Sticky Information and Economic Dynamics

by

Evan Hart

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Economics

Dissertation Committee:

Jeremy Piger, Chair

Bruce McGough, Co-Chair

George Evans, Core Member

Roberto Gutierrez, Institutional Representative

University of Oregon

Spring 2024

Dissertation Abstract

Evan Hart

Doctor of Philosphy in Economics

Title: Sticky Information and Economic Dynamics

In this dissertation, I investigate economic dynamics under the sticky information model assumption. First, I propose a novel method for evaluating the likelihood of a nonlinear model with time-varying parameters and endogenous variables. Using this method, I estimate model parameters and unobserved time-varying parameters of the sticky information Phillips curve. Finally, I adapt a bounded rationality assumption to an endogenous sticky information model, further enriching our understanding of economic behavior under these conditions.

In Chapter 1, I propose a method to evaluate the likelihood of a nonlinear model with time-varying parameters and endogenous variables. Existing techniques to estimate time-varying parameter models with endogenous variables are restricted to conditionally linear models. The proposed approach modifies a Sequential Monte Carlo filter to evaluate the likelihood of a nonlinear process with an endogenous variable. The modified filter augments the typical measurement and state equations with an equation incorporating instrumental variables. I evaluate the performance of a Bayesian estimator based on the likelihood calculation using simulations and find that the approach generates accurate estimates of both parameters and the unobserved time-varying parameter.

In Chapter 2, I analyze the empirical evidence of variation in a structural parameter of the sticky information Phillips curve. This involves scrutinizing both the statistical significance of the variation and its economic implications. Upon examination, I discover a systematic trend in firms' attention to relevant macroeconomic conditions, indicating a decline in attention over time.

In Chapter 3, I study the stability of equilibrium in a general equilibrium model with information frictions. The equilibrium attentiveness rate is stable under a decreasing gain adaptive learning scheme. This stability motivates a review of the transition between equilibrium rates; a drop in the cost of gathering and processing information is used to shift the equilibrium. The attentiveness rate immediately jumps and increases asymptotically, approaching the new equilibrium.

Contents

List of Figures

List of Tables

1 A Sequential Monte Carlo Approach to Endogenous Time Varying Parameter Models

1.1 Introduction

I propose a method to evaluate the likelihood of a nonlinear model with time-varying parameters and endogenous variables. Existing techniques to estimate time-varying parameter models with endogenous variables are restricted to conditionally linear models. The proposed approach modifies a Sequential Monte Carlo filter to evaluate the likelihood of a nonlinear process with an endogenous variable. The modified filter augments the typical measurement and state equations with an equation incorporating instrumental variables. I evaluate the performance of a Bayesian estimator based on the likelihood calculation using simulations and find that the approach generates accurate estimates of both parameters and the unobserved time-varying parameter.

Complex models are becoming more prevalent and essential to understanding macroeconomic dynamics. Many of these models allow for nonlinearities and shifts in parameters. Cogley and Sargent (2005) use evidence about parameter drift and stochastic volatility to infer that monetary policy rules have changed and that the persistence of inflation itself has drifted over time. Fernández-Villaverde et al. (2007) show how the pricing parameters' movements correlate with inflation, casting doubt on the empirical relevance of Calvo models.

Time-varying parameter (TVP) models have gained considerable traction in various domains. By allowing parameters to vary over time, TVP models provide a dynamic view of relationships within data. Works by Cogley and Sargent (2005), Primiceri (2005), and Koop and Korobilis (2013) have primarily delved into time-varying structural vector autoregressions (VAR). While scholars like Young (2004) and Fernández-Villaverde et al. (2007) have extended the TVP assumption into dynamic stochastic general equilibrium (DSGE) models, broadening its applicability. Notably, many of these models adopt a random walk for TVP, the most versatile choice, capturing various forms of parameter variation.

Within the realm of empirical macroeconomic research, scholars have explored conditionally linear TVP models incorporating endogenous regressors. For instance, Peersman and Pozzi (2004) and Kim (2006) have contributed significant insights in this area. Nonetheless, these studies often face challenges related to heteroscedasticity, stemming from substituting the fitted value from the instrumenting equation for the actual value. Addressing this concern, Kim and Kim (2011) propose a remedy utilizing a Kalman filter for linear models, although this method has not been extended to nonlinear models.

A method to estimate a TVP model with nonlinearities, where the time-varying parameter is on an endogenous variable and instrumental variables are available, is not previously defined in the literature. I fill this gap in the literature by adapting the Sequential Monte Carlo filter to run a joint estimation procedure. The Sequential Monte Carlo filter, also called a particle filter, finds the likelihood of the data conditioned on a given set of parameters. It does this by proposing a set of states for the time-varying parameter. The probability of each state is calculated and used to sample from the initial proposal. This procedure is iterated for each period. The likelihood of the data is the product across states of the average likelihood over the accepted particles for each state.

I refine the likelihood estimation by modeling it as a multivariate normal distribution, where potential errors are matched with instrumental errors. These potential errors are determined at each period through a dual-sampling approach, while the instrumental errors are derived from static parameters that initialize the particle filter and remain known during the sampling process. This methodological adjustment offers a novel approach for estimating TVP models with nonlinearities, endogenous variables, and instrumental variables, enhancing the analytical toolkit available for modeling complex systems.

The proposed methodology is assessed through Monte Carlo simulations. Initially, I illustrate that the unmodified Sequential Monte Carlo filter yields a biased estimation of the time-varying parameter in the presence of endogeneity. Subsequently, I demonstrate that the modified Sequential Monte Carlo filter produces estimates of the time-varying parameter that exhibit minimal bias. To ascertain the robustness of the extended particle filter, I conduct simulations using a simple linear model and contrast the outcomes when instrumental variables are and aren't included. Furthermore, I enhance this robustness assessment by comparing the filter's performance under misspecified versions of the underlying distribution and examining resultant discrepancies in the estimated parameters. This comprehensive evaluation underscores the efficacy and reliability of the proposed extended particle filter for estimating time-varying parameters in the presence of endogeneity and instrumental variables.

The rest of the paper has the following structure. First, I develop a TVP model with endogenous regressors. Then, I explain the traditional method that can be used to estimate such a model but fails to account for the endogeneity. I go on to explain a joint estimation modification to the existing particle filter. Section 5 presents Monte Carlo experiments, in which I compare the performance of the proposed joint method to the traditional method.

1.2 State Space Model

The state space model provides the foundation for time series data analysis. It is widely used for forecasting, structural, and unstructured modeling. First I lay out a standard system, with normal errors. Then I introduce the system of interest, in which a regressor is endogenous. A particle filter can be adapted to estimate such a system.

The particle filter estimates the likelihood of a data set given a state space model and a set of parameters to define that model. The particle filter does this with a number of assumptions. The time path is conditional on given parameters, the distribution assumptions of the stochastic terms, and the parameters to define those distributions. We must use a filter for estimation when an unobserved state variable follows a time series process. When the measurement or the transition equation is nonlinear, the particle filter can be used instead of the Kalman filter.

The Kalman filter stands as a preferred method in scenarios where errors conform to normal distribution and the model maintains conditional linearity. However, when encountering disturbances deviating from normality in either the measurement or transition equation, the particle filter emerges as a viable alternative to the Kalman filter. Its adaptability to nonnormal conditions renders it a valuable tool in situations where traditional methods may falter, offering robustness and flexibility in handling complex data dynamics.

Consider the following conditionally linear state space model with a random walk time varying parameter:

$$y_t = \lambda_t x_t + \nu_t, \tag{1}$$

$$\lambda_t = \lambda_{t-1} + \omega_t, \qquad \omega \backsim N(0, \sigma_\omega^2). \tag{2}$$

The measurement equation (1) is defined by a set of parameters: θ, state variable λ_t, and a stochastic error term: ν_t. This model includes a single exogenous regressor: x_t. The transition equation (2) gives structure to the process that governs the evolution of the time-varying parameter. The procedure for the traditional particle filter will be discussed in terms of normally distributed and independent identical draws of ν and ω for each period; the procedure can be extended to other distributions.

To incorporate endogeneity into the model, I introduce the following assumptions. Initially, I assert the existence of an instrumental variable z_t. The instrument is a relevant predictor of the endogenous variable and influences the dependent variable solely through this endogenous variable. Subsequently, I presume that the joint distribution between v_t and the disturbance term in the equation linking an endogenous variable x_t to the instrument z_t is known. Particularly, I adopt a joint normality assumption:

$$x_t = mz_t + \xi_t \tag{3}$$

$$\nu, \xi \backsim N(0, \Sigma), \qquad \Sigma = \begin{bmatrix} \sigma_\nu^2 & \rho * \sigma_\nu * \sigma_\xi \\ \rho * \sigma_\nu * \sigma_\xi & \sigma_\xi^2 \end{bmatrix} \tag{4}$$

Equation (3) links the endogenous variable to the instrumental variable. Equation (4) identifies the endogeneity in the system. I assume joint normality. This is a crucial assumption. In a model with a TVP on the endogenous regressor, a solution with a single compound error would suffer from heteroscedasticity, stemming from substituting the fitted value from the instrumenting equation for the actual value. Joint normality is a common assumption made in the literature. The control function literature approach uses a similar assumption about the error terms. An example of the control function literature: Heckman (1979) makes a similar assumption for a linear model with an endogeneity issue.

We can gain some intuition from this simple model in the following way. Suppose the covariance factor ρ is positive. As ξ gets larger, so does the conditional mean of ν. Therefore, as ξ gets larger, positive nonzero values of ν should have a higher likelihood. This means that states with error of zero will be selected less during the second sample period. A univariate distribution fixes the mean at zero and fails to adjust to the larger expected value of ν, biasing the estimate of λ_t upward.

1.3 Review of the Particle Filter

The particle filter can evaluate the likelihood of data given a state space model, a set of parameters to define that model, and exogenous independent variables. The state space model is represented by equations: (1), (2), and (3). The particle filter finds the likelihood by estimating a set of states for each period, then taking the average likelihood of the sample for each period and taking the product between periods. The particle filter finds a sample for each period's time-varying parameter through a double-sampling process. The first sample is a naive sample based on the parameters of the stochastic process governing the time series process. The second sample is an informed sample that uses the relative probability of the states to resample from the naive sample states.

The first sample of states is found using the transition equation, this sample is denoted: $\lambda_0^{t|t-1,i}$. Under the random walk assumption, each state is the sum of the previous value of λ and a random variable ω. This fact is expressed in our notation: $t|t-1$, each period λ_t is conditioned on the previous value of λ_{t-1}. For each period, N values are simulated for ω_t from the model distribution. The first sample of $\omega_0^{t,i}$ is denoted by i from 1 to N, notice that the first sample is not conditioned on the previous value of λ. The first sample of states now has the following mathematical expression:

$$\lambda_0^{t|t-1,i} = \lambda^{t-1|t-2,i} + \omega_0^{t,i}.$$

At time $t = 1$, the previous value: λ^0 is not conditioned on a previous value, and we assume all N paths start at the same value of λ^0.

The second sample is an informed sample that uses the relative probability of the states. The probability for each state is the likelihood of the error created by that state divided by the total likelihood from all the sample errors. The likelihood of an error is the value of the density function at the error: $f(\nu_0^{t|t-1,i})$. We must know the distribution and have a suggestion for the parameters that define the distribution. The error is the difference between each prediction and the observation value. The prediction is found using the measurement equation. From these predictions, we find the initial sample of errors $\nu_0^{t|t-1,i}$:

$$y_t = h(X_t, \theta, \lambda_0^{t|t-1,i}) + \nu_0^{t|t-1,i}.$$

We then resample based on the relative probability of the suggestions:

$$p(\lambda_0^{t|t-1,i}) = \frac{f(\nu_0^{t|t-1,i})}{\sum_{i=1}^{N} f(\nu_0^{t|t-1,i})}$$

This final sample is saved, now denoted $\lambda^{t|t-1,i}$ with corresponding values of $\nu^{t|t-1,i}$, and the process is repeated for period t+1.

Using the error terms from the final sample of the states, we can calculate the likelihood of the data. The per-period likelihood is the average likelihood of the final sample:

$$f(\nu^{t|t-1}|\theta) = \frac{1}{N}\sum_{i=0}^{N} f(\nu^{t|t-1,i}|\theta).$$

Independence of ν across periods allows us to take the product of the per period likelihood to find the likelihood of all the data:

$$L(Y^T|\theta) = \prod_{t=1}^{T} f(\nu^{t|t-1}|\theta).$$

The procedure defined above will provide a likelihood function that asymptotically peaks at the correct value of the model parameters provided that X_t is exogenous. However, if X_t is endogenous, the likelihood function will peak at the wrong values of model parameters. Estimation routines built on this likelihood calculation will then produce biased estimates of model parameters and the time-varying process λ_t. In the next section I present a modified version of the particle filter that corrects for this bias.

