



ELSEVIER

Available online at www.sciencedirect.com



ScienceDirect

Journal of Economic Dynamics & Control 31 (2007) 245–276

JOURNAL OF
Economic
Dynamics
& Control

www.elsevier.com/locate/jedc

Sticky information and model uncertainty in survey data on inflation expectations

William A. Branch*

Department of Economics, 3151 Social Science Plaza, Irvine, CA 92697, USA

Received 7 July 2005; accepted 11 November 2005

Available online 15 March 2006

Abstract

This paper compares three reduced-form 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. Sticky information versus sticky prices: a proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics* 117(4), 1295–1328]. 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. The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *Economic Journal* 114, 497]. Non-parametric evidence suggests that models which allow the degree of heterogeneity to change over time provide a better fit of the data.

© 2006 Elsevier B.V. All rights reserved.

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

Keywords: Heterogeneous expectations; Adaptive learning; Model uncertainty; Survey expectations

*Tel.: +1 949 824 4221; fax: +1 949 824 2182.

E-mail address: wbranch@uci.edu.

1. Introduction

Despite the prominence of rational expectations in macroeconomics there is considerable interest in its limitations. Recent approaches impose bounded rationality at the primitive level; see, for example, [Mankiw and Reis \(2002\)](#), [Ball et al. \(2005\)](#), [Reis \(2004\)](#), [Branch et al. \(2004\)](#) 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.

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, compare how well three competing reduced-form models of expectation formation fit the survey data. [Carroll \(2003\)](#) and [Mankiw et al. \(2003\)](#) provide indirect evidence of limited information flows in expectation formation. This paper 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.

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. In an innovative paper, [Mankiw et al. \(2003\)](#) 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. [Bryan and Venkatu \(2001a, b\)](#) 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 may be an aspect of expectation formation.¹

¹Other papers which show heterogeneity across forecasting models include [Baak \(1999\)](#), [Chavas \(2000\)](#), and [Aadland \(2004\)](#). Experimental evidence is provided by [Heemeijer et al. \(2004\)](#) and [Hommes et al. \(2005\)](#).

This paper extends Branch (2004) by enlarging the set of competing reduced form models. Our approach, like Mankiw et al. (2003), tests for a particular form of sticky information flows in agents' survey expectations. We also extend Mankiw et al. (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. Reis (2004), in a microfounded model of inattentiveness, shows that with costly updating of information firms may *choose* to be inattentive, and under certain conditions the distribution of information may be in line with Mankiw–Reis. Motivated by Reis' purposeful model of inattention, 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 reduced-form 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 compare these models with the approach presented in Branch (2004). In Branch (2004) agents make a discrete choice between (non-nested) 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 reduced-form approaches are statistically identical to the distribution of the survey data. We show, however, 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 how well each model fits the entire period by period distribution of survey expectations. Since it is not reasonable to expect that the simple empirical models, motivated by theory, will provide an exact match to the data, 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, with a time-varying distribution of agents across models, capture the time-variation reasonably well. 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.

Our results suggest that a structural approach to model uncertainty and inattentiveness along the lines of Reis (2004) may further enhance our understanding of the process of expectation formation. It is important to note, then, that our results are specific to classes of expectation formation models. Our approach prevents us from making more general statements about *all* classes of models or making broader structural claims about model uncertainty or inattentiveness.

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

Although the theoretical models of expectation formation do not provide a perfect fit, we argue that they capture important characteristics of the survey data. The results here provide the first empirical comparison of heterogeneous expectations models.

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

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 \(2004\)](#).⁴

2.1. Survey expectations

This paper characterizes the 12-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, because it covers a diverse spectrum of inflation volatility (e.g. great inflation and great moderation) and to keep the comparison sharp with our earlier work.⁵ This paper characterizes expectations of future inflation⁶ and the sample consists of 93 142 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%. Excluding these responses does not change the qualitative results. [Mankiw et al. \(2003\)](#) extend the sample to prior periods by inferring a distribution from the survey sample.

³Three 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.

⁴We use the terms sticky information and model uncertainty as descriptors of the reduced-form models. The underlying micro behavior which leads to these empirical models is not implied by these terms.

⁵There are some missing months in 1977, and so the sample is restricted only to continuous periods.

⁶The two relevant questions are:

1. During the next 12 months, do you think that prices in general will go up, down, or stay where they are now?
2. By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

In the Michigan survey each agent i at each date t reports their 12-month ahead forecast $\tilde{\pi}_{i,t}^e$. We assume that agents estimate a VAR of the form,

$$y_t = A_{1,t-1}y_{t-1} + \cdots + A_{p,t-1}y_{t-p} + \varepsilon_t,$$

which has the VAR(1) form,

$$z_t = A_{t-1}z_{t-1} + \tilde{\varepsilon}_t, \quad (1)$$

where $z_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$ 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.

Notice that in (1) we allow agents to estimate a VAR with time-varying parameters. This is in accordance with the adaptive learning literature where agents, uncertain about the true value of A , estimate A_{t-1} in real-time. Then (1) could also be a particular theory of model uncertainty. Below, we are interested in how survey expectations are distributed across several variants of this model. 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 12-month ahead forecasts, we assume agents look at past monthly inflation to forecast 12-month ahead inflation. In other words, the VAR in (1) consists of monthly data. This assumption is also made in Mankiw et al. (2003).⁷

2.2. Static sticky information

In a series of papers, Mankiw and Reis (2002), Ball et al. (2005), and Mankiw et al. (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. 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.⁸ Full information in this setting is not complete information since (1) is a reduced-form VAR that could represent model uncertainty (because of the time-varying parameters). In order to form this forecast the agent must generate a series

⁷The Labor Department in its press release reports first the monthly inflation figure. Thus, it is not unreasonable to assume it is an input to agents' forecasting model.

