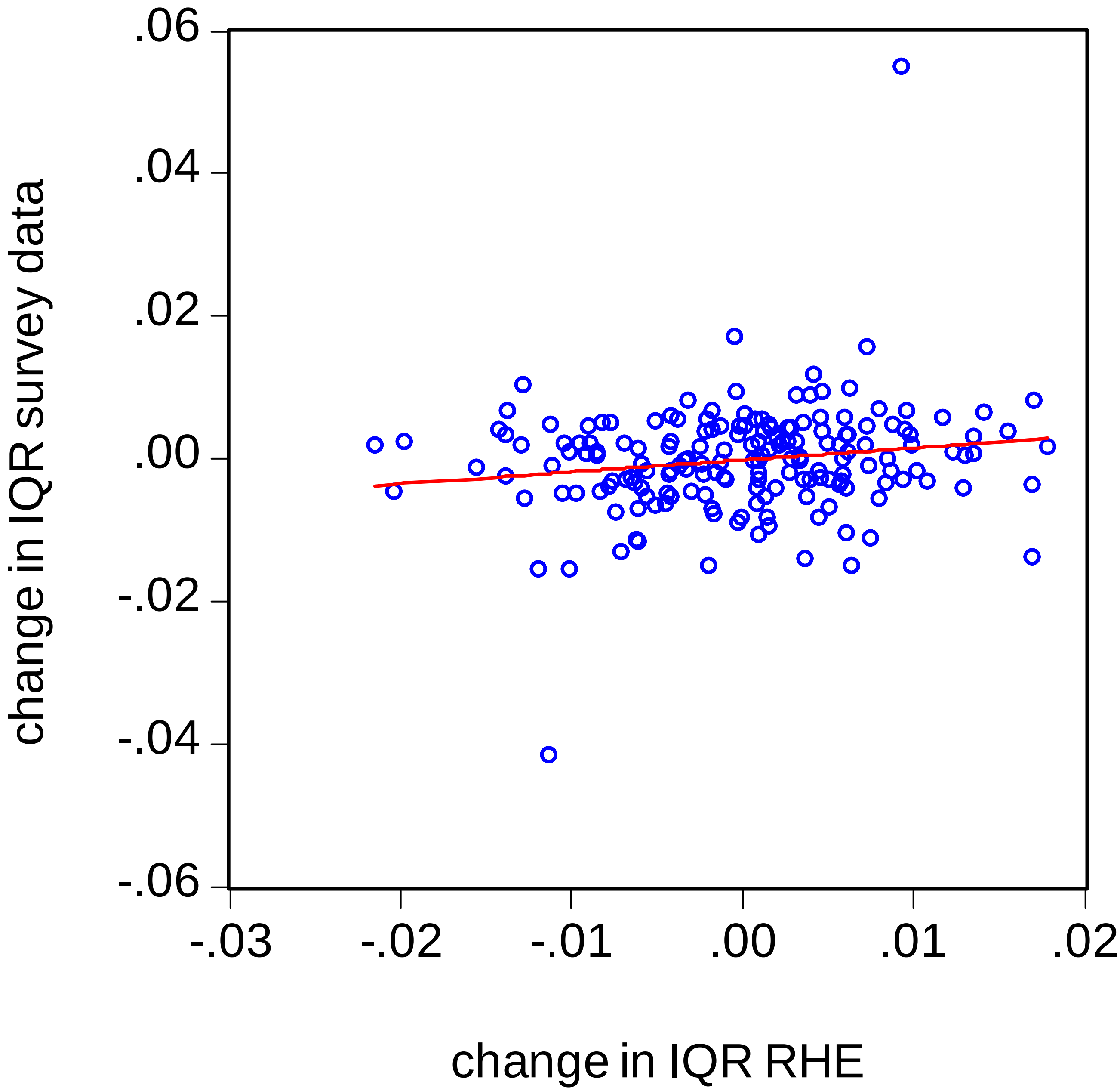

Figure 8. Scatter plot of change in Interquartile Ranges for RHE model uncertainty approach against the Michigan survey data, for each survey period. Trend line, with slope .167, also plotted.