1.4 A Modified Particle Filter Incorporating Instrumental Variables

If x_t is endogenous, an estimate of our time varying parameters using an uncorrected particle filter will be inconsistent. A particle filter using a joint distribution between the instrumental and measurement error will estimate consistent parameters. Endogeneity here is represented by a correlation between the variable and the measurement error. Assuming we have instrumental variables, we can estimate an instrumental error, ξ_t, that contains the part of x_t that is correlated with v_t:

$$x_t = mz_t + \xi_t$$

The particular filter procedure remains the same. However, the likelihood used for the second sample is a multivariate distribution. After taking the initial naive sample and finding the error term ν, we can use a joint distribution between the measurement and instrumental error to find the likelihood of each state. The probability that a state is selected now considers the correlation between ξ_t and $\nu^{t|t-1,i}$:

$$p(\lambda^{t|t-1,i}) = \frac{g(\nu^{t|t-1,i}, \xi_t)}{\sum_{i=1}^{N} g(\nu^{t|t-1,i}, \xi_t)}$$

Using the error terms from the final sample of the states and the instrumental error, we can calculate the per period likelihood as the average likelihood between the instrumental error and the final sample state errors:

$$g(\nu_t, \xi_t|\theta, m) = \frac{1}{N}\sum_{i=0}^{N} g(\nu_{t,i}, \xi_t|\theta, m),$$

Independence still holds across periods, allowing us to take the product of the per-period likelihood to find a likelihood that will give unbiased estimates of model parameters:

$$L(Y^T|\theta, m) = \prod_{t=1}^{T} g(\nu_t, \xi_t|\theta, m).$$

1.5 Benefits of an Extended Filter

To analyze the benefits of the particle filter, I will compare the output of an extended particle filter to a traditional particle filter. The comparison is made across models and levels of covariance. It is beneficial for the accuracy of the time-varying parameter and the accuracy of the static parameters to include instruments when there is endogeniety. When covariance is zero, the extended particle filter makes slightly less accurate estimates.

Any maximum likelihood method can be used to search over possible parameters; I choose to implement a Bayesian Markov Chain Monte Carlo. In this estimation method, we collect samples of the parameters of interest to specify a posterior distribution of the parameters. The sampler needs specific prior beliefs about those parameters. I choose to implement flat priors for all my variables. Flat priors will provide a clean starting point to see this bias. We have a burn-in period for each iteration of the MH sampler, after which we collect 2000 samples of the parameters. Then, I take the median of these samples.

To assess the bias in the traditional particle filter, I conduct simulations for 102 datasets with a length of T=250. At each iteration, I gather point estimates and a TVP path for both types of filters. In tables, I compile the mean of the estimates and the standard deviation between the samples to facilitate comparison. Furthermore, to evaluate the bias in the TVP, I graph and contrast the average deviation between the true value and the filtered estimate of the TVP.

1.6 Model I

The first test will use a nonlinear model with an additive error. This sample model was chosen because of its relation to a sticky information Phillips curve with drift in the rate of attentiveness. Endogeneity is known to be an issue when estimating the Phillips curve, and more sophisticated models will check for variation in the parameters. Each parameter will have a possible range shown in Table 12. The model has the following form:

measurement equation:

$$y_t = \frac{1-\lambda_t}{\lambda_t} x_{1,t} + \beta x_{2,t} + \nu_t,$$

transition equation:

$$\lambda_t = \lambda_{t-1} + \omega_t,$$

instrumental equation:

$$x_{1,t} = m z_t + \xi_t.$$

After gathering a deviation path for each dataset, I calculate the average deviation at each period and plot this deviation in Figure 1. In the time series graph where the correlation coefficient is zero, we can observe that both the particle filter and the extended particle filter exhibit deviations ranging from -0.3 to 0.2, with the particle filter showing a slightly closer deviation to 0. The extended particle filter estimates a slightly negative covariance, biasing the path downward, as observed in Table 1. However, when we introduce endogeneity, the traditional particle filter exhibits a significant downward skew, converging to a deviation of -2. In contrast, the extended particle filter fluctuates slightly around 0. The bias in the TVP introduces a bias in the static parameters under full estimation. This bias arises because the peak likelihood value is no longer at the correct parameters. This bias is also evident in Table 2; the traditional particle filter struggles to identify the static parameter β. Both filters perform well in estimating the initial value of the TVP, but they both overestimate the variance of the stochastic component of the random walk that forms the structure of the TVP.

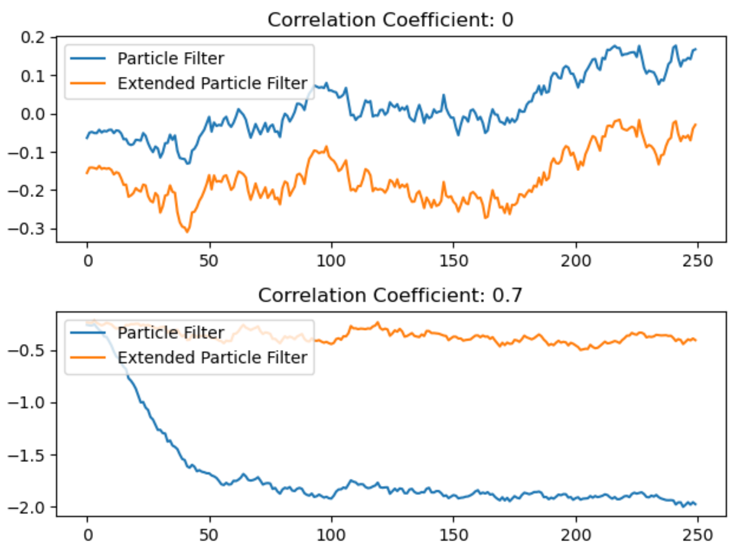

Figure 1: Absolute Deviation of λ_t, Model I.

	Extended Particle Filter	Particle Filter
β =1	0.959 (0.054)	0.976 (0.045)
$\sigma_\nu = 1$	0.957 (0.049)	0.963 (0.043)
λ_0 =4	3.845 (0.384)	3.934 (0.239)
σ_ω =0.1	0.167 (0.025)	0.151 (0.021)
σ_ξ =1	1.005 (0.049)	-
$\rho = 0$	-0.066 (0.074)	-
$m = 1$	0.981 (0.049)	-
Acceptance Rate	0.400	0.578

Table 1: Metropolis-Hastings Sampler Results for Model 1

	Extended Particle Filter	Particle Filter
$\beta = 1$	0.971 (0.041)	0.841 (0.036)
$\sigma_\nu = 1$	0.897 (0.054)	0.784 (0.036)
$\lambda_0 = 4$	3.735 (0.311)	3.736 (0.125)
$\sigma_\omega = 0.1$	0.181 (0.018)	0.167 (0.014)
$\sigma_\xi = 1$	1.000 (0.048)	-
$\rho = 0.7$	0.685 (0.041)	-
$m = 1$	1.055 (0.048)	-
Acceptance Rate	0.344	0.310

Table 2: Metropolis-Hastings Sampler Results for Model 1

1.7 The Role of Instruments

The joint normal assumption furnishes the necessary structure for estimating the models and potentially enables us to directly model a relationship between the endogenous variable and the measurement error. In this experiment, there actually is an instrument (m does not equal zero), but the researcher assumes that m equals zero. A useful tool if a researcher doesn't have access to instruments or available instruments are weak. To test the effect of excluding instruments, I will simulate a model with instruments and compare a particle filter that uses those instruments and one that does not. To compare the effectiveness of the two filters, I will again provide the estimation results of the static parameters and the per period mean squared error of the TVP. I will again use an MH sampler and over 102 independent data sets; priors are in Table 13.

The test will use a linear model with an additive error. This sample model was chosen because of its simplicity. The model has the following form:

measurement equation:

$$y_t = \lambda_t * x_{1,t} + \beta * x_{2,t} + \nu_t,$$

transition equation:

$$x_1 = mz_t + \xi_t,$$

instrumental equation:

$$\lambda_t = \lambda_{t-1} + \omega_t.$$

The exclusion of instruments has had a minimal effect on the parametrization I have chosen here. In Table 3, we can see that the estimation of static parameters is very close. However, the per period mean squared error of the TVP is higher for the model that doesn't implement the instruments, as shown in Figure 2. I interpret this as motivation for utilizing instruments if they are available. However, if instruments are not available or are weak, the extended particle filter would be a robust choice.

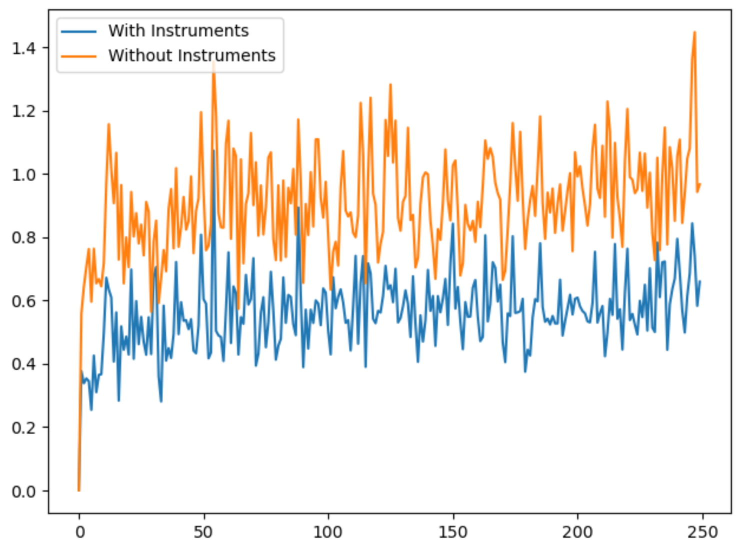

Figure 2: Per-Period Mean Squared Error of λ_t

	With Instruments	Without Instruments
$\sigma_\nu = 1$	0.933 (0.027)	0.995 (0.023)
β =1	1.001 (0.024)	1.000 (0.023)
λ_0 =1	1.000 (0.024)	0.999 (0.023)
σ_ω =1	1.009 (0.024)	1.002 (0.023)
σ_ξ =1	1.001 (0.024)	1.011 (0.023)
$\rho = 0.7$	0.823 (0.040)	0.707 (0.016)
m=1	1.000 (0.024)	-
Acceptance Rate	0.368	0.348

Table 3: Metropolis-Hastings Sampler Results for Linear Model

1.8 Robustness to Misspecification

The method relies heavily on the joint normal assumption between the variables. To test the robustness of the extended particle filter to misspecification, I will compare particle filters when the true error has a joint t-distribution with three degrees of freedom. One particle filter will use the normal assumption, while the other correctly specifies a t-distribution with three degrees of freedom. I will compare the particle filters using an MH sampler and the following priors using 102 independent data sets. The model again has the following form:

measurement equation:

$$y_t = \lambda_t * x_{1,t} + \beta * x_{2,t} + \nu_t,$$

transition equation:

$$x_1 = m z_t + \xi_t,$$

$$\nu, \xi \backsim T(0, \Sigma, df = 3), \qquad \Sigma = \begin{bmatrix} \sigma_\nu^2 & \rho * \sigma_\nu * \sigma_\xi \\ \rho * \sigma_\nu * \sigma_\xi & \sigma_\xi^2 \end{bmatrix} \tag{5}$$

instrumental equation:

$$\lambda_t = \lambda_{t-1} + \omega_t.$$

The change in the data-generating process of the ν and ξ to a t-distribution with three degrees of freedom has an interesting effect on the particle filter that incorrectly specifies a Normal Distribution. While the estimates are fairly accurate, the acceptance rate is very low, and the variation between the acceptance rates of independent draws is large. With accurate starting values, the extended particle filter is robust to misspecification.