⁸Of course, because we have not posed a model for the economy these expectations may not be rational. Instead, at best these are the optimal linear forecasts given beliefs in (1). In the terminology of Evans and Honkapohja (2001) these are restricted perceptions.

of i -step ahead forecasts of monthly inflation,

$$\pi_{j,t+i}^e = E(\pi_{t+i} | I_{t-j}) = (A_{t-j}^{j+i} z_{t-j})_{\pi},$$

where $(A_{t-j}^{j+i} z)_{\pi}$ is the inflation component of the projection, with π denoting monthly inflation.⁹ The agent then continues out-of-sample forecasting in order to generate the sequence $\pi_{j,t+1}^e, \pi_{j,t+2}^e, \dots, \pi_{j,t+12}^e$. The 12-month ahead forecast of annual inflation is generated according to,

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

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 $\hat{\pi}_{0,t+12}$, $\lambda(1 - \lambda)$ with $\hat{\pi}_{1,t+12}$, $\lambda(1 - \lambda)^2$ with $\hat{\pi}_{2,t+12}$, and so on. The mean forecast is,

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

Reis (2004) shows that, in a model where firms face costly updating of information, it is possible that aggregate inattentiveness could also yield this reduced-form. Below, in our empirical approach, we will sample from this sticky information distribution to generate a predicted survey sample and assess this model's ability to explain the entire distribution of Michigan survey responses.¹⁰

2.3. Rationally heterogeneous sticky information

The Mankiw–Reis approach in the previous subsection is a reduced-form heterogeneous expectations model with 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

⁹The implicit assumption in this projection is that agents assume A remains constant over the forecast horizon. An interesting extension would be to consider a setting where agents account for their uncertainty in A when constructing conditional forecasts.

¹⁰In 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.

with a list of forecasting models distinct in the frequency of recursive updating. In each period agents choose their expectations from this list. This approach can be viewed as a reduced-form generalization of Reis (2004). Reis' model is a first principled alternative to Mankiw–Reis in the sense that the choice of updating probabilities is purposefully chosen and (possibly) time-varying. One could envision the reduced-form in this paper as arising from Reis' model extended to include structural change.

There is a burgeoning literature on dynamic predictor selection. A prime example is the adaptively rational equilibrium dynamics (ARED) of Brock and Hommes (1997). In the ARED 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 .¹¹ Moreover, unlike in the previous subsection, we assume that each predictor $\hat{\pi}_{j,t+12}$ is recursive and updated each $(j+1)$ periods.¹² It should be noted that the predictors are updated each $j+1$ periods since $j=0$ was designated above as full-information.¹³ 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 reduced-form model 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

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

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 et al. (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

¹¹Brock et al. (2005) 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.

¹²In 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 $\hat{\pi}_{j,t+12}$.

¹³Below we will change this (unfortunate) notation so that j is descriptive and represents the frequency with which the predictor is updated.

the intuitively appealing assumption that the proportion of agents using a predictor is increasing in its accuracy.¹⁴

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 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.¹⁵ 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 period, how much ‘memory’ they have. This is intractable in the current framework and we leave this issue to future research.

2.4. Model uncertainty

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. We interpret model uncertainty as agents selecting from the 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. Agents are distributed across these predictors according to the MNL (2).

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

¹⁴The 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.

¹⁵Up to the noise in the random utility function.

set of predictors is meant to represent the classes of multivariate and univariate forecasting methods. 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 that are not special cases of the model uncertainty case.

Our use of the term ‘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, as switching between forecast models is meant to proxy for an underlying, deeper sense of model uncertainty. Because the RHE model uncertainty alternative is meant as a representation of more fundamental behavior, we cannot interpret poor empirical fit as an indictment against all forms of model uncertainty.

3. Empirical results

This section presents results of an empirical analysis of the two sticky information models in comparison to the survey data on inflation expectations.

3.1. Predictor estimation

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, we describe a process of recursive forecasting.

First, we describe the VAR which is the basis for forecasting.¹⁶ We follow Branch (2004) and Mankiw et al. (2003) in assuming that the VAR(1) consists of monthly inflation, unemployment, and 3-month t-bill rates. 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 choose a lag length of 12 in order to minimize the Akaike Information Criterion. This VAR is used to generate a 12-month ahead forecast of inflation.

There are several issues to pin down. The model of sticky information presented in this paper is an assumption on the information sets, which evolve stochastically. Each sticky information alternative makes specific assumptions on how agents’ information sets evolve. However, the parameters of the VAR model may be time-varying and unknown by agents. As a result, we assume that when agents update

¹⁶We focus on VAR forecasting because it is often cited as an approximation to rational expectations and is a frequently employed forecasting strategy.

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 et al. (2003) by fixing $\lambda = .1$. To ensure robustness of our results, we also let $.05 \leq \lambda \leq .25$.¹⁷ 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

$$\mathcal{H}_t = \{\hat{\pi}_{1,t+12}, \hat{\pi}_{3,t+12}, \hat{\pi}_{6,t+12}, \hat{\pi}_{9,t+12}\}.$$

That is, the available predictors update information every period, every third period, every sixth period, and once every nine periods. We make these restrictions in order to maximize the number of identifiable predictors. We omit VARs 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 VARs estimated less frequently than every 9 months. VARs 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.¹⁸ It is important to note that this bound does not affect the qualitative results. An advantage to the empirical procedure below is it will identify those agents that update less than once a year as $\hat{\pi}_9$ which still implies 'sticky information' in this context.

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,

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

¹⁷It 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). Empirical exploration suggests that the optimal λ falls in the range $[\.05, .25]$ based on minimizing a Kullback–Leibler distance measure. A smaller value of λ (.05) is preferred if the objective is to minimize the total distance from the actual data and a larger value (.25) if minimizing the mean distance across periods is the goal.

¹⁸This 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.

where C_j is a constant around which the mean predictor proportions vary. Agents are assumed to base decisions on how well 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. 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.