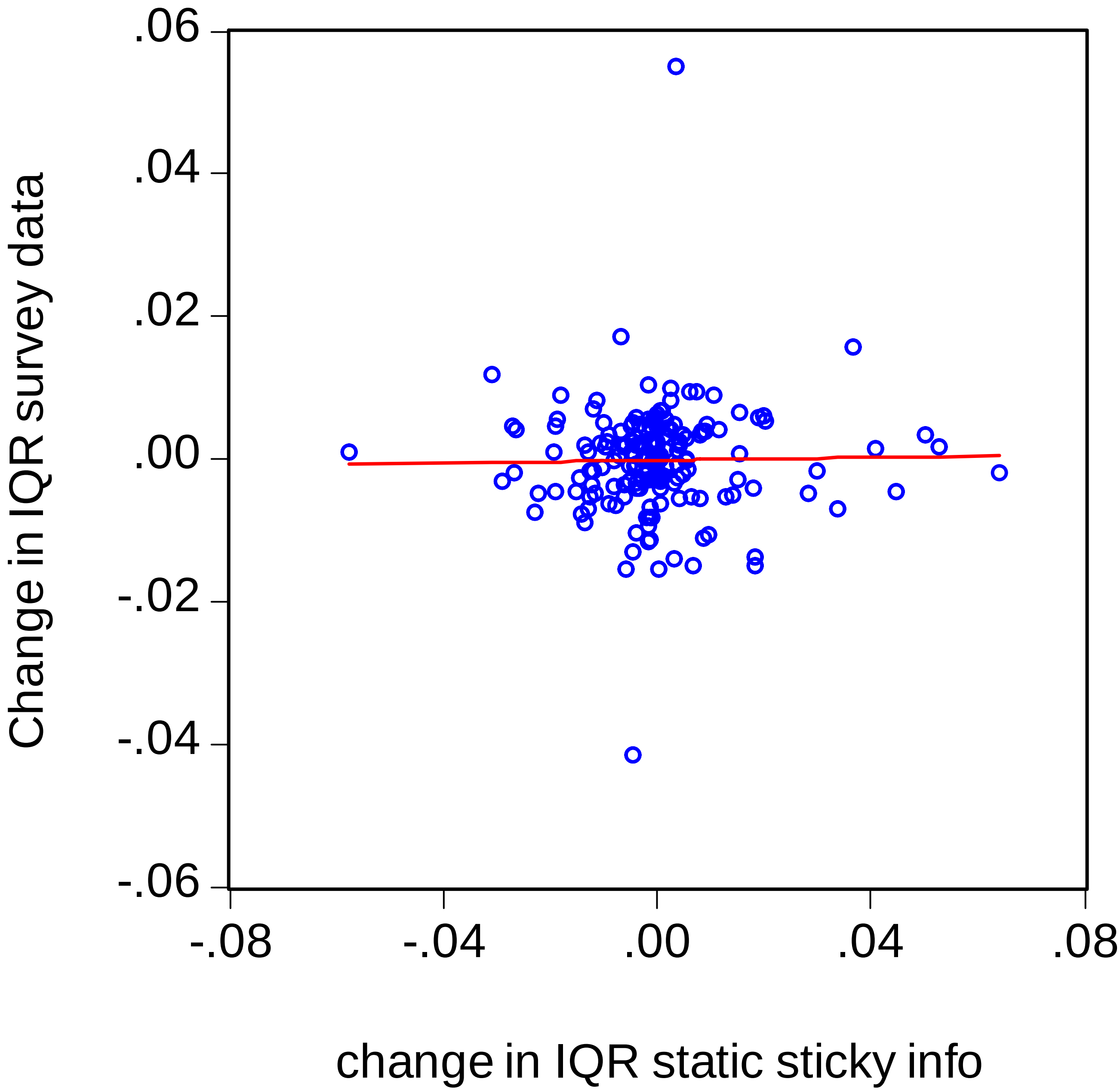

Figure 9. Scatter plot of changes in Interquartile Ranges for static sticky information approach against the Michigan survey data, for each survey period. Trend line, with slope .0099, also plotted.

	Survey Responses		
Period	3	5	10
1979.1-1993.12	17.89%	23.46%	14.69%
1979.5	2.6%	17.40%	38.73%
1982.7	0%	3.7%	13%
1983.3	23%	22%	11.7%

Table 3. Examples of time-varying "digit preferencing". Each value is percentage of survey sample reporting a given value in a given period.