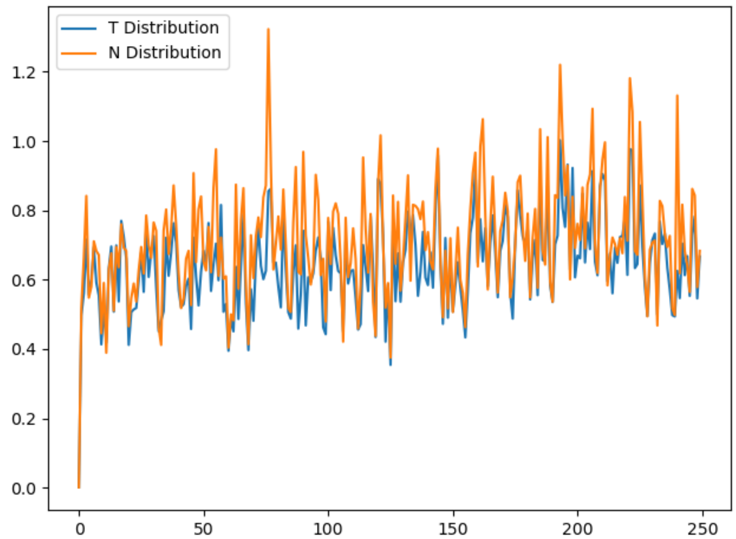

Figure 3: Per-Period Mean Squared Error of λ_t

	T Distribution	N Distribution
$\sigma_\nu = 1$	0.955 (0.026)	0.979 (0.023)
β =1	1.000 (0.024)	0.998 (0.023)
λ_0 =1	1.000 (0.024)	0.998 (0.023)
σ_ω =1	1.013 (0.024)	1.009 (0.023)
σ_ξ =1	1.012 (0.024)	1.052 (0.027)
$\rho = 0.7$	0.777 (0.031)	0.739 (0.021)
m=1	1.002 (0.024)	0.998 (0.023)
Acceptance Rate	0.268	0.129

Table 4: Metropolis-Hastings Sampler Results for Non-Normal Model

1.9 Conclusion

A method to estimate a TVP model with nonlinearities, where the time-varying parameter is on an endogenous variable and instrumental variables are available, is not previously defined in the literature. I fill this gap in the literature by adapting the particle filter to run a joint estimation procedure. The IV estimation was shown to correct for bias in the TVP and static parameters when there is endogeneity. In addition, I show results demonstrating the flexibility of this method. We can run an estimation procedure without instruments. Finally, I show the method's robustness to misspecification.

2 Time Variation in the Sticky Information Phillips Curve

2.1 Introduction

I analyze the empirical evidence of variation in the structural parameter of the sticky information model. This involves scrutinizing both the statistical significance of the variation and its economic implications. Upon examination, I discover a systematic trend in firms' attention to relevant macroeconomic conditions, indicating a decline in attention over time. This decline in attention is found to have significant economic implications, particularly in relation to the expected volatility of prices. By incorporating models proposed by Ball et al. (2003), I demonstrate that this trend in attention is associated with a decrease in the expected volatility of prices. Comparisons between fixed attention models and time-varying attention models reveal substantial differences in parameter estimates. A Bayesian model comparison favors a fixed parameter model but does not eliminate the time varying parameter model altogether. These findings underscore the importance of considering time-varying attention in economic modeling and policymaking, as it provides valuable insights into the evolving nature of decision-making processes in response to macroeconomic information.

In the Sticky Information Phillips curve, attention is used as a nominal rigidity. Firms strive to set a profit-maximizing price path based on the current aggregate price and output, but they do not have access to future data. Therefore, they must rely on forecasts of these variables to set future prices in their path. Attention is modeled as a probability that they are given a current history of relevant variables to use in their forecasts. Empirical research on attention to economic conditions holds relevance for both macroeconomic policy and firm strategy. Identifying this aggregate attention presents an opportunity for the Federal Reserve to enact policies that are more effective. In my analysis of attentiveness, I identify variations that are important economically and cannot be ruled out statistically.

Research into structural models employs various forms of imperfect knowledge. The first is Lucas (1972a), where agents possess limited knowledge about changes in prices. A more recent perspective introduced by Sims (2003) posits that agents suffer from limited attention, preventing them from processing all available information. In my study, I employ the Phillips Curve developed by Mankiw and Reis (2002), where the likelihood of setting a fixed price is substituted with the probability of receiving new information. (Henceforth referred to as SIPC.)

We may anticipate that the rate of attention changes over time. Branch et al. (2009) proposed a model where the attention rate is an equilibrium value determined by the current state of the economy. For instance, the attention rate declines if the costs associated with acquiring and processing information increase. These variations are significant because the level of attention influences the economy's responsiveness to external shocks. Therefore, I estimate a model that permits the rate of attention to follow a random walk process. This approach enables the rate to change in every period of the model, allowing the data to dictate the path instead of imposing a more structured pattern such as a structural break.

Shifts in attentiveness have been explored through various frameworks. Coibion and Gorodnichenko (2015) utilize a method that links ex post mean forecast errors to ex ante revisions in the average forecast among surveyed professionals, providing a benchmark for evaluating potential deviations by economic agents. On the other hand, Carroll (2003) constructs a model that compares a survey of household expectations to the forecasts from the Survey of Professional Forecasters.

The empirical approach employed in this paper closely resembles that implemented by Coibion (2010). The expectations of firms will be represented using data from the Survey of Professional Forecasters, replacing the conventional rational expectations theory, wherein historical values of the variable of interest would be used. In the SIPC, adopting historical values would result in an error process highly correlated with the regressors and instruments. In addition, this model uses an error term that is assumed to be additive to the inflation equation and drawn from an independent and identically distributed (IID) process. This approach deviates from the AR(1) process used to introduce uncertainty in the models of Ball et al. (2003) and Branch et al. (2009).

Using Bayesian estimation, I identify a downward trend in the rate of attention over the sample period. To determine if the changes are economically significant, I turn to a general equilibrium model developed by Ball et al. (2003). This enriched model uses a similar sticky information Phillips curve but creates a general equilibrium model by specifying a demand curve and an objective function by which monetary policy can be derived. Using this enriched model, I have uncovered economically significant changes. Specifically, the shift in attention translates to a substantial decrease in the expected volatility of aggregate price, ranging from 5.5% to 12.7%. These findings highlight the potential impact of implementing the proposed adjustments, suggesting a notable improvement in the stability of aggregate price movements.

Bayesian model comparison allows us to gauge how credible two models are based on our data. Chib and Jeliazkov (2001) provide the framework for estimating the marginal likelihood when implementing a Markov chain Monte Carlo estimation. The Bayesian model comparison places a 16% probability on the TVP model. This means it cannot be ruled out statistically.

Enriched attention models can inform us of the causes of these behaviors. In Branch et al. (2005), firms will decrease their attention rate in response to an increase in the cost of absorbing aggregate information. The increasing volume of available data is usually associated with increased information costs. In Hart (2023), I extended the Branch et al. (2005) model by introducing a bounded rationality assumption. In response to large shocks, firms will be temporarily more attentive. This response is caused by the belief that the economic shocks are drawn from a distribution with a higher standard deviation.

The rest of the paper has the following structure. First, I develop a TVP version of the SIGE. Then, I explain what data will be used and how it is structured. I go on to explain the estimation method and priors. Finally, I present and explore the results of the estimation.

2.2 Empirical Approach

2.3 Model

The assumption developed by Mankiw and Reis 2002 is used to model inflation. Firms use outdated information until they are randomly "selected" to receive new information. The parameter that measures firms' attention is (1-λ), the probability of being selected. Aggregate price is assumed to be a weighted average over firms' expectation of the optimal price:

$$\pi_t = \frac{(1-\lambda)}{\lambda}\alpha x_t + \sum_{j=0}^{J-1}(1-\lambda) * \lambda^j E_{t-1-j}(\pi_t + \alpha \Delta x_t) + \nu_t.$$

The weighting system is formed by the fixed probability of receiving new information. The proportion of firms using the newest information set I_{t-1} is $(1-\lambda)$. The proportion of firms using the information set I_{t-2} is $(1-\lambda)\lambda$, the probability of receiving new information last period multiplied by the probability of staying in that information set. As an information set becomes more outdated it will shrink by the probability of staying in that information set each period.

If we allow the probability of new information to change over time the weighting system is formed by the history of probabilities of receiving new information. The proportion of firms using the newest information set I_{t-1} is $(1-\lambda_t)$. The proportion of firms using the information set I_{t-2} is $(1-\lambda_{t-1})\lambda_t$, the probability of receiving new information last period multiplied by the probability of staying in that information set. This time-varying parameter model has the following form of the SIPC:

$$\pi_t = \frac{(1-\lambda_t)}{\lambda_t}\alpha x_t + \sum_{j=0}^{J-1} a_{t,j+1} E_{t-1-j}(\pi_t + \alpha \Delta x_t) + \nu_t, \tag{6}$$

s.t

$a_{t,1} = (1-\lambda_t)$

$a_{t,2} = \lambda_t * (1-\lambda_{t-1})$

$a_{t,3} = \lambda_{t-1} * \lambda_t * (1-\lambda_{t-2})$

$a_{t,4} = \lambda_{t-2} * \lambda_{t-1} * \lambda_t * (1-\lambda_{t-3})$

$a_{t,5} = \lambda_{t-3} * \lambda_{t-2} * \lambda_{t-1} * \lambda_t * (1-\lambda_{t-4})$.

$$x_t = \delta Z_t + \xi_t, \tag{7}$$

$$\lambda_t = \lambda_{t-1} + \omega_t. \tag{8}$$

The transition equation is modeled as a random walk. The selection of a random walk stems from uncertainty regarding the underlying process. Prior to this study, there has been no estimation of an SIPC with a time-varying parameter. The lack of structure placed on the random walk assumption allows the data to inform us on the underlying form.

2.4 Data

The data is naturally divided into two groups. First, we have realized times series data on inflation, output and the natural rate of output. Second, we have firm's expectation. Expectation data of firms will be proxied using expectation data from the survey of professional forecasters.

The GDP price index forecasts are the median from the survey of Professional Forecasters. Collected from the Philadelphia Federal Reserve website. This data set runs from 1968-10-01 to 2022-10-01. The file includes 9 columns, two are dedicated to the date, two are annual average forecasts. The other 6 contain our expectation data. For a row labeled time t, we have the expectations using information set $t-1$. Using this information set, the median forecast is provided starting with P_{t-1} ending with P_{t+4}. Forecasts for the quarterly level of the chain-weighted GDP price index can be used to construct an expectation of inflation using the following formula:

$$E_{t-1}\pi_{t+i} = log(E_{t-1}P_{t+i}) - log(E_{t-1}P_{t+i-1}) \text{ for } i \in \{0, 1, 2, 3, 4\}$$

The Real GDP forecasts are the taken from the same source and have the same format. No expectation data on the change in natural output is proxied with the actual change. The expected change in the output gap can be broken into two pieces given our simplifying assumption. The expectation data will form the first part. While the second part will use the actual change in the output gap. Therefore, the expected change in the output gap has the following form:

$$E_{t-1}\alpha\Delta x_{t+i} = \alpha * (E_{t-1}\Delta y_{t+i} - \Delta y^N_{t+i})$$

s.t.

$$E_{t-1}\Delta y_{t+i} = log(E_{t-1}(Y_{t+i})) - log(E_{t-1}(Y_{t-1+i})),$$

$$\Delta y^N_{t+i} = log(Y^N_{t+i}) - log(Y^N_{t+i-1}) \text{ for } i \in \{0, 1, 2, 3, 4\}.$$

Inflation is measured using the implicit GDP price deflator. The output gap is the log difference between real gross domestic product and the CBO measure of potential output.

2.5 Estimation Method

A particle filter will be used to evaluate the likelihood of the model, and the Metropolis-Hastings algorithm used to simulate from the posterior. The highly nonlinear nature of the measurement equation means that traditional estimation methods cannot be used. The Metropolis Hastings algorithm is a simple algorithm for producing samples from the distribution that otherwise would be hard to characterize.

The MH sampler can be used to take a sample of parameter values to characterize the posterior distribution. The prior beliefs of this model use a gamma distribution to characterize the standard deviation parameters and a normal distribution for α and λ_0. There are 1,000 burn in draws and 2,000 post convergence draws. Before beginning the sampler, we must choose a θ_0 and the variance-covariance matrix R, used to create variation from the previously accepted value of theta.

The stochastic terms are assumed to be normally distributed. Standard deviations of the stochastic terms in model equations are assumed to have a Gamma distribution with hyper-parameters α and β.

$$\sigma_{\nu}^{-1} \backsim Gamma(\alpha_{\nu}, \beta_{\nu}) \quad st \quad \alpha_{\nu} = 100 \quad \beta_{\nu} = 4$$

$$\sigma_{\omega}^{-1} \backsim Gamma(\alpha_{\omega}, \beta_{\omega}) \quad st \quad \alpha_{\omega} = 80 \quad \beta_{\omega} = 1$$

$$\sigma_{\xi}^{-1} \backsim Gamma(\alpha_{\xi}, \beta_{\xi}) \quad st \quad \alpha_{\xi} = 45 \quad \beta_{\xi} = 2$$

All the coefficients are assumed to be drawn for independent normal distributions. Hyper parameters for Normal distribution describing prior for conditional mean parameters:

$$\alpha \backsim N(\mu_{\alpha}, \sigma_{\alpha}^2) \quad st \quad \mu_{\alpha} = 0.1 \quad \sigma_{\alpha}^2 = .001$$

$$\lambda_0 \backsim N(\mu_{\lambda_0}, \sigma_{\lambda_0}^2) \quad st \quad \mu_{\lambda_0} = 0.7 \quad \sigma_{\lambda_0}^2 = 1$$

The priors for the conditional mean parameters in the instrumental variable equation are established in this section. A diffuse prior is employed, wherein the hyperparameters dictate a mean of 0 and a variance of 1. This choice aims to minimize the influence of the prior on our posterior estimates.

$$\delta_i \backsim N(\mu_{\delta_i}, \sigma_{\delta_i}^2) \quad st \quad \mu_{\delta_i} = 0 \quad \sigma_{\delta_i}^2 = 1 \quad \forall i$$

Th covariance between the instrumental error: ξ and the measurement error: ν is also assumed to be drawn from a normal distribution:

$$\sigma_{\nu,\xi} \backsim N(\mu_{\sigma}, \sigma_{\sigma}^2) \quad st \quad \mu_{\sigma} = 0 \quad \sigma_{\sigma}^2 = 1$$

The assessment for weak instruments typically employs an F-test. The leading evaluation for weak instruments in time series models is Olea and Pflueger (2013) approach. Nonetheless, this technique hinges on the linear Ordinary Least Squares (OLS) estimates of the structural equation. An alternative test, Stock and Yogo (2005), has been completed for this model by Coibion (2010) and stands as the next viable option.