We now describe how the forecasts are actually computed. We assume agents engage in real-time learning by recursively updating their prior parameter estimates.¹⁹ In VARs 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 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

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

where

$$A_t = A_{t-1} + t^{-1}R_t^{-1}z_{t-1}(z'_t - z'_{t-1}A'_{t-1}),$$

$$R_t = R_{t-1} + t^{-1}(z_{t-1}z'_{t-1} - R_{t-1}).$$

Note that R_t is the sample second moment of z_{t-1} .²⁰ Since there are three 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} .

¹⁹Some authors advocate forecasting based on real-time data sets (see Croushore and Stark, 2003). Such an undertaking is beyond the scope of this paper.

²⁰See Evans and Honkapohja (2001) for an overview of real-time learning using recursive least squares.

One (potential) 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.

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 j th period according to,

$$A_{j,t} = \begin{cases} A_{j,t-1} + t^{-1} R_{j,t}^{-1} z_{j,t-1} (z'_t - z'_{j,t-1} A'_{j,t-1}) & \text{every } j\text{th } t, \\ A_{j,t-1} & \text{otherwise.} \end{cases}$$

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 12 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 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 $t > 1976.11$. Implicit in the RHE specification of sticky information is that when agents update information they only acquire the most recent data point. 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. As mentioned above, 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 .

3.2. Discussion of forecasting models

Forecasts based on VARs like (1) are employed extensively in macroeconomics. 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. 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.

It is important to note that the results in this section are conditional on particular classes of reduced-form 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 extend the earlier paper to these two additional models. Our results do not extend to classes of model uncertainty or sticky information not considered here. Because the models are reduced-form it is not possible to draw structural conclusions from the empirical results. A time-varying VAR for example could represent uncertainty about the structural model parameters. Univariate forecasting models, on the other hand, could arise from a model where agents are ‘inattentive’ to other factors, such as output growth.

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, thus ensuring our results are not sensitive to choice of gain sequence (see Branch and Evans, 2005).

3.3. Maximum likelihood estimation of RHE sticky information model

In this subsection we address the paper’s primary objective: 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. To this end, we specify a backwards MNL approach. Each survey response is assumed to be reported as

$$\tilde{\pi}_{i,t}^e = \hat{\pi}_{j,t+12} + v_{i,t}, \quad (4)$$

where $v_{i,t}$ is distributed normally and $j \in \{1, 3, 6, 9\}$. A survey response is a 12-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, $v_{i,t}$, and report an expectation that is their perception of future inflation. The stochastic term $v_{i,t}$ has the interpretation of individual idiosyncratic shocks and is the standard modeling approach in RHE models. These idiosyncratic shocks represent unobserved heterogeneity and are unrelated to our interpretation of model uncertainty or sticky information. In particular, it is consistent with the findings of Bryan and Venkatu (2001a, b) and Souleles (2004) who emphasize the difference between an expectation and a perception. Our model assumes that agents form expectations and report perceptions. In particular, $v_{i,t}$ can account for many of the idiosyncrasies reported in Bryan and Venkatu (2001a, b) as well as difference in market baskets, etc.

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

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

where

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

and $MSE^l = \{MSE_{1,t-k}, MSE_{3,t-k}, MSE_{6,t-k}, MSE_{9,t-k}\}_{k=0}^l$. 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 $\{\tilde{\pi}_{i,t}^e\}_{i,t}$ is

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

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

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.

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

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

Table 1
Maximum likelihood estimation results

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

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 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 9-month predictor carrying the highest cost is not paradoxical. In the ARED 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

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

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. However, the finding of a non-geometric distribution is a primary finding of this paper.

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.

Fig. 1 plots the estimated predictor proportions. These proportions are estimated by simulating (2) and (3) with the parameter estimates in 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 Fig. 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

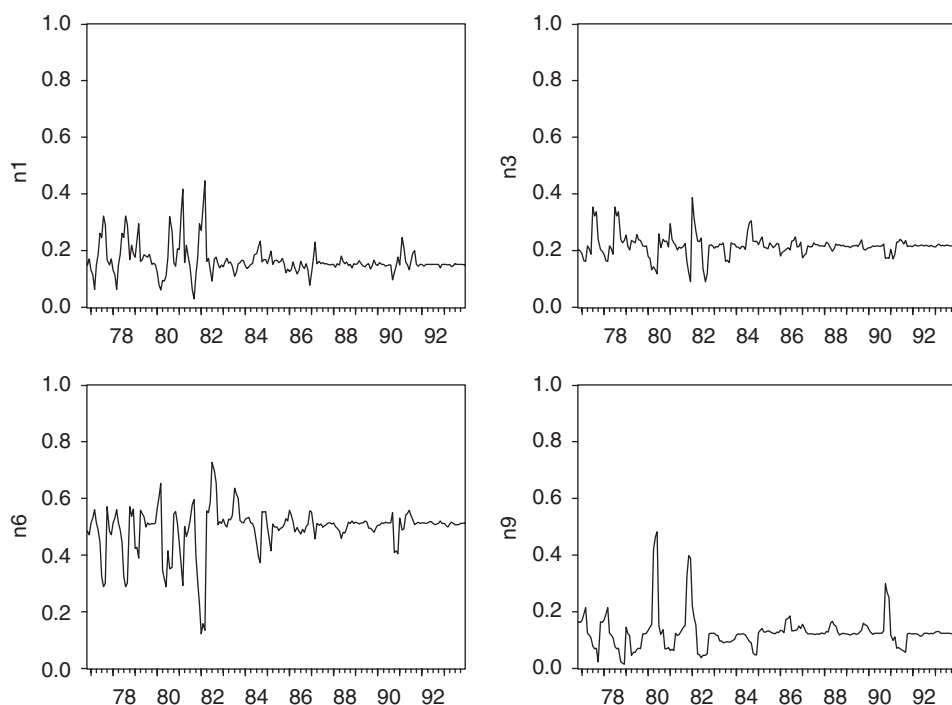


Fig. 1. Estimated proportions.