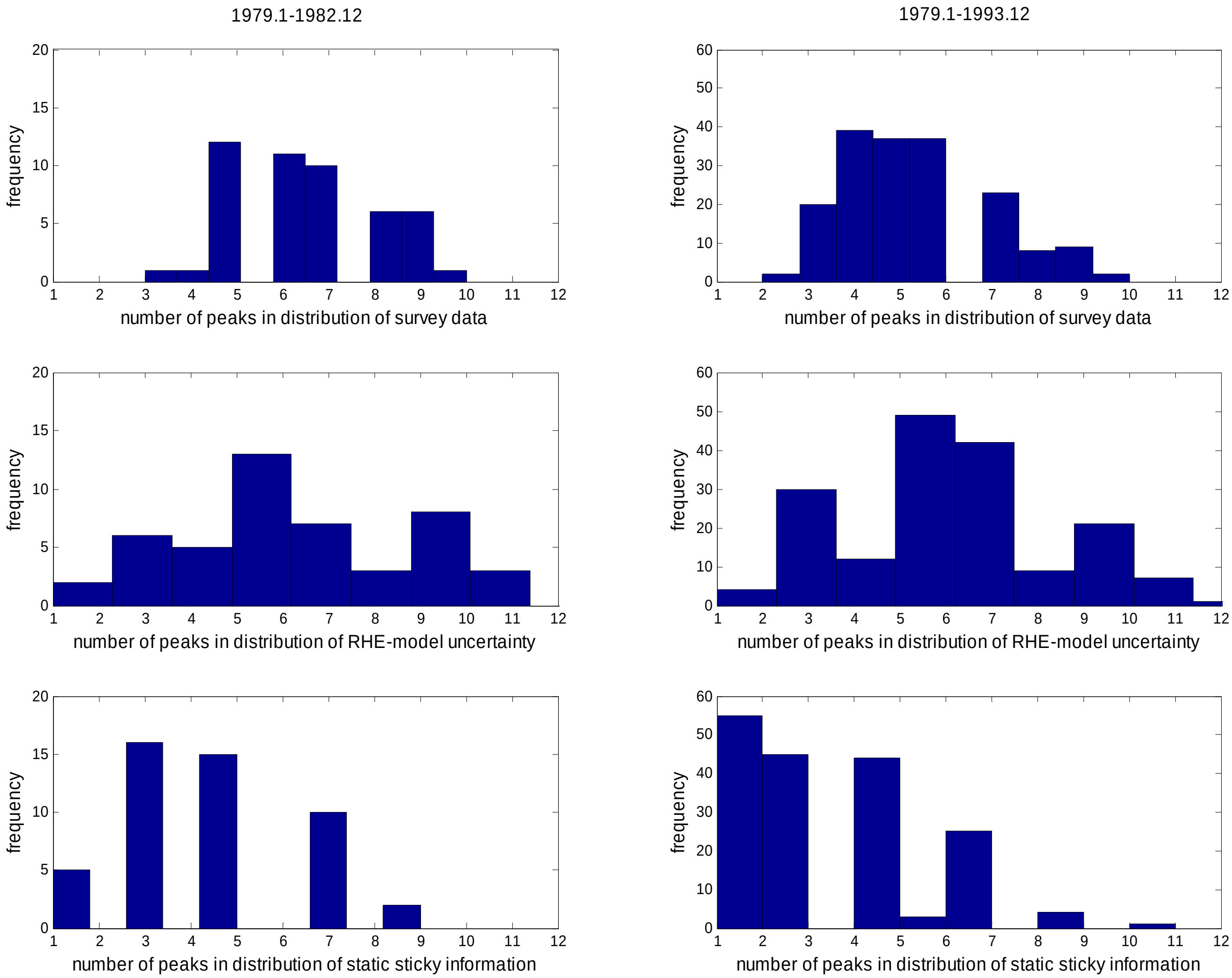

Figure 10. Plots the number of peaks in the distributions of survey data, density functions of RHE-model uncertainty and static sticky information. The left column of histograms plots the number of peaks in a given time-period's distribution over the period 1979.1-1982.12. The right column plots the number of peaks in a given time period's distribution over the period 1979.1-1993.12.