2.6 TVP Estimation

Estimating the SIPC with variation in the rate of attention gives us parameter estimates and a time series of firms' attentiveness to information relevant to price setting. The Metropolis Hasting sampler drew 2000 samples of our parameters. These samples can be used to understand an estimate of the parameter and the accuracy of the estimates.

The results of modeling time-varying attention to macroeconomic information are as follows. The final acceptance rate of the draws stood at 0.44, surpassing the recommended threshold in the literature, yet remaining within an acceptable range. The estimated values for α and $1 - \lambda_0$ are 0.06 and 0.46, respectively. The ensuing figures illustrate the generated samples. Each sample of the static parameter models gave rise to a corresponding sample path of λ. By aggregating these sample paths using a time-wise mean, a composite graph was constructed. Subsequently, the 95 % confidence bounds per period were determined.

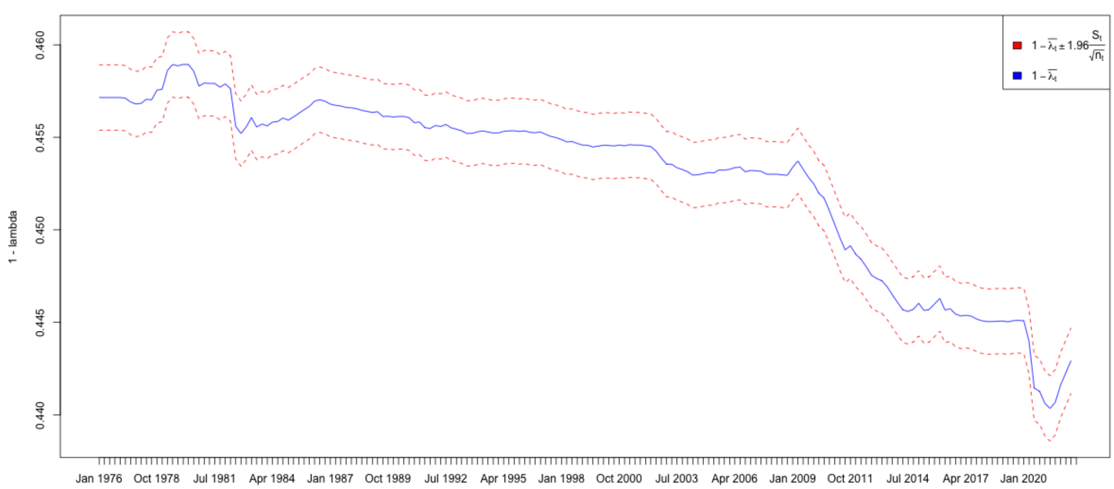

Figure 4: Proportion of firms using the newest information over time.

	Alpha	Lambda0	HV	HW	HE	Covariance
Mean	0.06	0.54	329.01	77.92	90.32	0.00
SD	0.01	0.03	17.88	8.97	5.07	0.00

Table 5: Measure Parameters in TVP Model

	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5
Mean	-0.00	0.81	0.20	-0.14	-0.11
SD	0.00	0.05	0.04	0.02	0.05

Table 6: IV Parameters in TVP Model

The time-varying parameter estimation reveals a consistent trend of declining attention over the analyzed period. Commencing from January 1976, the estimated value of $1-\lambda$ stands at 0.457, progressively decreasing to 0.443 by April 2022. To gauge the economic implications of this trend, I will refer to the model introduced by Ball et al. (2003).

The changes in attention may be driven by shifts in other parameters. The standard deviation of the error used in the measurement equation could be shifting over time. To test for this possibility, another model could be estimated and compared to the ones presented here.

Prior research conducted by Coibion and Gorodnichenko (2015) found that the frequency of forecasts by professional forecasters surveyed does not change over time. Therefore, the shifts in attention are isolated to the model presented and not the data.

2.7 Economic Implications

To analyze the economic impact of the identified shifts in the rate of attention, I use the model developed in Ball et al. (2003). A similar model to the one estimated that incorporates a demand and policy equation to form a general equilibrium model. The demand equation is a simple quantity equation in which nominal spending is proportional to the money supply. The policy rule is derived from maximizing the Woodford and Woodford and Walsh (2005) approximation of utility. In the rational expectations equilibrium, the variance of price can be expressed in terms of the model parameters and the rate of attention. The equation for price variance is given here:

$$\sigma_p^2 = Var(\varepsilon_t)\sum_{j=0}^{\infty}\phi_j^2$$

$$\phi_j = \frac{\rho^j}{\alpha^2\omega + \frac{(1-\lambda)^{j+1}}{1-(1-\lambda)^{j+1}}}$$

Using the parameterization in Branch et al. (2009) , the initial volatility registers at 1.66, tapering down to 1.57 towards the end of the analyzed period. A discernible shift in volatility is observed, a decrease of 5.5% in the variance of price. An alternative parameterization found in Ball et al. (2003) increases the magnitude of the shift to -12.7%. This suggests that policymaking may need adjustments in response to the time-varying estimates of attention. The parameterization taken from that model is $\alpha = 0.1$, $\rho = 0.8$, $\omega = 1$, and $\sigma_\varepsilon^2 = 1$. The parameterization from Branch et al. (2009) is $\alpha = 0.1$, $\rho = 0.8$, $\omega = 20$, and $\sigma_\varepsilon^2 = 0.1$.

2.8 Fixed Parameter Estimation

Fixed estimation is done using a Metropolis Hasting Sampler and a joint likelihood function equivalent to the particle filter with zero variation. This likelihood method was chosen to provide an equivalent measure of the marginal likelihood to compare to the time varying model. The results of modeling fixed attention to macroeconomic information are as follows: the final acceptance rate of the draws stood at 0.33, within the recommended threshold in the literature. The estimated values for α and $1-\lambda$ are 0.07 and 0.45, respectively. The estimated value of λ implies that most firms are expected to update their information twice a year.

Comparing the results for the two model types using the values $1-\lambda$, I find that the mean value of the TVP estimate is 0.45, similar to the fixed estimate. The fixed estimate of 0.55 is significantly different to the value found by Coibion (2010), but similar to that found by Reis (2009). A Bayesian model comparison will shed some light on which estimate is best.

	Alpha	Lambda	HV	HE	Covariance
Mean	0.07	0.55	325.01	91.61	0.00
SD	0.01	0.03	17.56	4.55	0.00

Table 7: Measure Parameters in Fixed Model

	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5
Mean	-0.00	0.87	-0.03	0.02	0.03
SD	0.00	0.05	0.05	0.02	0.05

Table 8: IV Parameters in Fixed Model

2.9 Bayesian Model Comparison

Bayesian model comparison can provide the relative plausibility of two models given our data. To calculate the probability of a time-varying attention model, I find the ratio between the marginal likelihood of the TVP model and the total marginal likelihood of both the TVP model and the fixed parameter model. Chib and Jeliazkov (2001) provide the framework for estimating the marginal likelihood when implementing a Markov chain Monte Carlo estimation.

The marginal likelihood of each model provides the foundation for a Bayesian model comparison:

$$p(Y|M_i) = \frac{p(Y|\theta)p(\theta)}{p(\theta|Y)}$$

The probability distribution of each model is equal to the marginal likelihood multiplied by the prior probability assigned to each model. In this comparison, I set the prior probabilities equal to each other. Thus, the relative probability of each model is the marginal likelihood of that model divided by the sum of all the marginal likelihoods:

$$p_i = \frac{p(M_i|Y)}{p(M_1|Y) + p(M_2|Y)}$$

The Metropolis-Hasting Sampler is used to sample from the posterior distribution, creating N samples. I calculate the likelihood function and the prior density at the median of the samples. Then to estimate the posterior ordinate we use the following equation:

$$\widehat{p(\tilde{\theta}|Y)} = \frac{\frac{1}{g}\sum_{g=1}^{G} \alpha(\theta^{[g]}, \tilde{\theta}) q(\tilde{\theta}|\theta^{[g]})}{\frac{1}{J}\sum_{j=1}^{J} \alpha(\tilde{\theta}, \theta^{[g]})},$$

in which $\alpha(\theta^{[g]}, \tilde{\theta})$ is the probability of accepting the median value instead of sample g and $q(\tilde{\theta}|\theta^{[g]})$ is the likelihood of the variation between the median and the sample. J draws are taken by random variation from the median, then $\alpha(\tilde{\theta}, \theta^{[g]})$ is the probability of accepting that draw instead of the median. The probabilities are then averaged to get the numerator and denominator to get an estimate of the ordinate.

A Bayesian model comparison between the fixed model and the TVP model favors the fixed parameter model. The probability placed on the fixed parameter model is 0.84. This analysis favors a model that excludes a time-varying parameter for attention but it cannot be ruled out statistically.

2.10 Conclusion

In conclusion, the estimation of the sticky information Phillips curve model with variation in the rate of attention provides valuable insights into parameter estimates and the dynamics of firms' attentiveness to macroeconomic information relevant to price setting.

The results from modeling time-varying attention reveal a declining trend in attention over the analyzed period, indicating potential economic implications for inflation volatility. Incorporating models proposed by Ball et al. (2003) suggests that policymakers may need to adjust their strategies in response to these time-varying estimates of attention.

Overall, these findings underscore the importance of considering time-varying attention in economic modeling and policymaking, as it provides valuable insights into the evolving nature of decision-making processes in response to macroeconomic information.

3 Bounded Rationality in an Endogenous Inattention Model

3.1 Introduction

This paper contributes to the existing literature by adding adaptive learning to an endogenous inattention model. The equilibrium must be learned using a recursive system. In this model with bounded rationality, the level of attentiveness is the probability of receiving optimal price forecasts using the most recently estimated forecasting model and data.

The yeoman farmer model uses an average of all the firm's prices to form the aggregate price, and the firms set their prices according to the optimal price. In the sticky information system, firms can set prices each period according to a forecast of the optimal price. The friction faced by firms is the availability of these forecasts. Therefore some firms set prices according to forecasts that do not utilize the most recent information. The rate of attentiveness is the probability of receiving a forecast path using the most recent information.

In a sticky information system with bounded rationality, the equilibrium coefficients of the forecasting model must be learned. Agents learn these values by gathering data and estimating a forecasting model recursively. The information, in this case, is a set of the realizations of a stochastic process.

An agent in the economic system calculates the rate of attentiveness by minimizing an expected loss function. The equilibrium level of information attentiveness can be found using adaptive learning. I replace rational expectations with a recursively estimated autoregressive moving average model to demonstrate this. The agent then uses the estimated coefficients to find an implied best level of attentiveness.

The adaption of complex tools for analysis has allowed companies to gather and process information more quickly. An increased rate of information processing is a possible explanation for a shift in the volatility of aggregate price marked by the end of the Great Moderation. This motivates the use of a permanent decrease in the cost of obtaining and processing information in my simulation studying transition dynamics.

The Great Moderation is a period during which the volatility in output and inflation decreased considerably in the United States and other major industrial countries. Reduced volatility in inflation makes financial planning easier for people. Lower output volatility stabilizes the job market by reducing the uncertainty faced by firms. This period would not last forever.

The end of the Great Moderation is marked by the bursting of the US housing bubble and the subsequent financial crisis. There is no consensus on this new era of macroeconomic volatility. Check and Piger (2021) find evidence in time series data that the Great Moderation continues with only a burst in volatility during the financial crisis. However, this finding did not include data from the Covid-19 epidemic.

3.2 Literature Review

This paper draws on a range of literatures, adaptive learning, information frictions and information processing costs.

This study builds on the information processing literature. Mankiw and Reis (2002) develop an exogenous system of information frictions to replace the sticky price assumption developed in work by Taylor (1980), Rotemberg (1982), and Calvo (1983). Sims (2003) developed a rational inattention model in which agents endogenously decide how much information to process. Sims' model uses the Shannon capacity limit mechanism to model how agents use up-to-date information with limited capacity. This model has a major advantage over the Mankiw and Reis (2002) model, in which agents endogenously decide what rates to absorb information. Branch et al. (2009) endogenize the rate of inattention in the Ball et al. (2005) model (a micro-founded version of Mankiw and Reis (2002)). This forms the major divergence in the information frictions literature. The Sims' model has agents constantly observing the most up to date information with some capacity for using the information and the Ball et al. (2005) and Branch et al. (2009), in which agents sometimes use extremely outdated information to make predictions. The former being more realistic and the latter a more tractable model.