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). Greater volatility in proportions during periods of high economic volatility is predicted by the model. This is because agents are assumed to respond to the most recent squared forecast error, and it is more likely that predictors with higher average errors will coincidentally produce the lowest error in a given period in such times.²³ 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. It would be interesting for future research to investigate whether these dynamics could be implied by an extension of Reis (2004) to a setting with structural change.

4. Fitting the full distribution

The results above 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.

A complete model comparison should study which reduced-form models of expectation formation track 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 et al. (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

²³One concern may be that the higher volatility, observed at the beginning of the survey sample is because of transient behavior. As detailed above, we used a pre-sample initialization period to eliminate transient dynamics.

dynamic predictor selection mechanism, at its very core, is a theory of time-varying distributions.

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.

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.

To estimate the density functions we make 465 draws from the distribution defined by each model in each time period.²⁴ 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 et al. \(2003\)](#) in fixing $\lambda = .10$. Along these lines, we draw from the same distribution estimated in [Branch \(2004\)](#).

The first question to address is whether these three models of expectation formation have distinct implications about the economy. To this end, [Fig. 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.

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 $\hat{d}, \hat{f}, \hat{g}, \hat{h}$ and hypothesis tests. We are interested in the

²⁴A sample size of 465 is approximately the mean size each period of the Michigan survey.

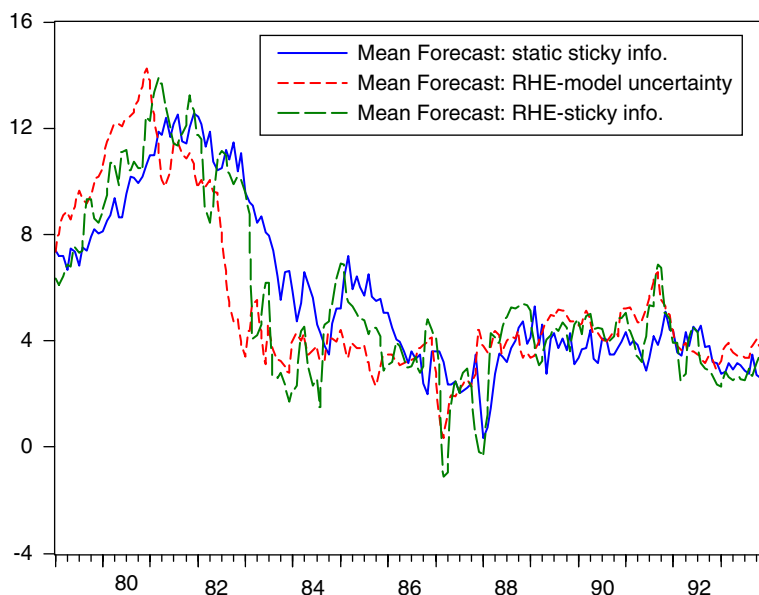


Fig. 2. Plot of mean forecasts from various models.

following hypotheses,

$$H_0 : \hat{d}(\pi^e) = \hat{h}(\pi^e),$$

$$H_0 : \hat{f}(\pi^e) = \hat{h}(\pi^e),$$

$$H_0 : \hat{g}(\pi^e) = \hat{h}(\pi^e).$$

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 and 1982.12. We report results for this period because it is emphasized in Mankiw et al. (2003) who hypothesize that sticky information should be most evident during the great disinflation. In each case, we 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.

That none of our alternative expectation formation models match up statistically with the survey data is not surprising. We now turn to other measures of fit besides hypothesis testing. 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

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

²⁶Details are in the Appendix.

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

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. Using the Kullback–Leibler preferred model is an appealing model selection criteria because there is a complete axiomatic foundation supporting it (see Shore and Johnson, 1980; Csiszar, 1991).

Fig. 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. Fig. 3 demonstrates that which model is the ‘closest’ to the survey data is time-varying and period specific. On average, the two RHE models fit best but not in every period. The RHE model uncertainty has the lowest average distance but this is less evident during the great moderation (post 1984). The RHE sticky information provides a better average fit than the static sticky information model, in line with the maximum likelihood results of Section 3. We conclude that the two RHE cases provide a closer fit than the static sticky information case.

Mankiw et al. (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 Fig. 4 plots the estimated densities and survey histogram for particular times during 1979.1–1982.12. Periods 1985.1, 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. The top panel of Fig. 4 is the case where RHE sticky information dominates, the middle panel is the case where the Mankiw–Reis approach provides the best fit, and the bottom panel is one period where the RHE of Branch (2004) is the closest to actual data. In the top panel there is a double peaked shape to the static sticky information model as found in Mankiw et al. (2003) while 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. These plots, 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.²⁷

We turn to an examination of how well the confidence intervals ‘cover’ the Michigan survey data. We first construct 95% confidence intervals for the empirical distributions of the RHE model uncertainty, RHE sticky information, and static sticky information approaches. 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 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 each reduced-form model. 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

²⁷Fry and Harris (2005) present an alternative theory for spikes based on ‘digit preferencing.’

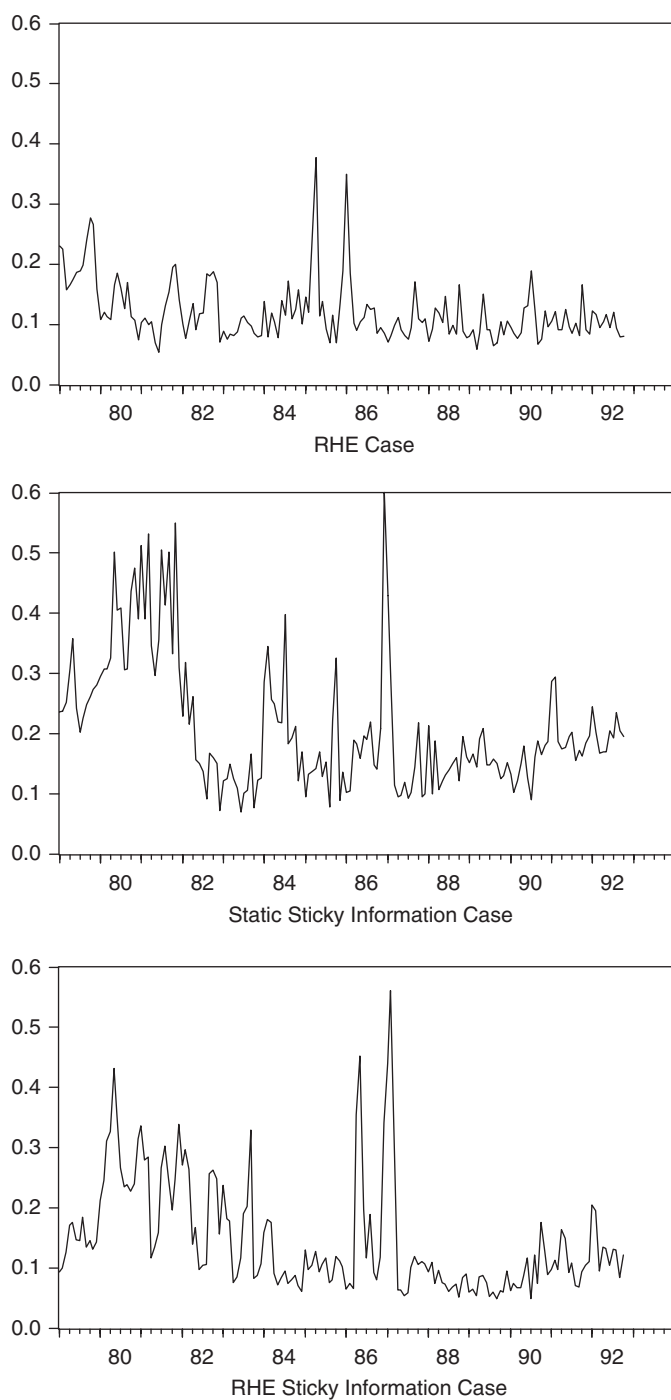


Fig. 3. Kullback–Leibler distance measures for various models.

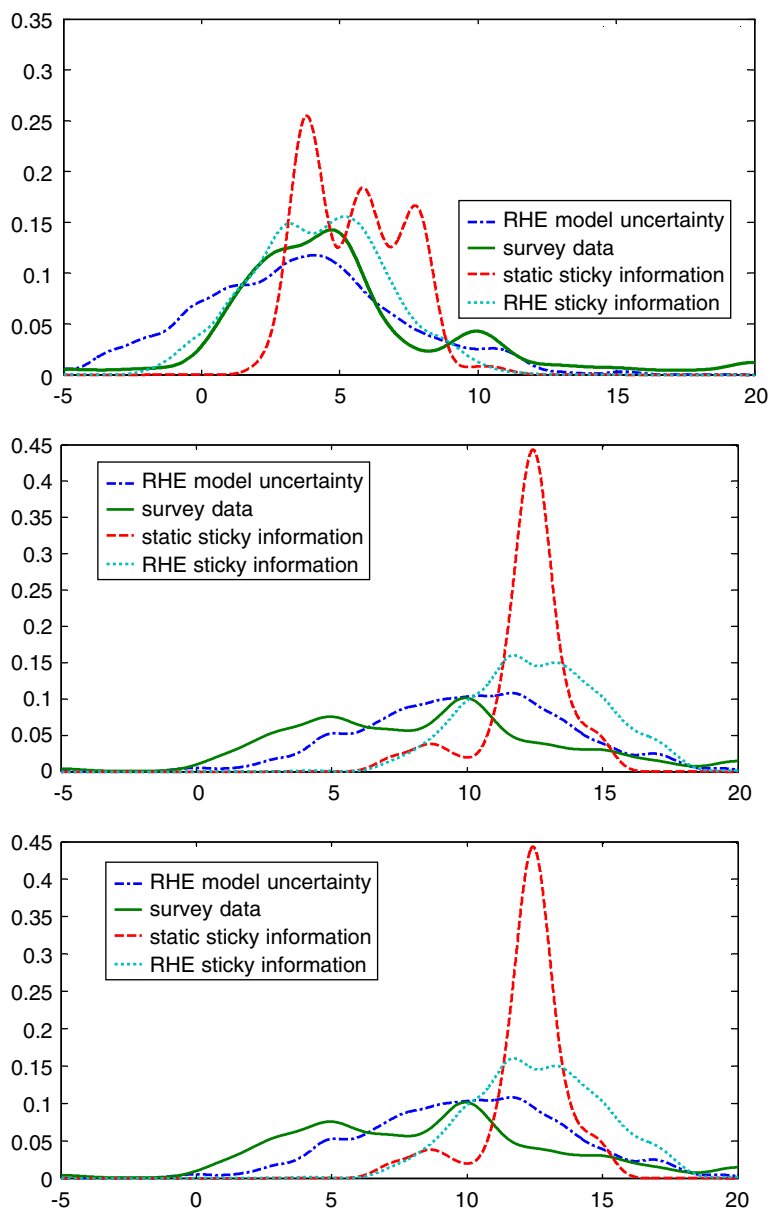


Fig. 4. Comparisons of histograms. Top panel (1985.1) is where RHE sticky information provides best fit. Middle panel (1982.12) is where static sticky information provides best fit. Bottom panel (1981.4) is where RHE model uncertainty fits best.