Reis (2005) builds a micro-founded model that uses a basic assumption that agents face a cost of thinking through information and set prices based on this process. This model makes many contributions to justifying the foundations of time-contingent sticky information models. This simple model outperforms autoregressive models at forecasting inflation, and it is robust to the Lucas critique in that it can account moderately well for the inflation data under a different policy regime in the pre-war United States.

Carroll (2003) uses literature on epidemiology to construct a model in which agents are exposed to information. The foundational argument can be used to argue the distribution of information in the Mankiw and Reis (2002) model. However, Carroll develops an alternative model to study expectations instead of a model to study realized inflation. He estimates this model using the University of Michigan monthly survey of households and the Survey of Professional Forecasters.

Branch et al. (2009) endogenize the rate of inattention by modeling the choice of attentiveness as a symmetric Nash equilibrium. This theoretical step forward allows the authors to demonstrate a mechanism in which a more active monetary policy can decrease the rate of attentiveness, which in turn affects price volatility and perhaps output volatility. This offers a potential explanation for the decline and output and price volatility seen in the data during the great moderation. In addition, they show the effect of information costs on the rate of information acquisition.

The sticky information hypothesis generalizes easily to replace other frictions. Mankiw and Reis (2006) and Mankiw and Reis (2007) began work on a general equilibrium model to test hypothetical monetary policy questions. Reis (2009) estimates an alternative to the standard DSGE model using the sticky information assumption as the only friction in the savings market, the goods market, and the labor market. Carlstrom et al. (2015) study forward guidance at the zero lower bound, similar to the influential papers by Eggertsson and Woodford (2003) and Werning (2011). Kiley (2016) compares the sticky information assumption with the sticky price assumption when studying different puzzles in macroeconomics, using the standard model from Woodford (2003) containing a pricing equation (aggregate supply), a simple IS curve, and a range of policy rules. First, the author demonstrates that the sticky information model may provide a solution to the forward guidance puzzle studied by Levin et al. (2010), Laséen and Svensson (2011) and Negro et al. (2012). Next, the authors show conflicting conclusions on the welfare effects of fiscal expansion studied by Christiano et al. (2011) and Werning (2011). They also study the paradox of toil and the paradox of volatility

Coibion and Gorodnichenko (2015) consider both types of models in an empirical study. Using data from the US Survey of professional forecasters, the authors reject the null hypothesis of rational expectations. Their estimates also point to economically significant estimates of information rigidities. They go on to draw the interesting conclusion that the great moderation played a causal role in generating the great recession.

Sims (2003) has spawned a large series of articles. Sims (2010) surveys early work and Maćkowiak et al. (2022) surveys later work. Woodford (2001) applies this limited capacity hypothesis to the

Lucas island model developed in Phelps and Cagan (1984) and Lucas (1972b) to show it is possible to explain not only real effects of purely nominal disturbances but real effects that may persist for a substantial period of time. Paciello and Wiederholt (2014) use the rational inattention framework to study optimal policy when firms must choose how much attention to devote to aggregate conditions. Mackowiak and Wiederholt (2021) introduce rationally inattentive firms into the RBC model; this allows the model to reproduce the persistence of productivity shocks that is found in the data. The authors also find that rational inattention causes an increase in demand for labor and investment on the impact of a positive news shock. Ilut and Valchev (2020) find that an incomplete market model where agents have limited cognitive perceptions generates a more realistic distribution of beliefs and actions than the standard model. Afrouzi and Yang (2020) build on previous work Maćkowiak et al. (2018) to develop an efficient method to solve rational inattention models that can be used for quantitative work. They use this method to study the effect of different monetary policy regimes on the slope of the long and short-run Phillips curve. Mackowiak and Wiederholt (2009) find that the inattentive model allows prices to respond at different speeds to different shocks.

Forward-looking expectations have long been included in economic analysis; Thornton (1802) details the effect of forward-looking expectations on the current value of government-issued paper. Simple models that incorporate forward-looking expectations explicitly into a model are Hicks (1939) and Cagan (1956).

Current mainstream beliefs follow rational expectations Muth (1961). The key idea is that economic agents use all available information and understand the structure of the economy. Lucas and Sargent (1981) collect much of the early work. Rational expectations are an essential step in modeling the formation of expectations. However, rational expectations invoke a strong set of assumptions about the agents in the economy. While the econometrician must estimate parameters, agents are assumed to know their exact value. The rational expectations operator is replaced with a forecasting rule to bridge the gap between economists and agents. Sargent (1993) surveys a range of forecasting models, Evans and Honkapohja (2001) provide a foundation for using adaptive learning in macroeconomic models.

Evans and McGough (2020) provide a survey of the most recent work in adaptive learning. Application to monetary policy, the empirical fit of existing models, equilibrium when agents have a restricted perception of the economy, and equilibrium when agents have heterogeneous expectations. Bullard and Mitra (2002) highlight the need for learnable equilibria when setting policy rules—highlighting the coordination issue and how the failure of private sector expectation coordination can cause the economic system to diverge. Orphanides and Williams (2004) show that monetary policy that performs efficiently when agents form rational expectations and perform poorly when expectations are imperfect. Adaptive learning is also a plausible way to increase the empirical fit of models. Milani (2007) finds that a DSGE model estimated with adaptive learning no longer needs habits in consumption and price indexation to match the data. Using likehood-based Bayesian methods to estimate the model, he finds the values of habits and indexation are close to zero. Branch and McGough (2004) study equilibria when agents have heterogeneous beliefs.

Horaguchi (1996) shows that information processing costs in a one-shot game leads to boundedly rational behavior: the arrival at (not confess, not confess) in prisoner's dilemma. Dong et al. (2016) find a decrease in information processing costs reduces the delay in the effect of new information on stock prices.

3.3 Existing Model

The Ball et al. (2005) model uses the "yeoman farmer economy" used in Rotemberg and Woodford (1998) and Woodford (2003). The yeoman farmer economy allows us to derive pricing equations using micro-foundations; this allows for a utility-based measure of deadweight loss.

3.3.1 Market Structure

The economy is populated by a continuum of monopolistically competitive farmers. These farmers produce a good to sell to other agents and buy goods to consume. These farmers have risk aversion: $\sigma > 0$, and a marginal disutility of labor: $\psi > 0$. Farmer i's period t utility is given by the following equation:

$$U(C_{it}, Y_{it}) = \frac{C_{it}^{1-\sigma} - 1}{1-\sigma} - \frac{\hat{A}Y_{it}^{1+\zeta}}{1+\zeta}.$$

The elasticity of substitution between different goods, defined by (γ), is larger than one. The first term uses a constant elasticity substitution aggregator C_{it} for the agent's consumption of different goods:

$$C_{it} = [\int_0^1 (C_{it}^j)^{\frac{\gamma-1}{\gamma}} dj]^{\frac{\gamma}{\gamma-1}}.$$

The second term is the disutility from producing Y. The production function is defined by Y = AL, such that L is labor and A is technology. A follows a stochastic process and $\hat{A} = A^{-(1+\gamma)}$.

Agents maximize their expected discounted utility stream subject to their budget constraint (including proportional sales tax τ_t) by choosing consumption and labor. This utility maximization leads to the known log demand function y_{it} for good i at time t:

$$y_{it} = y_t - \gamma(p_{it} - p_t),$$

where p_i is the log price charged for good i. Not all farmer's wages will be the same since not all prices are set using the same information set. We assume there exists complete financial markets such that these risks are shared. Taking the demand we can solve for producer's optimal price to maximize their utility function:

$$p_{it}^* = p_t + \alpha y_t + u_t.$$

The parameter α is equal to $(\zeta + \sigma)/(1 + \gamma\zeta)$. In this model we choose to set the technology to a constant such that the natural output level is zero. There exists a proportional sales tax τ on all goods. These tax revenues fund equal lump sum transfers back to all agents. Assuming τ follows some stationary stochastic process is a convenient way to generate time-varying monopoly markups. The shock u is equal to $\frac{log((1-\bar{\tau})/(1-\tau))}{1+\psi\gamma}$. We focus on mark-up shocks modeled by AR(1) process with the structure:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad \text{with } 0 < \rho < 1.$$

3.3.2 Frictions

The Ball et al. (2005) model assumes that farmers are able to adjust prices each period. However, only a fraction of the farmers use the most recent information set to forecast the optimal price. Other firms must use a forecast based on outdated information. The current price level is a weighted average of all farmer's prices, therefore depends on past expectations of the current optimal price.

Formally, each period a fraction of farmers are selected to use an updated price path; these farmers are selected randomly from the continuum. The rest of the population must use outdated information to set prices. The fraction of agents that obtain new information is denoted λ; each firm has the same probability of being selected for new information. The result of this assumption is what we expect to see, an exponential distribution of firms over the information sets, with the largest group updating this period and the size of the group decreasing as information becomes more outdated.

Fraction λ of the farmers set a price according to time t information, where u_t is observable at time t and ρ is known. That implies that the fraction $\lambda(1-\lambda)$ set price according to time $t-1$ information, the fraction $\lambda(1-\lambda)^i$ set price according to time $t-i$ information. The aggregate price is given by:

$$p_t = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j E_{t-j}(p_t + \alpha y_t + u_t).$$

The authors derive the second order approximation of farmers' utility function:

$$-Var(y_t) - wE[Var_i(p_{it} - p_t)] + t.i.p., \tag{9}$$

where $w = \frac{\psi+\gamma^{-1}}{(\psi+\sigma)}$ and $t.i.p.$ stands for terms independent of policy. The second term is the variation of output across firms, this variation at the firm level causes loss in utility due to the departure from the efficient level in the labor supply.

In this case, we use a simple demand; the quantity theory approach. Nominal GDP is equal to the aggregate price level times output, presented here in logs:

$$y = m - p - e.$$

We study monetary policy with the objective to minimize equation (9), by controlling the quantity of money m. Here, w is considered to represent the relative importance the policy maker places on cross-sectional price variance as opposed to output variance. The first order conditions for minimization give us the following:

$$E_{t-1}y_t = -\alpha w E_{t-1}p_t.$$

Combine this equation with aggregate demand equation gives a policy rule based on a fixed quantity of money:

$$m_t = (1 - \alpha w)E_{t-1}p_t.$$

Assuming policy makers follow this rule, the rational expectations equilibrium paths for optimal price, aggregate price and output are as follows:

$$p_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j},$$

$$y_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j},$$

$$p_t^* = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}.$$

3.3.3 Nash Equilibrium

Above, we considered λ to be exogenous, this is a convenient assumption, but optimizing agents would adjust their rate of inattention in response to changes in the economy. Branch et al. (2009) develop a model in which λ is the result of a symmetric simultaneous Nash equilibrium, successfully combining elements of time dependence and state dependence into a single model. The firms in their model minimize a loss function to choose λ; this allows information updating to be affected by other factors in the model economy.

All firms are choosing λ independently at the same time. From any specific firm's perspective, there exists an economy-wide probability $\bar{\lambda}$ chosen by other firms. We arrive at a symmetric Nash equilibrium when the best choice of λ equals to the economy wide choice $\bar{\lambda}$.

$p_t(\bar{\lambda})$ and $y_t(\bar{\lambda})$ are the equilibrium price and output given that all agents have used $\bar{\lambda}$ for the history of the economy. In order to solve for equilibrium, Branch et al. (2009) first redefine the optimal price p_t^* in terms of $\bar{\lambda}$:

$$p_t^*(\bar{\lambda}) = p_t(\bar{\lambda}) + \alpha y_t(\bar{\lambda}) + u_t$$

We must proceed carefully here, firms are not choosing a length of time in which to update their information set. They are choosing a probability to update. This process can be thought of in the following way. Each period every firm flips a coin, this coin says "update" (U) on one side and "don't update" (DU) on the other. The probability the coin lands on update is the firms choice of λ. Let S be a function such that $S_t(U) = 1$ and $S_t(DU) = 0$, this function is a random variable. Let $\hat{p}_t(\lambda)$ be the price set by a firm at time t that is determined by the random variable in the following way:

$$\hat{p}_t(\lambda) = \begin{cases} p_t^*(\bar{\lambda}) & \text{if } s_t = 1 \\ E_{t-k}p_t^*(\bar{\lambda}) & \text{if } s_{t-k} = 1, s_{t-k+1} = 0, ..., s_t = 0. \end{cases}$$

The expected profit loss is proportional to the expected squared error between the price set by the firm $\hat{p}_t(\lambda)$ and the actual optimal price $p_t^*(\bar{\lambda})$:

$$L(\lambda, \bar{\lambda}) = E[(\hat{p}_t(\lambda) - p_t^*(\bar{\lambda}))^2].$$

If it were free to adjust price, firms would choose λ to equal one. This would reduce the loss to zero. Sticky information is motivated by a cost of data gathering and reoptimization. To approximate the cost agents face, $C\lambda^2$ is used.