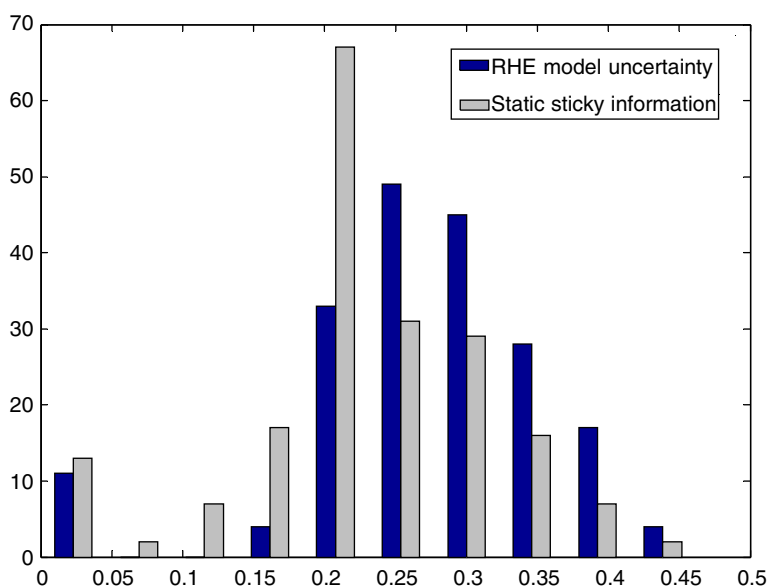


Fig. 5. Comparison of coverage: percent of price expectations lying in the 95% confidence intervals of RHE and static sticky information models.

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 $-5, \dots, 0, \dots, 20$ then there are 26 cases considered and we calculate the proportion of these cases which lie in the various confidence intervals. Fig. 5 reports the results.²⁸

Fig. 5 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. Fig. 5 instead illustrates the percentage of the survey data's histogram, in each period, that lies within the confidence interval of the empirical distributions.

Fig. 3 suggests that the relative fit was period specific, with the two reduced-form sticky information models producing poor fit prior to the great moderation. Fig. 6 makes the coverage comparisons for two subsamples: the pre great moderation (1977.11–1984.12, top panels) and the great moderation (1985.1–1993.12, bottom

²⁸We focus on these two models to economize space. As suggested by Fig. 3, and verified by our own explorations, the two RHE models produce similar qualitative results relative to the static sticky information model.

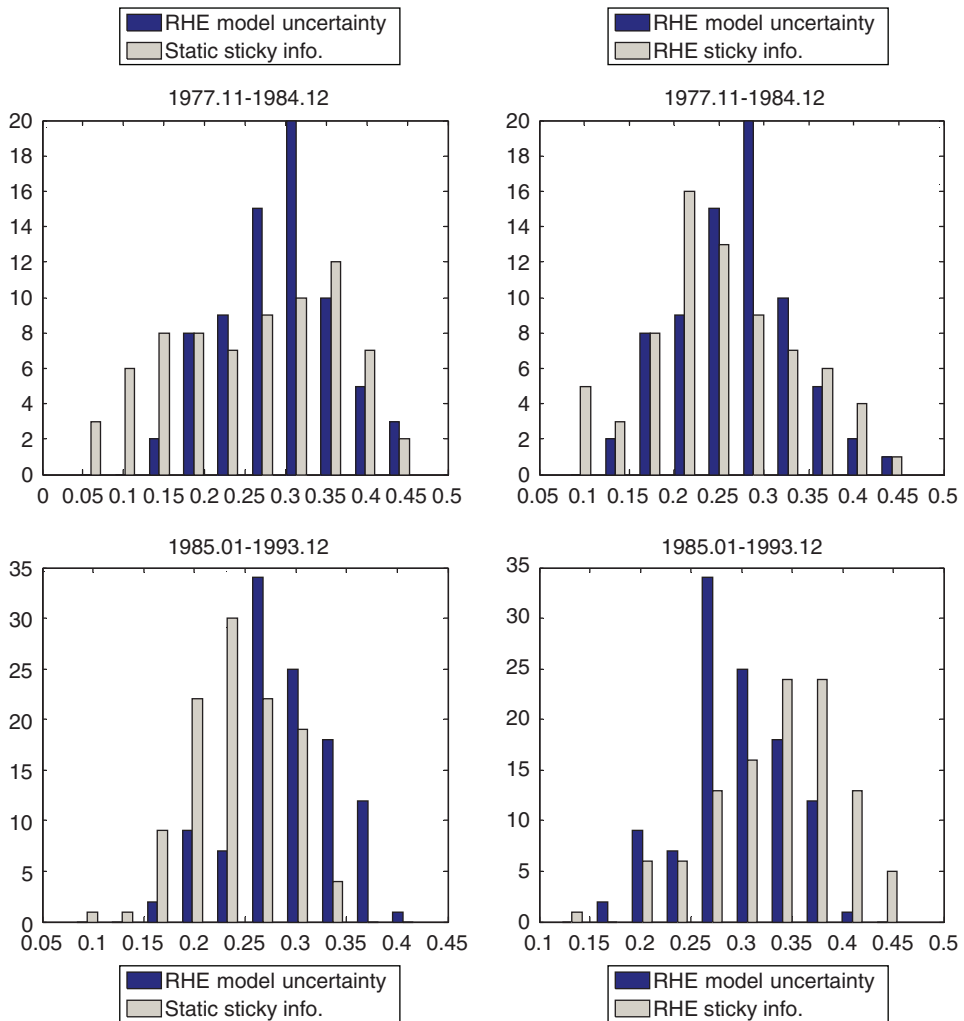


Fig. 6. Comparison of coverage. Left panels for the pre-Great Moderation period, right panels for the Great Moderation. Upper panels compare RHE model uncertainty and static sticky information. Lower panels compare RHE model uncertainty and RHE sticky information.

panels). The left panels compare RHE model uncertainty to static sticky information and the right graphs compare the two RHE models. As a metric of fit we say that a model provides better coverage if the number of periods in which the coverage percentage is greater than .25 is higher than the alternative. Using this measure, Fig. 6 suggests that the two models with time-varying distribution have greater coverage than the static sticky information alternative. This is particularly true during the great moderation. Fig. 6 also suggests the RHE sticky information has greater coverage than the RHE model uncertainty during the great moderation.

These findings suggest an interpretation along the following lines: during periods of economic volatility, like the 1970s, the RHE model uncertainty alternative fits best; during the great moderation of the 1980s and 1990s the RHE sticky information model provides a closer fit.

Ultimately, this paper argues, through parametric and non-parametric evidence, in favor of those models that are able to capture the evolution across time of the distribution of survey data. Fig. 7, top panel, provides further evidence on how well the RHE models capture the time-varying dispersion of the survey responses. Fig. 7 reports 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. The top panel of Fig. 7 scatters changes in the IQR for the RHE model against changes in the IQR for the survey data, and the bottom graph does similarly for the static sticky information model.²⁹ Each figure also plots the trend line. The slope of the trend in the top and bottom panels are .167 and .0099, respectively.

In Fig. 7, top panel, there is an upward trend suggesting that as the survey data becomes more disperse over time, then the RHE model uncertainty approach predicts a greater dispersion of the survey data as well. In the bottom panel, the trend line is increasing slightly, but the slope is not statistically different from zero. 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 IQRs for the RHE and static sticky information approach 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.

5. Conclusion

This paper examined three reduced-form models of heterogeneous expectations. This paper achieved two objectives: first, a characterization of sticky information in survey data; second, an examination of how well each model fits the full distribution of survey responses. We compared two models of sticky information against the RHE 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 (1997) where we assume agents make a

²⁹Plotting the level of IQR produces similar results as Fig. 7. Also, the plot for changes in IQR for the RHE sticky information model produces a similar, though less steep, trend as the RHE model uncertainty case.

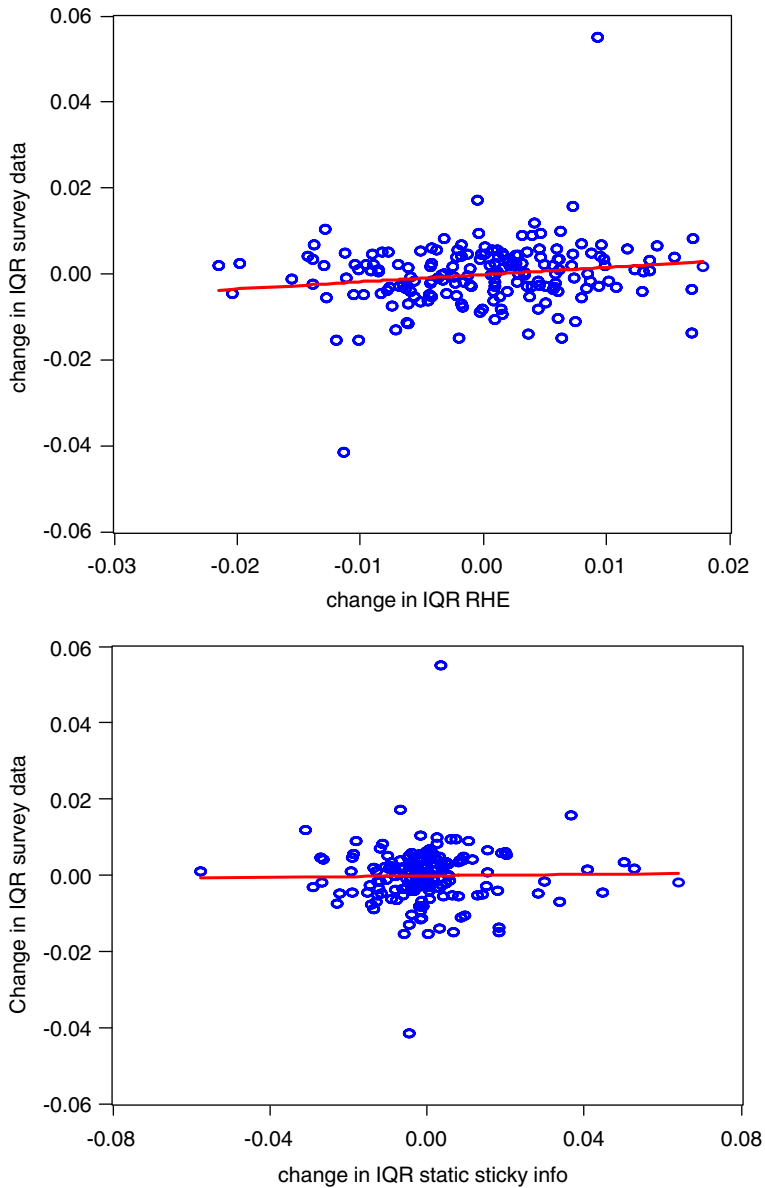


Fig. 7. Scatter plots of change in interquartile ranges for each period, top graph is RHE model uncertainty against the Michigan survey data, bottom graph is for static sticky information. Trend lines (.167 and .0099, respectively) also plotted.

discrete choice between recursive forecasting functions which differ by the frequency with which they are updated. This is a reduced-form approach ‘in the spirit’ of Reis (2004).

We first characterized limited information flows in the survey data by restricting agents to a particular class of sticky information models. In the paper's main results, 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 3–6 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 measures of fit. Non-parametric evidence suggests that the two rationally heterogeneous expectations models provide a better fit than the static sticky information model. We construct estimates of the density functions for each model and compare them to the histogram of the actual survey data. However, this result holds, on average, and there are periods, particularly during the late 1980s and early 1990s, in which static sticky information provides the best fit.

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 may be elements of the data. Most significantly, the distribution of heterogeneity is non-geometric and time-varying. The models presented here, though, do not allow for sticky information across competing models of the economy. Future research should address structural explanations of these findings.

Acknowledgments

This paper has benefited enormously from comments and suggestions by the Editor, Peter Ireland, two anonymous referees, Bruce McGough, Fabio Milani, and Ken Small.