Now, using this cost of information accrual and the loss firms face when setting λ we have the following minimization problem:

$$T(\bar{\lambda}) = \arg\ \min_{\lambda}[L(\lambda, \bar{\lambda}) + C\lambda^2].$$

$T(\bar{\lambda})$ is a best-response function, it maps the economy choice of lambda to the optimal choice by the firm. We arrive at a symmetric Nash equilibrium when $\lambda^* = T(\lambda^*)$.

Branch et. al. (2009) show in equilibrium this loss is equivalent to:

$$L(\lambda, \bar{\lambda}) = (1-\lambda)\bar{\psi}_0 - \lambda \sum_{j=1}^{\infty}(1-\lambda)^j \bar{\psi}_j.$$

3.4 Incorporating Learning into the Existing Model

Branch, Carlson, Evans and McGough studied what value of λ agents would choose if the equilibrium values for MA process that governs optimal price and aggregate price are known, this is the rational expectations equilibrium (REE). Expanding the existing model by adding adaptive learning will allow us to study the stability of equilibrium, study structural changes and potentially help with the empirical fit of the model. Structural changes are convenient to study using adaptive learning because the equilibrium values of the price coefficients are a function of all past realization of λ, solving for an infinite number of coefficients using an infinite history of lambdas would be very challenging for the agents in the model and for the researcher.

In REE both the firms and the policy maker are making forecasts of a variable that follows an MA(∞) process. Though the moving-average process is a basic time series model, estimation via maximum likelihood is intractable. I follow a similar method to the classical two step process outlined in Durbin (1959). First, both types approximate the true process using an ARMA(1,5) model estimated via recursive least squares. Next, we can find MA expression of the estimated ARMA model.

I will implement adaptive learning at the aggregate level; all agents must update with the same probability λ_t. It is convenient to assume a consultant is calculating optimal price paths, λ_t and randomly selecting firms to give a price path estimated using the most up to date coefficients. This consultant is collecting data and running a regression to estimate ARMA(1,5) forecasting model. The key assumption in the Ball et al. (2005) model is that firms follow a pre-set price path. The consultant uses the estimated model to create this price path for firms. Using the estimated MA coefficients she can calculate λ using the loss function, this λ is calculated for all the firms by the consultant.

It is also assumed that both the consultant and policy makers can observe markup shocks and know the form of the shocks and the values in the process. This assumption gives the agents access to the history of ε. Given this assumption, agents are able to run a regression to estimate coefficients using collected data.

3.4.1 Policy

For any particular period t, policy makers implement optimal policy; $\hat{m}_t$. This policy take a period to implement. Therefore, the policy maker must forecast p_t using the prior information set $t-1$. Aggregate price follows an $MA(\infty)$ process in REE, here policy makers will use an ARMA(1,5) model. This model has a single autoregressive coefficient to capture the long term effects of shocks and five moving average coefficients for the most recent shocks in the series, this allows us to accurately approximate the effect of all shocks:

$$\hat{E}_{t-1}p_t = a_{t-1}p_{t-1} + b_{1,t-1}\varepsilon_{t-1} + b_{2,t-1}\varepsilon_{t-2} + b_{3,t-1}\varepsilon_{t-3} + b_{4,t-1}\varepsilon_{t-4} + b_{5,t-1}\varepsilon_{t-5}.$$

This estimation gives us the following policy:

$$\hat{m}_t = (1-\alpha\omega)\hat{E}_{t-1}p_t$$

or

$$\hat{E}_{t-1}p_t = c'_{p,t-1}x_{p,t},$$

s.t.

$$c'_{p,t-1} = [a_{t-1}, b_{1,t-1}, b_{2,t-1}, b_{3,t-1}, b_{4,t-1}, b_{5,t-1}],$$

$$x'_{p,t} = [p_{t-1}, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \varepsilon_{t-4}, \varepsilon_{t-5}],$$

that is estimated via recursive least squares,

$$c_{p,t} = c_{p,t-1} + \gamma_t R^{-1}_{p,t}x_{p,t-1}(p_t - c'_{p,t-1}x_{p,t-1})$$

$$R_{p,t} = R_{p,t-1} + \gamma_t(x_{p,t-1}x'_{p,t-1} - R_{p,t-1}).$$

Coefficients are estimated recursively each period when new data becomes available. $c_{p,t}$ is a vector of the ARMA(1,5) coefficients, $R_{p,t}$ is the second moment matrix of the regressors. γ_t is the gain sequence, under RLS this gain is equal to t^{-1}, because all past data points count equally. A fixed gain sequence will weight more recent observation more than older observations, this is useful if there are structural changes in the system.

3.4.2 Firms

The consultant will use an ARMA(1,5) model to approximate an MA(∞) process. The consultant forms expectation $\hat{E}_t p^*_{t+k}$. When $k = 0$, we have the following forecast rule:

$$p^*_t = \alpha p^*_{t-1} + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_3 \varepsilon_{t-3} + \beta_4 \varepsilon_{t-4}$$

or

$$\hat{E}_t p^*_t = c'_{c,t-1} x_{c,t},$$

s.t.

$$c'_{c,t} = [\alpha_t, \beta_{0,t-1}, \beta_{1,t-1}, \beta_{2,t-1}, \beta_{3,t-1}, \beta_{4,t-1}],$$

$$x'_{c,t} = [p^*_{t-1}, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \varepsilon_{t-4}],$$

that is estimated via recursive least squares,

$$c_{c,t} = c'_{c,t-1} + \gamma_t R^{-1}_{c,t} x_{c,t-1}(p^*_{t-1} - c'_{c,t-1} x_{c,t-1}),$$

$$R_{c,t} = R_{c,t-1} + \gamma_t (x_{c,t-1} x'_{c,t-1} - R_{c,t-1}).$$

This model has an optimal price for each period: p^*_t . Firms would like to set their price equal to this optimal price each period. A consultant must forecast the optimal price given a fixed information set. The consultant gives this forecast path of the optimal price to selected firms. The firms then use this price path to set their price each period. The forecasting technique of an ARMA model is simple. First, a forecast is made using the available information:

$$\hat{E}_t p^*_t = \alpha_t p^*_{t-1} + \beta_{t,0} \varepsilon_t + \beta_{t,1} \varepsilon_{t-1} + \beta_{t,2} \varepsilon_{t-2} + \beta_{t,3} \varepsilon_{t-3} + \beta_{t,4} \varepsilon_{t-4}.$$

To forecast past p^*_{t+1}, we replace the realized value of p^*_{t-1} with the predicted value $\hat{E}_t p^*_t$ and replace ε_t with the expected value: 0 and shift the realized shocks down in the following manner:

$$\hat{E}_t p^*_{t+1} = \alpha_t \hat{E}_t p^*_t + \beta_{t,0} 0 + \beta_{t,1} \varepsilon_t + \beta_{t,2} \varepsilon_{t-1} + \beta_{t,3} \varepsilon_{t-2} + \beta_{t,4} \varepsilon_{t-3}.$$

This technique can be continued to form an optimal price path:

$$\hat{E}_t p^*_{t+2} = \alpha_t \hat{E}_t p^*_{t+1} + \beta_{t,1}0 + \beta_{t,2}0 + \beta_{t,3}\varepsilon_t + \beta_{t,4}\varepsilon_{t-1} + \beta_{t,5}\varepsilon_{t-2}$$
$$\hat{E}_t p^*_{t+3} = \alpha_t \hat{E}_t p^*_{t+2} + \beta_{t,1}0 + \beta_{t,2}0 + \beta_{t,3}0 + \beta_{t,4}\varepsilon_t + \beta_{t,5}\varepsilon_{t-1}$$
$$\hat{E}_t p^*_{t+3} = \alpha_t \hat{E}_t p^*_{t+3} + \beta_{t,1}0 + \beta_{t,2}0 + \beta_{t,3}0 + \beta_{t,4}0 + \beta_{t,5}\varepsilon_t$$
$$\hat{E}_t p^*_{t+3} = \alpha_t \hat{E}_t p^*_{t+4} + \beta_{t,1}0 + \beta_{t,2}0 + \beta_{t,3}0 + \beta_{t,4}0 + \beta_{t,5}0$$
$$\vdots$$

3.4.3 Optimal λ

The consultant must now use the conditional expected squared error between the price set by the firm $\hat{p}_t(\lambda)$ and the actual optimal price p^*_t to find optimal λ_t:

$$L(\lambda_t) = E_t[(\hat{p}_t(\lambda_t) - p^*_t)^2],$$

$$\lambda_t = \arg\min_{\lambda_t}[L(\lambda_t) + C\lambda_t^2].$$

We can find an infinite-order MA model representation of the recursively estimated ARMA model. The consultant uses an ARMA(1,k) model and must find the MA equivelant: $\hat{\theta}$ for the calculation of the optimal lambda. Explicitly the model has the following form:

$$Ep^*_t = coef_{AR} * p^*_{t-1} + coef_1 * \varepsilon_t +, ..., + coef_k * \varepsilon_{t-k}.$$

We can solve for the implied values:

$$\hat{\theta_0} = coef_1.$$

$$\hat{\theta_1} = coef_2 + coef_{AR} * coef_1.$$

$$\hat{\theta_2} = coef_3 + coef_{AR} * coef_2 + coef^2_{AR} * coef_3.$$

Now, we can summarize the pattern for coeffecient i such that i $\in \{0, 1, .., k-1\}$

$$\hat{\theta}_i = \sum_{j=0}^{i} coef_{AR}^{j} * coef_{i-j+1},$$

and for coeffecient i such that $i \geq k$

$$\hat{\theta}_i = \sum_{j=0}^{k} coef_{AR}^{j+(i-k)} * coef_{k-j+1}.$$

Using the estimated MA values, we can follow the steps in Branch et al. (2009) to obtain the following:

$$L(\lambda_t) = (1-\lambda_t)\hat{\psi}_{t,0} - \lambda_t \sum_{j=1}^{\infty}(1-\lambda_t)^j \hat{\psi}_{t,j},$$

s.t.

$$\hat{\psi}_{t,j} = \sigma_{\varepsilon}^2 \sum_{k=j}^{\infty} \hat{\theta}_{t,k}^2.$$

We must assume that firms are told how often to update information by the consultant. This means that firms have access (second hand) to the most up-to-date coefficients. Since firms are choosing a probability to update, they cannot receive a 1 step ahead prediction for each period without cost. The cost is "paid" each period given probability λ_t. Another critical assumption made is that the consultant has perfect knowledge of σ_{ε}^2.

3.4.4 Economic System

Aggregate price is now a function of a history of λ. Denote Λ_t as the set that contains the history of λ_t s.t.

$$\lambda_{t-i} \in \Lambda_t \qquad \forall i \geq 0.$$

Now we can compute aggregate and optimal price:

$$p_t(\Lambda_t) = \sum_{j=0}^{\infty} \lambda_{t-j} * \Pi_{i=0}^{j-1}(1-\lambda_{t-i}) * \hat{E}_{t-j}p_t^*,$$

$$p_t^*(\Lambda_t) = (1-\alpha)p_t(\Lambda_t) + \alpha m_t + u_t.$$

The recursive causal ordering that describes the model economy:

$$\hat{E}_{t-1}p_t = a_{t-1}p_{t-1} + b_{1,t-1}\varepsilon_{t-1} + b_{2,t-1}\varepsilon_{t-2} + b_{3,t-1}\varepsilon_{t-3} + b_{4,t-1}\varepsilon_{t-4} + b_{5,t-1}\varepsilon_{t-5},$$

$$m_t = (1-\alpha\omega)E_{t-1}p_t,$$

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

$$\hat{E}_t p^*_{t+k} \qquad \forall k \geq 0,$$

$$\lambda_t = \text{argmin}_{\lambda_t \in [0,1]}[L(\lambda_t) + C\lambda_t^2],$$

$$p_t(\lambda_t, \Lambda_{t-1}) = \sum_{j=0}^{\infty} \lambda_{t-j} * \Pi_{i=0}^{j-1}(1-\lambda_{t-i}) * \hat{E}_{t-j}p^*_t,$$

$$p^*_t(\lambda_t, \Lambda_{t-1}) = (1-\alpha)p_t(\lambda_t, \Lambda_{t-1}) + \alpha m_t + u_t.$$

3.5 Calibration

The primary goal of calibration is to provide a qualitative example. However, the initial values are chosen to roughly match variance in price and output to those observed in the data. The model parameters are as follows; $C = 5$, $\alpha = 0.1$, $\sigma^2_\varepsilon = 0.1$, $\sigma^2_e = 0$, and $\rho = 0.85$. I examine the dynamics when C shifts to 4 from 5, holding other parameters fixed.

The magnitude of this shift has a heterogenous effect on the costs of updating at different rates. This is due to the parametric cost function: $\lambda^2 C$. Below we can see that the cost of never updating information remains constant. While the cost of updating with a probability of 1 decreases by 20%.

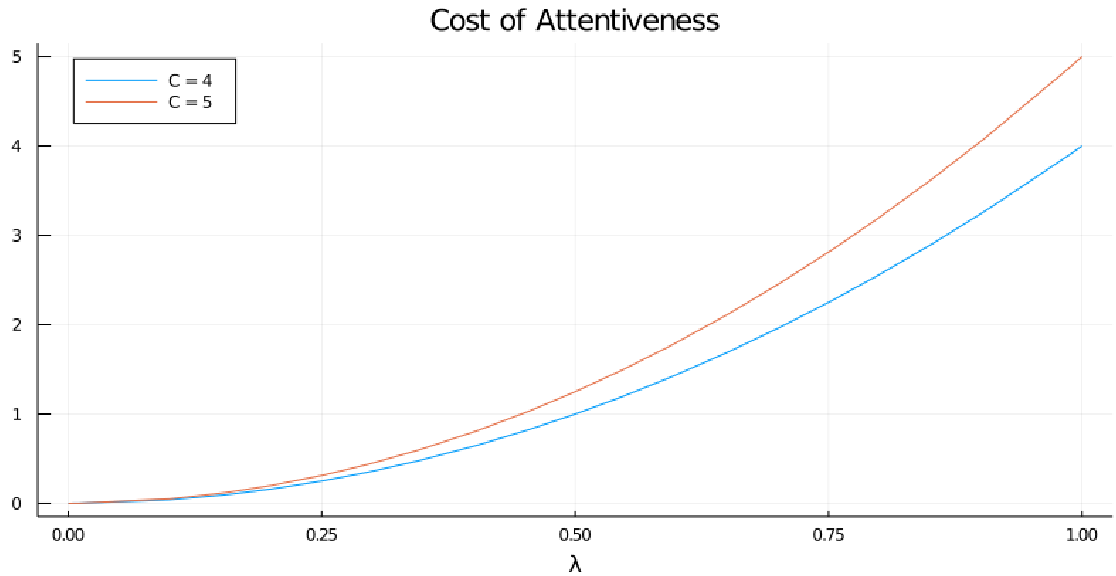

Figure 5: Shift in Cost

3.6 Stability and Restricted Perceptions Equilibrium

To test the stability of the chosen calibration, I simulate the system and check the outcome visually for convergence. This result will be studied using a decreasing gain: $\gamma_t = t^{-1}$. In the initial calibration with $C = 5$ shown in figure 1 below, λ_t converges to approximately 0.10. In the secondary calibration with $C = 4$ shown in figure 2, λ_t converges to approximately 0.14. A visual inspection indicates stability at both equilibrium, this motivates a review of the transition between.

	RPE Coefficients at ARMA(1, 5)
RPE Coefficients for Firms	[0.852 1.082 -0. 0.004 0.004 0.002]
RPE Coefficients for Policy Maker	[0.918 0.067 0.046 0.034 0.018 0.015]

Table 9: Equilibrium values at C= 5

	RPE Coefficients at ARMA(1, 5)
RPE Coefficients for Firms	[0.874, 1.119, 0.001, 0.007, 0.007, 0.004]
RPE Coefficients for Policy Maker	[0.929, 0.099, 0.069, 0.052, 0.029, 0.024]

Table 10: Equilibrium values at C= 4

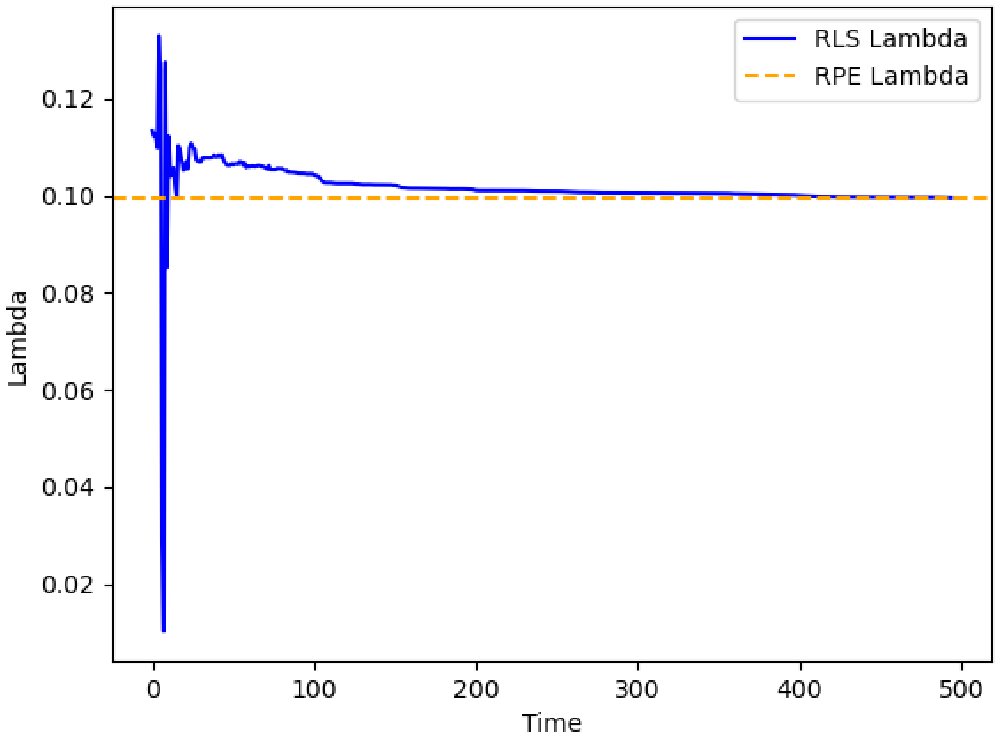

Figure 6: Stability with C set to 5

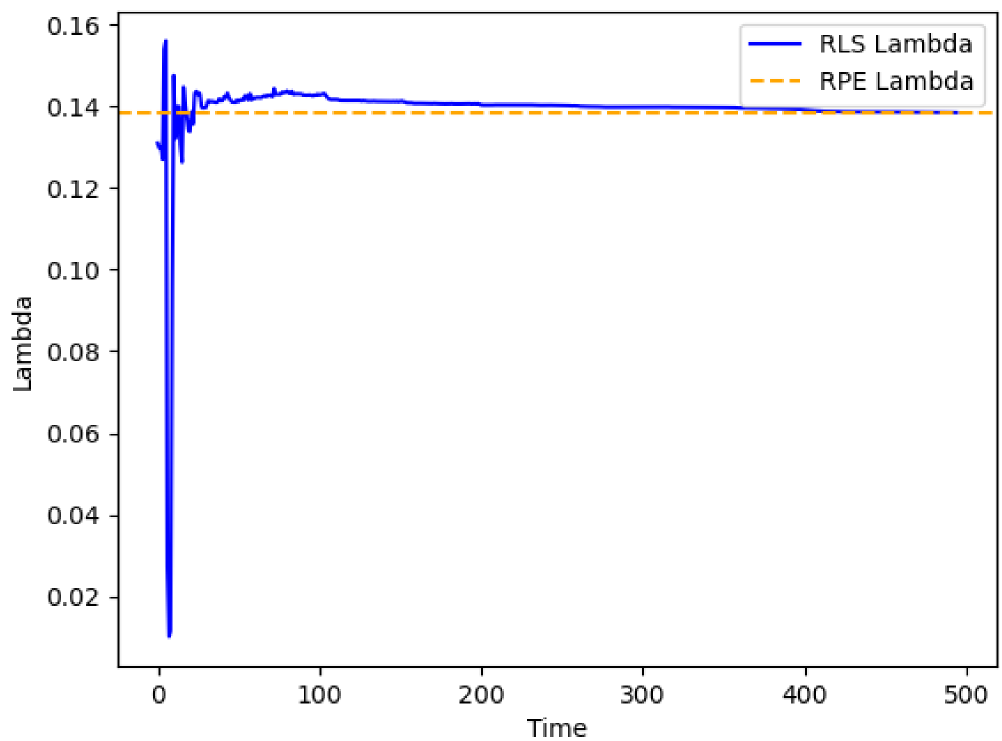

Figure 7: Stability with C set to 4

To investigate the robustness of my analysis, I repeated it using an AR(1) forecasting rule. However, this approach proves to be too volatile for the system, and convergence is not consistent when employing an AR(1) forecasting rule. Introducing a single MA coefficient allows for convergence into an equilibrium.

3.7 Transition Dynamics

To examine the transition between Nash equilibrium, I set C to 4 for an initial period length of 300 periods. At that time, C shifts to 5. This causes an immediate jump in λ. A increase in cost makes choosing a smaller λ immediately a better choice. Subsequently, λ slowly converges to the new Nash equilibrium value as the ARMA coefficients converge to their new equilibrium value. This result is robust to a range of gain selections.

First, I test the standard decreasing gain sequence t^{-1}. This gain does a good job converging to the initial equilbrium value. We see a sizeable immediate jump in the value of λ, but the algorithm fails to converge to the equilibrium value quickly.

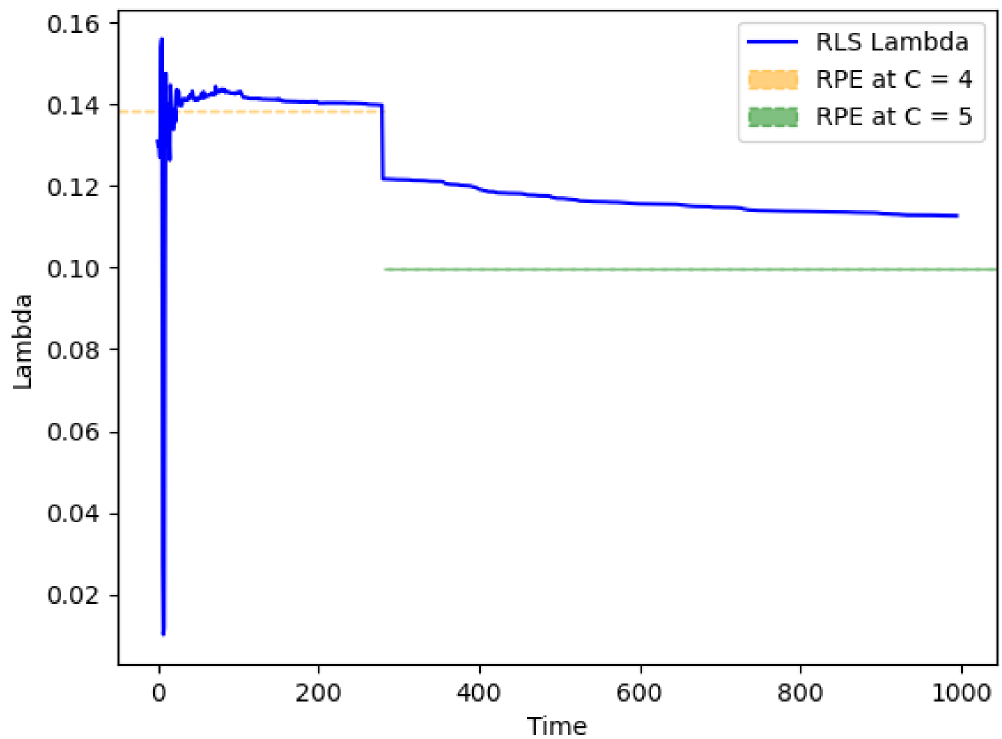

Figure 8: Transition Standard Decreasing Gain

A learning algorithm using a modified decreasing gain of $t^{-0.8}$ is able to pick up better on changes in the economy. However, it doesnt accuratly converge to the intial equilbrium value. After the the shift in model parameters it converges quickly towards the new equilibrium.

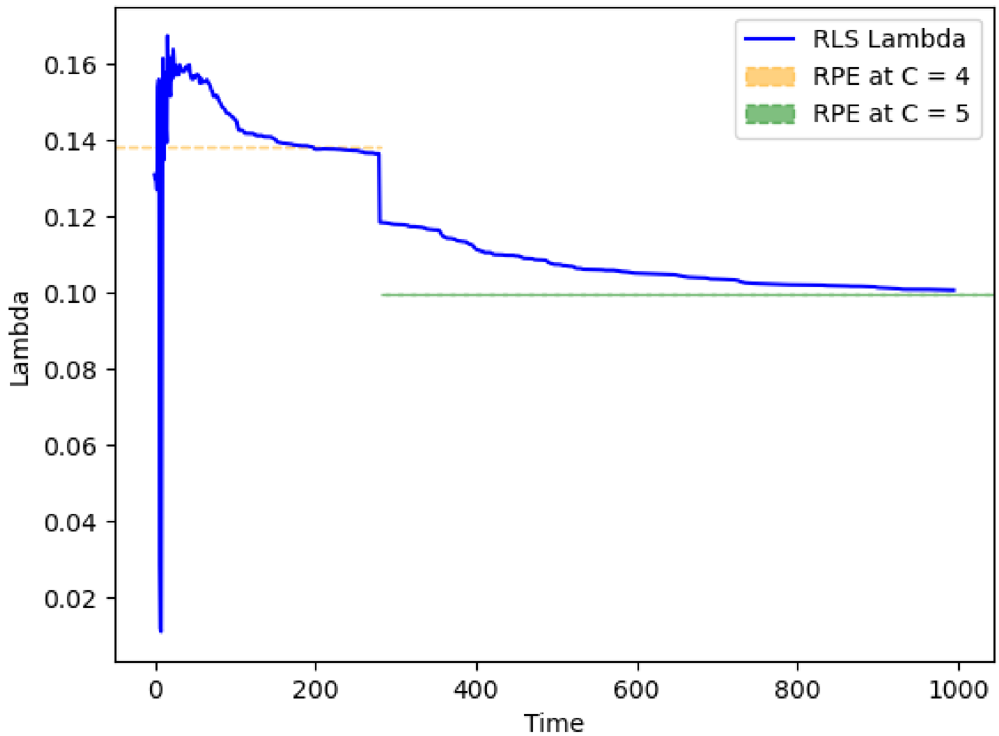

Figure 9: Transition Modified Decreasing Gain

3.8 Convergence to REE

A forecasting rule for which the RPE is close to the REE would justify the use of studying this system using the REE values. To do this, I graph a selection of $ARMA(1, k)$ models. In addtion to the ARMA values, I have plotted the REE values to study the convergence of our PLM. The value of $k = 10$ converges closely to the REE value.

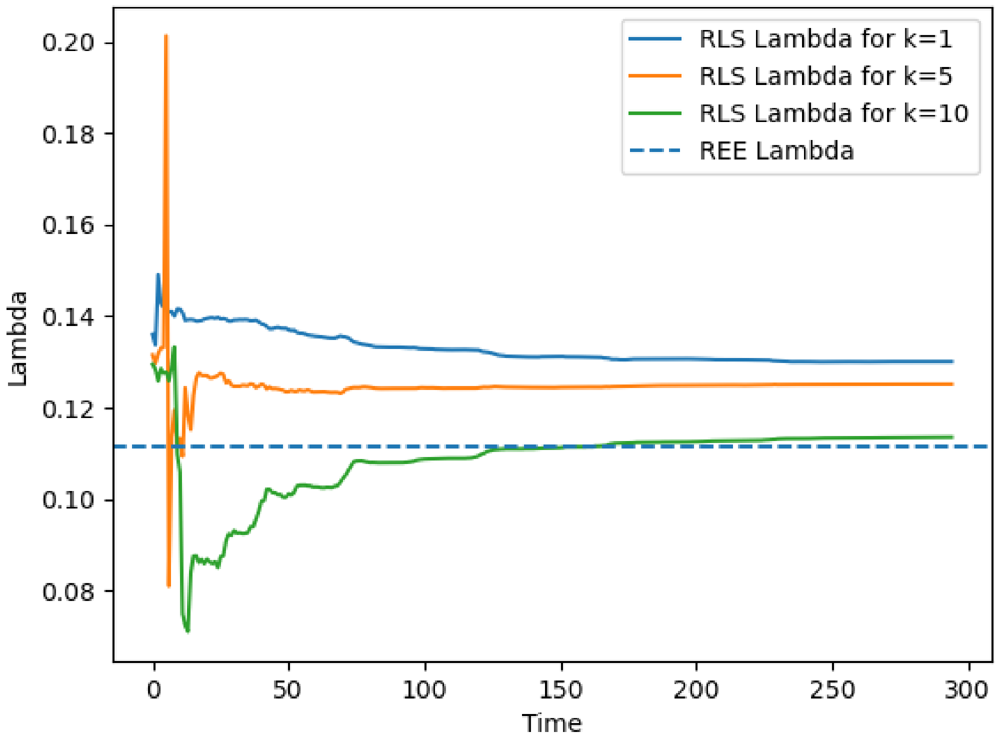

Figure 10: Transition Modified Decreasing Gain

Model	Equilibrium Lambda
1	0.130030
5	0.125100
10	0.113555
REE	0.111600

Table 11: RPE values for different k

3.9 Conclusion

This study focuses on the addition of adaptive learning to an endogenous attention model. These modifications enable us to explore stability and transitions to equilibrium. I find that the equilibrium is stable under learning and that the rational expectations equilibrium can be approximated with an ARMA(1,10) model.

4 References

Afrouzi, H. and Yang, C. (2020). Dynamic rational inattention and the phillips curve. *Unpublished working paper. Columbia University and UT Austin.*

Ball, L., Gregory Mankiw, N., and Reis, R. (2005). Monetary policy for inattentive economies. *Journal of Monetary Economics*, 52(4):703–725.

Ball, L., Mankiw, N. G., and Reis, R. (2003). Monetary policy for inattentive economies. Working Paper 9491, National Bureau of Economic Research.

Branch, W. and McGough, B. (2004). Multiple Equilibria in Heterogeneous Expectations Models. *The B.E. Journal of Macroeconomics*, 4(1):1–16.

Branch, W. A., Carlson, J., Evans, G. W., and McGough, B. (2009). Monetary policy, endogenous inattention and the volatility trade-off. *The Economic Journal*, 119(534):123–157.

Bullard, J. and Mitra, K. (2002). Learning about monetary policy rules. *Journal of Monetary Economics*, 49(6):1105–1129.

Cagan, P. (1956). *The monetary dynamics of hyperinflation.* Chicago: University of Chicago Press.

Calvo, G. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Carlstrom, C. T., Fuerst, T. S., and Paustian, M. (2015). Inflation and output in new keynesian models with a transient interest rate peg. *Journal of Monetary Economics*, 76:230–243.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118(1):269–298.

Check, A. and Piger, J. (2021). Structural breaks in u.s. macroeconomic time series: A bayesian model averaging approach. *Journal of Money, Credit and Banking*, 53(8):1999–2036.

Chib, S. and Jeliazkov, I. (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American Statistical Association*, 96(453):270–281.

Christiano, L., Eichenbaum, M., and Rebelo, S. (2011). When Is the Government Spending Multiplier Large? *Journal of Political Economy*, 119(1):78–121.

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

Coibion, O. (2010). Testing the sticky information phillips curve. *The Review of Economics and Statistics*, 92(1):87–101.

Coibion, O. and Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78.

Dong, Y., Li, O. Z., Lin, Y., and Ni, C. (2016). Does information-processing cost affect firm-specific information acquisition? evidence from xbrl adoption. *Journal of Financial and Quantitative Analysis*, 51(2):435–462.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, 46(3/4):306–316.

Eggertsson, G. B. and Woodford, M. (2003). The Zero Bound on Interest Rates and Optimal Monetary Policy. *Brookings Papers on Economic Activity*, 34(1):139–235.

Evans, G. W. and Honkapohja, S. (2001). *Learning and Expectations in Macroeconomics*. Princeton University Press.

Evans, G. W. and McGough, B. (2020). Adaptive learning in macroeconomics. In *Oxford Research Encyclopedia of Economics and Finance*.

Fernández-Villaverde, J., Rubio-Ramírez, J. F., Cogley, T., and Schorfheide, F. (2007). How structural are structural parameters? [with comments and discussion]. *NBER Macroeconomics Annual*, 22:83–167.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Hicks, J. R. (1939). The foundations of welfare economics. *The Economic Journal*, 49(196):696–712.

Horaguchi, H. (1996). The role of information processing cost as the foundation of bounded rationality in game theory. *Economics Letters*, 51(3):287–294.

Ilut, C. L. and Valchev, R. (2020). Economic agents as imperfect problem solvers. Working Paper 27820, National Bureau of Economic Research.

Kiley, M. T. (2016). Policy paradoxes in the new keynesian model. *Review of Economic Dynamics*, 21:1–15.

Kim, C.-J. (2006). Time-varying parameter models with endogenous regressors. *Economics Letters*, 91(1):21–26.

Kim, Y. and Kim, C.-J. (2011). Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *The Econometrics Journal*, 14(3):487–497.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198.

Laséen, S. and Svensson, L. E. (2011). Anticipated Alternative policy Rate Paths in Plicy Simulations. *International Journal of Central Banking*, 7(3):1–35.

Levin, A., López-Salido, D., Nelson, E., and Yun, Y. (2010). Limitations on the Effectiveness of Forward Guidance at the Zero Lower Bound. *International Journal of Central Banking*, 6(1):143–189.

Lucas, R. E. (1972a). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2):103–124.

Lucas, R. E. (1972b). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2):103–124.

Lucas, R. E. and Sargent, T. (1981). *Rational Expectations and Econometric Practice: Volume 1*. University of Minnesota Press, ned - new edition edition.

Mackowiak, B. and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803.

Mackowiak, B. and Wiederholt, M. (2021). Rational inattention and the business cycle effects of productivity and news shocks. *Unpublished working paper. European Central Bank and Sciences Po.*

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve*. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Mankiw, N. G. and Reis, R. (2006). Pervasive stickiness. *American Economic Review*, 96(2):164–169.

Mankiw, N. G. and Reis, R. (2007). Sticky Information in General Equilibrium. *Journal of the European Economic Association*, 5(2-3):603–613.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2018). Dynamic rational inattention: Analytical results. *Journal of Economic Theory*, 176:650–692.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2022). Rational expectations and the dynamics of hyperinflation. *Forthcoming Journal of Economic Literature.*

Milani, F. (2007). Expectations, learning and macroeconomic persistence. *Journal of Monetary Economics*, 54(7):2065–2082.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3):315–335.

Negro, M. D., Giannoni, M., and Patterson, C. (2012). The forward guidance puzzle. Technical report.

Olea, J. L. M. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics*, 31(3):358–369.

Orphanides, A. and Williams, J. (2004). Imperfect knowledge, inflation expectations, and monetary policy. In *The inflation-targeting debate*, pages 201–246. University of Chicago Press.

Paciello, L. and Wiederholt, M. (2014). Exogenous information, endogenous information, and optimal monetary policy. *The Review of Economic Studies*, 81(1 (286)):356–388.

Peersman, G. and Pozzi, L. (2004). Determinants of consumption smoothing. Working papers of faculty of economics and business administration, ghent university, belgium, Ghent University, Faculty of Economics and Business Administration.

Phelps, E. S. and Cagan, P. (1984). *The trouble with "rational expectations" and the problem of inflation stabilization*, page 31–46. Cambridge University Press.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852.

Reis, R. (2005). Inattentive producers. Working Paper 11820, National Bureau of Economic Research.

Reis, R. (2009). A sticky-information general-equilibrium model for policy analysis. Working Paper 14732, National Bureau of Economic Research.

Rotemberg, J. J. (1982). Sticky prices in the united states. *Journal of Political Economy*, 90(6):1187–1211.

Rotemberg, J. J. and Woodford, M. (1998). An optimization-based econometric framework for the evaluation of monetary policy: Expanded version. Working Paper 233, National Bureau of Economic Research.

Sargent, T. (1993). *Bounded rationality in macroeconomics.* Clarendon Press. Includes bibliographical references (pages 171-180) and indexes.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Sims, C. (2010). Rational inattention and monetary economics. In Friedman, B. M. and Woodford, M., editors, *Handbook of Monetary Economics*, volume 3, chapter 04, pages 155–181. Elsevier, 1 edition.

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.

Taylor, J. B. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy*, 88(1):1–23.

Thornton, H. (1802). *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain.* George Allen and Unwin.

Werning, I. (2011). Managing a liquidity trap: Monetary and fiscal policy. Working Paper 17344, National Bureau of Economic Research.

Woodford, M. (2001). Imperfect Common Knowledge and the Effects of Monetary Policy. NBER Working Papers 8673, National Bureau of Economic Research, Inc.

Woodford, M. (2003). *Interest and Prices*. Princeton University Press, Princeton.

Woodford, M. and Walsh, C. E. (2005). Interest and prices: Foundations of a theory of monetary policy. *Macroeconomic Dynamics*, 9(3):462–468.

Young, A. (2004). Labor's Share Fluctuations, Biased Technical Change, and the Business Cycle. *Review of Economic Dynamics*, 7(4):916–931.

5 Appendix of Chapter 1

Hyperparameters

	beta	σ_ν	λ_0	σ_ω	M	σ_ξ	ρ
start	0	0	3	0	0	0	-1
end	2	2	5	0.2	2	2	1

Table 12: Metropolis-Hastings Priors Model 1

Hyperparameters

	beta	σ_ν	λ_0	σ_ω	M	σ_ξ	ρ
start	0	0	0	0	0	0	-1
end	2	2	2	2	2	2	1

Table 13: Metropolis-Hastings Priors, Role of Instruments