Appendix A. Log-likelihood function derivation

Recall, that the actual observed survey response is given by

$$\tilde{\pi}_{i,t}^e = \hat{\pi}_{j,t+12} + v_{i,t}, \quad (6)$$

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 j th predictor was given by the theoretical model as a MNL,

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

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

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

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

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

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

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

Taking logs leads to form (5).

Appendix B. 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 [Figs. 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.

Appendix C. 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:

$$H_0 : d(x) = h(x),$$

$$H_0 : f(x) = h(x),$$

$$H_0 : g(x) = h(x).$$

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

$$T = nh^5 \frac{(\tilde{I} - c_2(n))}{\hat{\sigma}},$$

$$T_1 = nh^5 \frac{\tilde{I}_1}{\hat{\sigma}},$$

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 $\omega_i = ((x_j - x_i)/h)((y_j - y_i)/h)$.³⁰

Pagan and Ullah note that T, T_1 are distributed $N(0, 1)$ if $h \rightarrow 0$ and $hn \rightarrow \infty$. 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

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

³⁰See Pagan and Ullah (1999) for a justification.

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.

References

- Aadland, D., 2004. Cattle cycles, heterogeneous expectations, and the age distribution of capital. *Journal of Economic Dynamics and Control* 28(10), 1977–2002.
- Baak, S.J., 1999. Tests for bounded rationality with a linear dynamic model distorted by heterogeneous expectations. *Journal of Economic Dynamics and Control* 23, 1517–1543.
- Ball, L., Mankiw, N.G., Reis, R., 2005. Monetary policy for inattentive economies. *Journal of Monetary Economics* 52, 4.
- Branch, W.A., 2004. The theory of rationally heterogeneous expectations: evidence from survey data on inflation expectations. *Economic Journal* 114, 497.
- Branch, W.A., Evans, G.W., 2005. A simple recursive forecasting model. *Economics Letters*, forthcoming.
- Branch, W.A., Carlson, J., Evans, G.W., McGough, B., 2004. Monetary Policy, Endogenous Inattention, and the Inflation-Output Variance Tradeoff, Mimeo.
- Brock, W.A., Durlauf, S.N., 2004. Macroeconomics and Model Uncertainty, Mimeo.
- Brock, W.A., Hommes, C.H., 1997. A rational route to randomness. *Econometrica* 65, 1059–1160.
- Brock, W.A., Hommes, C.H., Wagener, F.O.O., 2005. Evolutionary dynamics in markets with many trader types. *Journal of Mathematical Economics* 41(1–2), 7–42.
- Bryan, M.F., Venkatu, G., 2001a. The demographics of inflation opinion surveys. *Economic Commentary*, Federal Reserve Bank of Cleveland.
- Bryan, M.F., Venkatu, G., 2001b. The curiously different inflation perspectives of men and women. *Economic Commentary*, Federal Reserve Bank of Cleveland.
- Calvo, G., 1983. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 12, 393–398.
- Carroll, C.D., 2003. The epidemiology of macroeconomic expectations. *Quarterly Journal of Economics* 118, 1.
- Chavas, J.-P., 2000. On information and market dynamics: the case of the U.S. beef market. *Journal of Economic Dynamics and Control* 24, 833–853.
- Croushore, D., Stark, T., 2003. A real-time data set for macroeconomists: does the data vintage matter? *Review of Economics and Statistics* 85, 605–617.
- Csiszar, I., 1991. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics* 19 (4), 2032–2066.
- Evans, G.W., Honkapohja, S., 2001. *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, NJ.
- Fry, T.R.L., Harris, M.N., 2005. The dogit ordered generalized extreme value model. *Australian and New Zealand Journal of Statistics*, forthcoming.
- Heemeijer, P., Hommes, C.H., Sonnemans, J., Tuinstra, J., 2004. Forming Price Expectations in Positive and Negative Feedback Systems, Mimeo.
- Hommes, C.H., Sonnemans, J., Tuinstra, J., van de Velden, H., 2005. A strategy experiment in dynamic asset pricing. *Review of Financial Studies* 18, 955–980.
- Mankiw, N.G., Reis, R., 2002. Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve. *Quarterly Journal of Economics* 117 (4), 1295–1328.
- Mankiw, N.G., Reis, R., Wolfers, J., 2003. Disagreement about inflation expectations. In: Gertler, M., Rogoff, K. (Eds.). *NBER Macroeconomics Annual* 2003.
- Manski, C.F., McFadden, D., 1981. *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA (Trilligator, Elsevier).
- McFadden, D., 1984. Econometric analysis of qualitative response models. *Handbook of Econometrics*, vol. II. In: Griliches, Z., Intriligator, M.D. (Eds.). Elsevier, Amsterdam.
- Pagan, A., Ullah, A., 1999. *Nonparametric Econometrics*. Cambridge University Press, Cambridge.

- Pesaran, M.H., Timmermann, A., 1995. Predictability of stock returns: robustness and economic significance. *The Journal of Finance* 50 (4), 1201–1228.
- Reis, R., 2004. Inattentive Producers, Mimeo.
- Shore, J.E., Johnson, R.W., 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum-cross entropy. *IEEE Transactions on Information Theory* 26, 26–37.
- Sims, C.A., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50 (3), 665–690.
- Sims, C.A., Zha, T., 2005. Were there regime switches in U.S. monetary policy? *American Economic Review*, forthcoming.
- Souleles, N.S., 2004. Consumer sentiment: its rationality and usefulness in forecasting expenditure – evidence from the Michigan micro data. *Journal of Money, Credit, and Banking* 36(1), 39–72.
- Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14 (1), 11–30.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.
- Williams, J.C., 2003. Discussion of ‘Disagreement about Inflation Expectations’. In: Gertler, M., Rogoff, K. (Eds.). *NBER Macroeconomics Annual* 2003.
- Woodford, M., 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ.