Online Appendix for "Learning and Subjective Expectation Formation: A Recurrent Neural Network Approach"

Chenyu(Sev) Hou

May 10, 2023

Appendices

C Examples

C.1 Example: Standard Noisy Information Model in Generic Learning Framework

In this subsection I take the standard noisy information model as an example and show how it can be represented by the Generic Learning Framework. The purpose of this example are three folds. First it gives an example of essential elements in the Generic Learning Model including hidden states $\Theta_{i,t}$, Average Structural Function and the transformed dynamic system (6) in the context of a familiar learning model. Secondly it illustrates how RNN performs in approximating the ASF (in this case linear) and estimating marginal effect without knowledge of the exact functional form of learning model. Lastly as I consider a special case when the expectation formation structure is still linear but OLS is mis-specified and show the performance of RNN in estimating the average marginal effect. This exercise illustrates the possible improvement in using RNN even in a linear case.

*I'm grateful to Jesse Perla, Paul Schrimpf, Paul Beaudry, Michael Devereux, and Amartya Lahiri for their invaluable guidance and support on this project. I thank Vadim Marmer, Fabio Milani, Monika Piazzesi, Henry Siu, Hassan Afrouzi, and many others for their insightful comments. I'm also thankful to the computational support of Compute Canada (www.computecanada.ca). All the remaining errors are mine.

†Chenyu Hou: Vancouver School of Economics, University of British Columbia. Email: sevhou1989@gmail.com

Data Generating Process: Consider agents want to predict inflation one period from now denoted as $\pi_{i,t+1|t}$. At time t, they can observe two signals $\{\pi_{i,t}, s_{i,t}\}$. There are two latent variables $\{\pi_t, L_t\}$ that they need to make inference of to form expectation of inflation. Represent the Actual Law of Motion as a Gaussian Linear State Space Model:

$$\begin{bmatrix} \pi_t \\ L_t \end{bmatrix} \equiv X_t = \boldsymbol{A} X_{t-1} + \epsilon_t \tag{42}$$

Where $\boldsymbol{A}$ describes how latent states X_t evolves along time, ϵ_t is i.i.d shock each period. Assume for simplicity the agent's Perceived Law of Motion is the same as (42). Agents do not observe X_t directly, instead they observe a noisy signals about it. Their observational equation is:

$$\begin{bmatrix} \pi_{i,t} \\ s_{i,t} \end{bmatrix} \equiv O_{i,t} = \boldsymbol{G} X_t + \nu_{i,t} \tag{43}$$

Both shock ϵ_t and $\nu_{i,t}$ are i.i.d and follow normal distribution with covariance matrix R and Q:

$$\epsilon_t \sim N(0, R) \quad \nu_{i,t} \sim N(0, Q)$$

This describes the standard noisy information model with two latent states. They use a stationary Kalman Filter to form prediction of the latent variable $X_{i,t+1|t}$, where $\boldsymbol{K}$ is the Kalman Gain.

$$\begin{bmatrix} \pi_{i,t+1|t} \\ L_{i,t+1|t} \end{bmatrix} \equiv X_{i,t+1|t} = \boldsymbol{A}(X_{i,t|t-1} + \boldsymbol{K}(O_{i,t} - \boldsymbol{G} X_{i,t|t-1})) \tag{44}$$

The Generic Learning Formulation The stationary Kalman Filter is a special case of Generic Learning Model. First notice the i.i.d error $\nu_{i,t}$ satisfies assumption 2. The expectation is also formed by filtering step and updating step:

$$X_{i,t|t} = X_{i,t|t-1} + \boldsymbol{K}(O_{i,t} - \boldsymbol{G} X_{i,t|t-1}) \quad \text{(Filtering Step)}$$

$$X_{i,t+1|t} = \boldsymbol{A} X_{i,t|t} \quad \text{(Forecasting Step)}$$

Replace $X_{i,t+1|t}$ with $\hat{Y}_{i,t+1|t}$ and define the "now-cast" variable $X_{i,t|t}$ as latent state variable $\Theta_{i,t}$ in Generic Learning Model, we can re-write Kalman Filter (44) as equation (45) and (46), which reflect the generic formulation of updating step (2) and forecasting step (3). It is obvious that in the stationary Kalman Filter case, both $F(.)$ and $H(.)$ are linear.

$$\hat{Y}_{i,t+1|t} = \boldsymbol{A} \Theta_{i,t} \tag{45}$$

$$\Theta_{i,t} = (\boldsymbol{A} - \boldsymbol{KGA})\Theta_{i,t-1} + \boldsymbol{KG} X_t + \boldsymbol{K} \nu_{i,t} \tag{46}$$

Average Structural Function I then turn to the ASF implied by Kalman Filter (45) and (46). This is simply done by taking expectation of $\hat{Y}_{i,t+1|t}$ conditional on observables X_t. The goal is to integrating out the i.i.d noise term $\nu_{i,t}$ which is not observable by econometrician. Now we can define the sufficient statistics for $\Theta_{i,t}$ as:

$$\theta_{i,t} = \mathbb{E}[\Theta_{i,t} | \{X_\tau\}_{\tau=0}^t] \tag{47}$$

Taking the expectation of (45) and (46) conditional on history of the observable $\{X_\tau\}_{\tau=0}^t$ it immediately follows:

$$\begin{aligned} y_{i,t+1|t} \equiv \mathbb{E}[\hat{Y}_{i,t+1|t} | \{X_\tau\}_{\tau=0}^t] &= \boldsymbol{A}\theta_{i,t} \\ \theta_{i,t} &= (\boldsymbol{A} - \boldsymbol{KGA})\theta_{i,t-1} + \boldsymbol{KG}X_t \end{aligned}$$

This illustrates the link between ASF with the underlying expectation formation model: in the linear case with mean zero error $\nu_{i,t}$, the function form from ASF, $f(.)$ and $h(.)$ are linear and are identical to those from the underlying expectation formation model.

Estimation with Simulated Sample Now suppose as econometricians we want to estimate marginal effect of two signals $\{\pi_t, s_{i,t}\}$ on $\pi_{i,t+1|t}$. The standard approach is to directly estimate the reduced-form equation derived from (44) with OLS. This requires $X_{i,t+1|t}$ observed for each t and the learning model is correctly specified. However in reality it is possible that expectation on latent state $L_{i,t+1|t}$ is not observable or not considered in the model[1]. If this is the case OLS with only lag term $\pi_{i,t|t-1}$ is included in the regression suffers from omitted variable problem.

On contrary, estimation with RNN does not require a correct specification on latent variable $\Theta_{i,t}$, and it doesn't need $L_{i,t|t-1}$ to be observable at all. To show this I simulated 100 random samples according to the Kalman Filter as in (44). In this experiment I consider three different models to estimate marginal effect of the two signals $\{\pi_t, s_{i,t}\}$: (1) the RNN with sequence of $\{\pi_\tau, s_{i,\tau}\}_{\tau=0}^t$ and lag expected inflation $\pi_{i,t|t-1}$ as input[2]; (2) mis-specified OLS that uses the same set of variables as dependent variable, the OLS is mis-specified because $L_{i,t+1|t}$ is not available to econometricians; (3) correctly specified OLS with $L_{i,t+1|t}$ observable, which is typically not available. I'll show RNN can still recover the linear relationship between signal

[1]For example, when agent form expectation on inflation, if they believe in a three equation New Keynesian Model, they may also want to infer demand and supply shocks as unobserved states. In a Kalman Filter that takes only inflation as unobserved state, OLS will suffer from omitted variable problem.

[2]Interestingly, for estimating ASF and marginal effect, one do not need to include the lag expectation $\pi_{i,t|t-1}$ in RNN, only history of signals are sufficient. The results without lag expectation are similar to these results I include here.

and expectational variable as well as obtain comparable estimate on signals as the correctly specified OLS estimator (BLUE in this case), whereas mis-specified OLS is heavily biased.

I first depict the recovered average structural function between inflation expectation $\pi_{i,t+1|t}$ and signals π_t, $s_{i,t}$ in Figure 6. The red solid line is the true Average Structural Function implied by the Kalman Filter (44) and the black solid line is the mean of estimated ASF from 100 random samples using RNN. I also plot estimated ASF for each sample in grey color. The top panel in Figure 6 is the ASF along dimension of realized inflation π_t and the bottom panel is along signal $s_{i,t}$. It is obvious that the estimated ASF all indicate linear relationship between signals and expected inflation. This means RNN will recover a linear function if the underlying expectation formation model is indeed linear. It also shows the stability of the performance of RNN: with 100 random samples it recovers the ASF relative close to the truth.

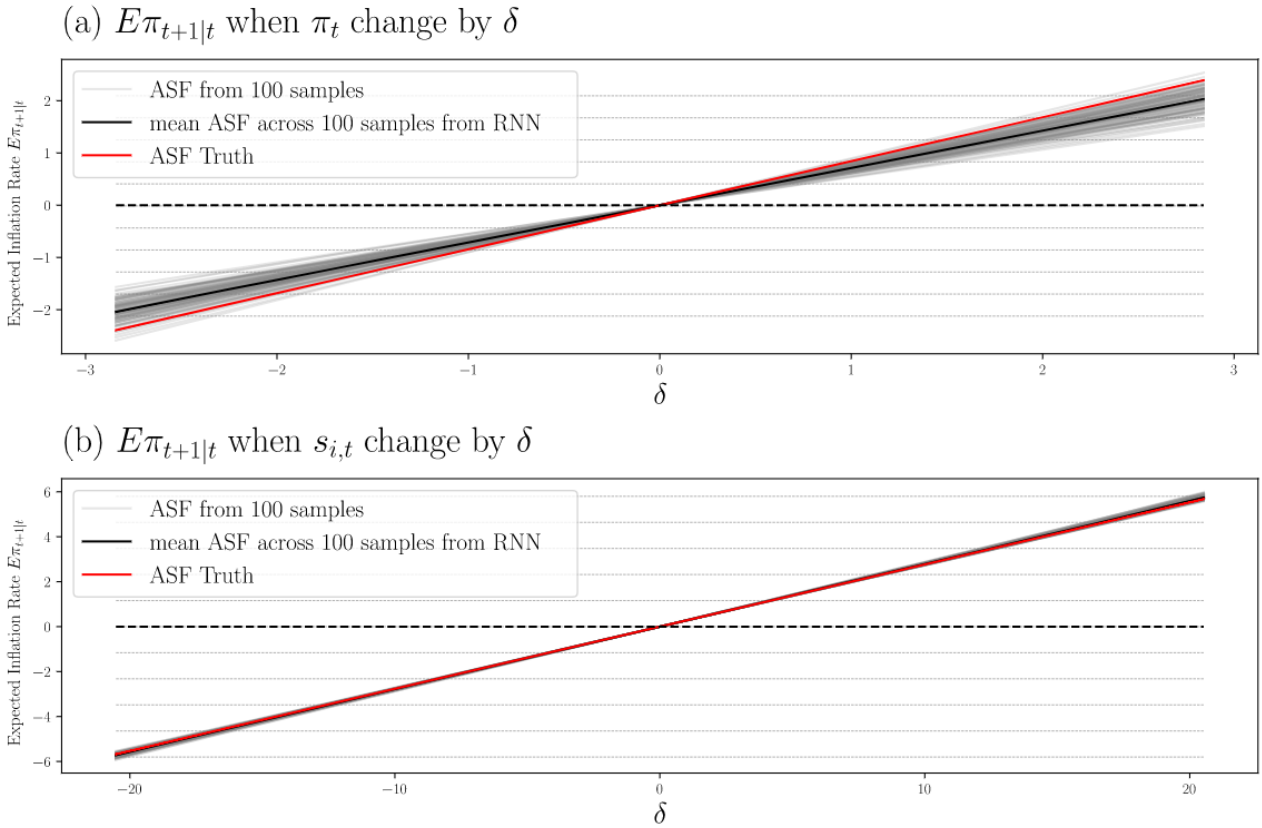


Figure 6: Estimated Average Structural Function from random samples using RNN. Function depicts change of expected variable in response to corresponding signal change by δ. Panel (a): expected inflation as function of inflation signal π_t. Panel (b): expected inflation as function of private signal $s_{i,t}$. Red solid line is the actual ASF implied by linear Kalman Filter. Solid black line is the mean of estimated ASF from 100 random samples. Grey lines are estimated ASFs from each random sample.

I then report the (naive) estimates of marginal effects from RNN and compare them to those from the other two models considered. The following table shows the estimation result from RNN, mis-specified OLS and correctly specified OLS. In this table, the first column is mean squared error on the whole sample, the second column is estimated marginal effect on signal π_t and third column is estimated marginal effect on signal $s_{i,t}$. In brackets I report the standard deviation of the estimate using 100 simulated random samples. Not surprisingly, correctly specified OLS is BLUE in this case with unbiased estimates and small standard deviations. However the key thing to notice here is that mis-specified OLS is biased due to the omitted latent state, whereas RNN has result that is consistent with the true marginal effect, with acceptable standard deviations across 100 samples.

Table 7: Performance of RNN v.s. OLS

	MSE	π_t	$s_{i,t}$
(1) RNN	2.91	0.82	0.276
	(0.054)	(0.037)	(0.003)
(2) OLS mis-specified	3.296	0.720	0.279
	(0.023)	(0.033)	(0.001)
(3) OLS correct	2.835	0.841	0.277
	(0.014)	(0.005)	(0.001)
Truth		0.842	0.277

* The first column is mean squared error on the whole sample, the second column is estimated marginal effect on signal π_t and third column is estimated marginal effect on signal $s_{i,t}$. In brackets I report the standard deviation of the statistics using 100 simulated random samples.

C.2 Example: Constant Gain Learning in Generic Learning Framework

In this subsection I will illustrate how a standard Constant Gain Learning model can be analytically expressed in the form of the Generic Learning Framework. An example of such model is from Evans and Honkapohja (2001). For simplicity I consider the one dimensional case, where an agent observes realized inflation π_t at each time t and try to form forecast about π_{t+1}. I also drop the individual indicator i to same some notations, but the framework can be easily generalized to multi-dimensional multi-agent case. The agent believes in a "Perceived Law of Motion" (PLM) about how inflation is evolving in time and try to estimate the relevant parameters in the PLM using observed data. To do this, she will run OLS at every period and apply a constant weight to the newly available data. With this learning scheme agent perceives different values for parameters in their PLM and form expectation accordingly. The model features a constant gain γ, which represents the weight the agent put on newly observed data. Let's assume the PLM the agent believes in is an AR(1) process:

$$\pi_{t+1} = b_0 + b_1\pi_t + \eta_{t+1} \quad \text{(PLM)}$$

In this setup, the parameters agent try to learn from realized data are b_0 and b_1. η_{t+1} stands for the mean zero i.i.d random shock realized in each period. The agent uses and OLS method to estimate b_0 and b_1 every period, and this process can be formulated recursively such that in each period the agent forms a different estimate b_t:

$$b_t = b_{t-1} + \gamma R_t^{-1}\boldsymbol{X}_{t-1}(\pi_t - b'_{t-1}\boldsymbol{X}_{t-1})$$
$$R_t = R_{t-1} + \gamma(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1} - R_{t-1})$$

$$\boldsymbol{X}_t = [1 \quad \pi_t]' \quad b = [b_0 \quad b_1]'$$

At time t, the agent then forms expectation about future inflation using the PLM, with some i.i.d noise attached on top of the endogenous component that comes from constant gain learning process, ϵ_t. This exogenous component is sometimes interpreted as "sentiment", for example in Cole and Milani (2020).

$$E_t\pi_{t+1} = b'_t\boldsymbol{X}_t + \epsilon_t \tag{48}$$

Now suppose the agent is learning with the above set-up. As observers we see: $\boldsymbol{X}_t, E_t\pi_{t+1}$ up to each time t. We do not see the hidden variables such as b_t and R_t. We also don't know the function form that connects the hidden variables, observables and expectational variables. The goal now is to represent the system described by this constant gain learning model in terms of the Generic Learning Framework. Define the hidden states $\Theta_t = [\boldsymbol{X}_t, b_t, R_t, \epsilon_t]'$. The recursive mapping from observables (and previous hidden states) to hidden states $H(.)$ then can be given by:

$$\Theta_t = H(\boldsymbol{X}_t, \Theta_{t-1}, \epsilon_t)$$

Where

$$\boldsymbol{X}_t \equiv H_1(\boldsymbol{X}_t, \Theta_{t-1}, \epsilon_t) = \boldsymbol{X}_t$$
$$R_t \equiv H_2(\boldsymbol{X}_t, \Theta_{t-1}, \epsilon_t) = R_{t-1} + \gamma(\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1} - R_{t-1})$$
$$b_t \equiv H_3(\boldsymbol{X}_t, \Theta_{t-1}, \epsilon_t) = b_{t-1} + \gamma R_t^{-1}\boldsymbol{X}_{t-1}(\pi_t - b'_{t-1}\boldsymbol{X}_{t-1})$$
$$\epsilon_t \equiv H_4(\boldsymbol{X}_t, \Theta_{t-1}, \epsilon_t) = \epsilon_t$$

Notice here, as Θ_t can be any measurable function of $\boldsymbol{X}_t$, Θ_{t-1} and ϵ_t, it can certainly contain elements such as the input $\boldsymbol{X}_t$. Although $\boldsymbol{X}_t$ is actually observable, it remains "hidden" to econometrician as without further knowledge on expectation formation process, one will not know what the exact mapping from observables to elements of Θ_t is. Then the expectation formation model $F(.)$ is given by:

$$E_t\pi_{t+1} \equiv F(\Theta_t) = b_t'\boldsymbol{X}_t + \epsilon_t$$

Now I show that the expectation formed by constant gain learning can be analytically represented by the Generic Learning Framework described by updating step (2) and forecasting step (3). The Average Structural Function implied by this setup is straight forward: one can define $\theta_t = [\boldsymbol{X}_t, b_t, R_t]'$ and obtain $f(.)$ and $h(.)$ by integrating out the i.i.d random variable ϵ_t.

D Appendix on Empirical Findings

D.1 More on Time-varying Marginal Effect

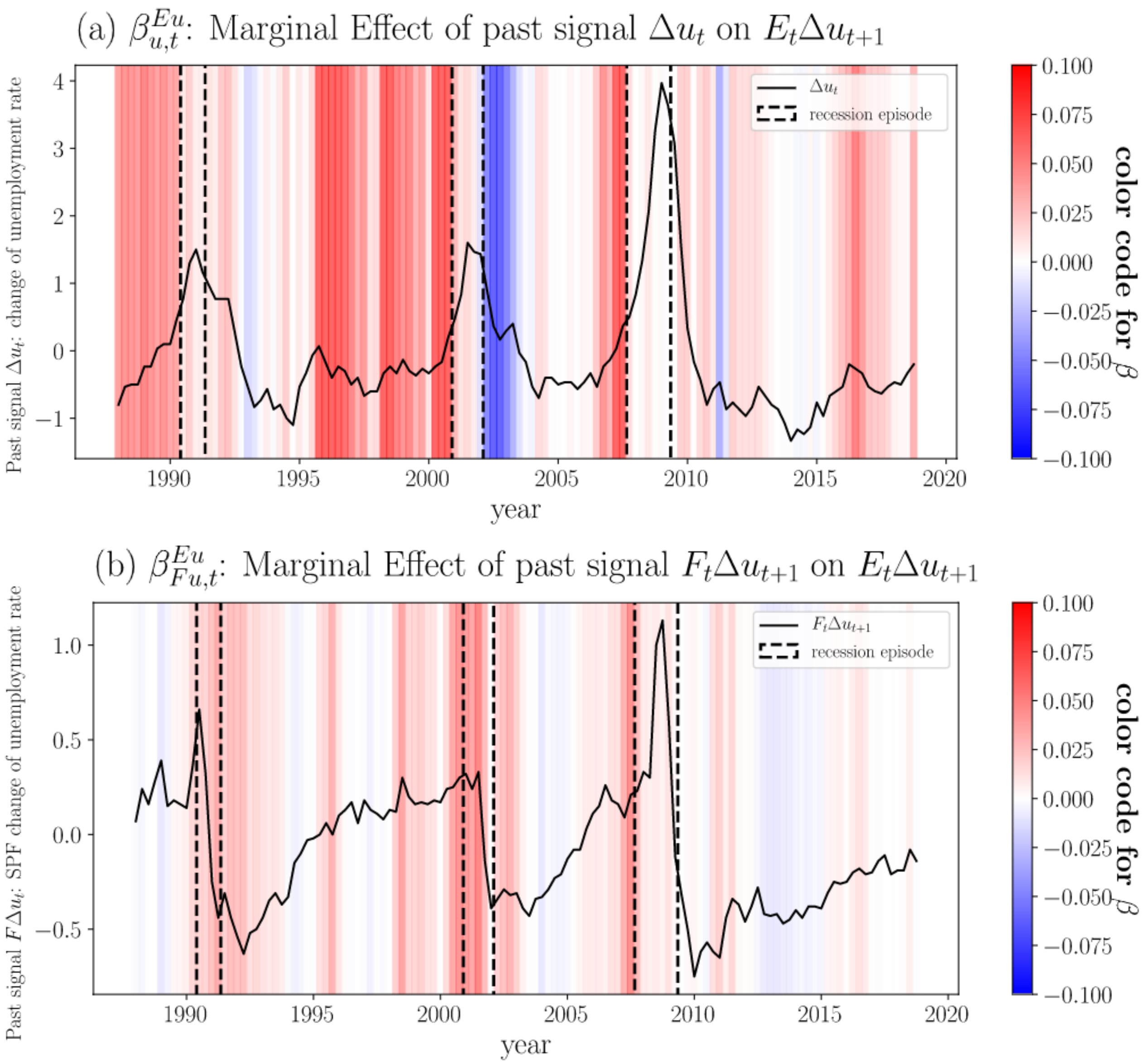


Figure 7:

To show the same attention shift pattern holds for all signals and expectations related to economic condition, I first plot the same heatmap for marginal effect of unemployment signals on expectation on unemployment change. This is Figure 7 below. It shows the same pattern holds as in Figure 3: in recession marginal effect of future signal is bigger and the opposite is true for past signal.

For marginal effects of cross-signals, for example, the impact of unemployment signal on economic condition expectation. These results are shown in Figure 8 below. It shows first unemployment signals generally have negative impact on expectation of economic condition. Furthermore, when looking at marginal effects of past signals, such an impact is again weak during recession periods whereas the marginal effects of future signals are again with bigger magnitudes during recessions.

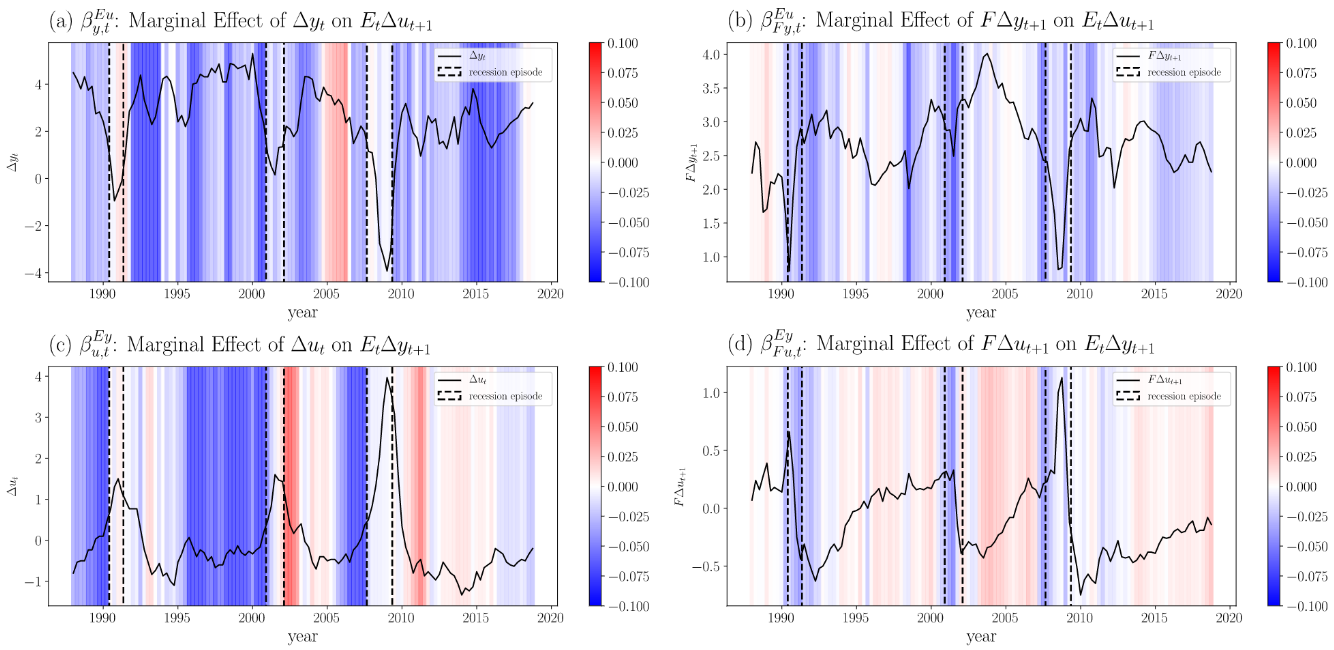


Figure 8:

However these attention shift during recession and ordinary period only holds significantly for expectations and signals related to indicators about economic conditions. Figure 9 plots the time-varying marginal effects for indicators on inflation and interest rate, there is no such attention shift at presence. The DML estimator also suggest the average marginal effects in recession and ordinary periods are not significantly different.

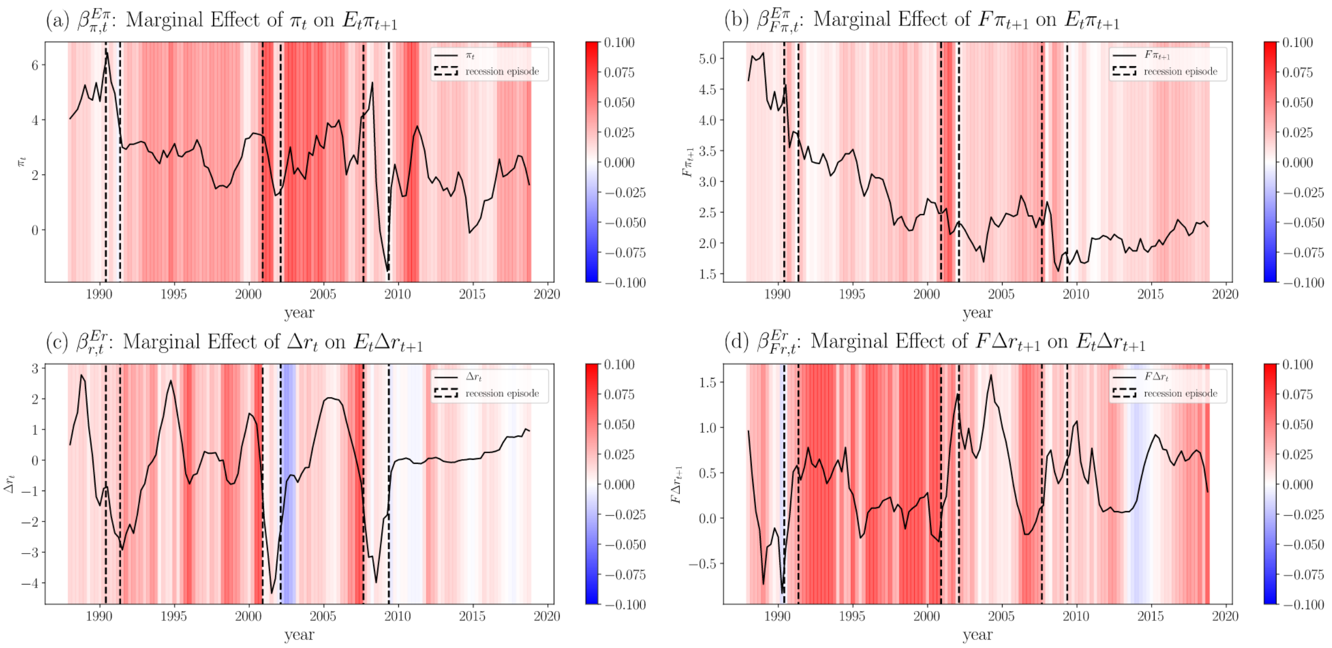


Figure 9:

D.2 Robustness of DML using NBER Recessions

Table 8: Average Marginal Effect of Past and Future Signals: NBER Recession

Expectation:			$E\Delta y_{t+1\|t}$			$E\Delta u_{t+1\|t}$	
	Signal	β_{bad} (std)	β_{ord} (std)	$\beta_{bad}=\beta_{ord}$ (p-val)	β_{bad} (std)	β_{ord} (std)	$\beta_{rec}=\beta_{ord}$ (p-val)
Future Signal	$F_t\Delta u_{t+1}$	**−0.047*** **	0.005	¡0.01	**0.033*** **	0.009***	¡0.01
		(0.006)	(0.002)		(0.004)	(0.002)	
	$F_t\Delta y_{t+1}$	**0.05*** **	0.02***	¡0.01	**−0.024*** **	−0.01***	¡0.01
		(0.007)	(0.003)		(0.003)	(0.001)	
Past Signal	Δu_t	−0.016*	−0.018***	0.86	0.012***	0.01***	0.74
		(0.008)	(0.003)		(0.005)	(0.002)	
	Δy_t	0.003	**0.015*** **	0.05	−0.004**	**−0.01*** **	0.04
		(0.004)	(0.002)		(0.002)	(0.001)	

* ***,**,*: Significance at 1%,5% and 10% level. β_{bad} is average marginal effect in bad periods defined by NBER recession dates, β_{ord} is average marginal effect in ordinary period. $\beta_{bad}=\beta_{ord}$ is test on equality between average marginal effects, its p-value is reported for each expectation-signal pair. Bold estimates denote the marginal effect with significantly bigger magnitude. Standard errors are adjusted for heteroskesticity and clustered within time.

Table 8 shows the DML estimates for marginal effects of past and future signals on real GDP growth and unemployment rate, during or out of recession. And the recession dates in use are those from NBER. Although "bad times" defined in **Section 4.2.2** are considered more plausible for reasons discussed before, using NBER recession dates won't qualitatively change the DML estimates much. Future signals still significantly have higher weights during bad periods and the weights on past signals are usually with bigger magnitude in ordinary period.

D.3 Decompose Time-varying ME with other signals

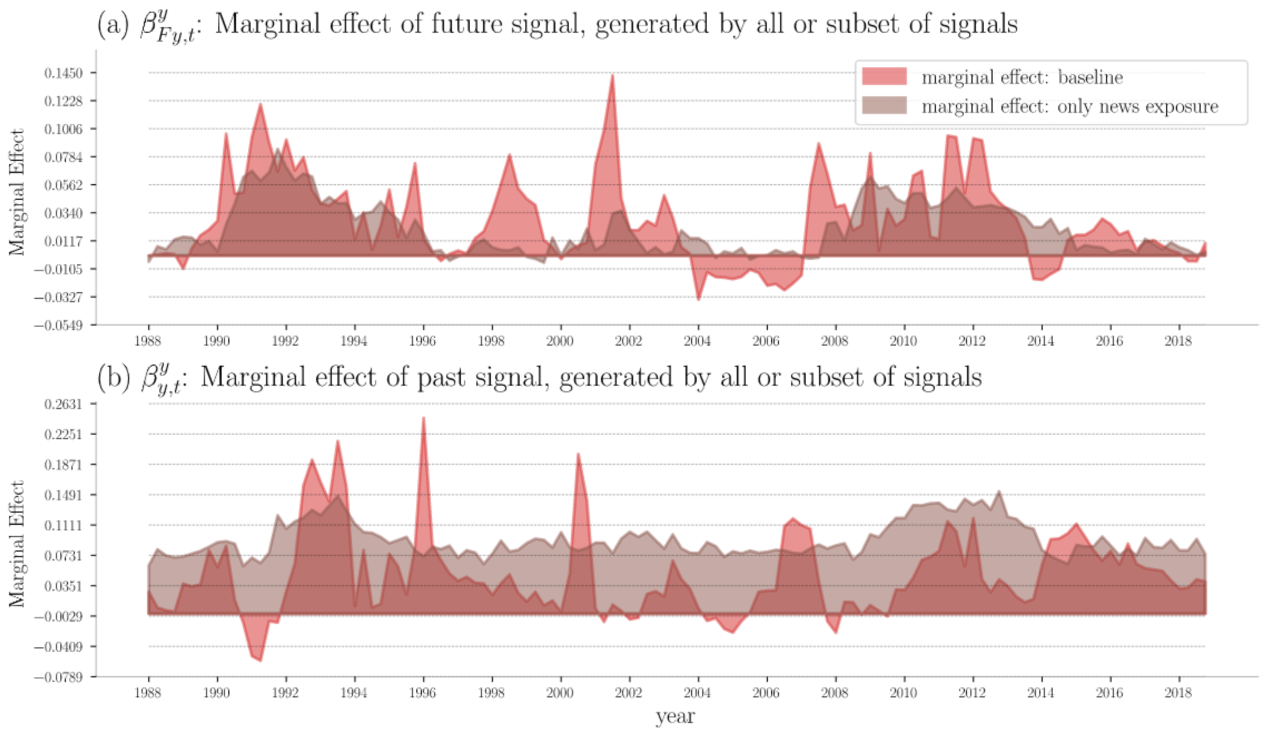


Figure 10: Time-varying marginal effect of past and future signal on real GDP growth. Top panel: marginal effect of future signal, $\beta^{Ey}_{Fy,t}$; bottom panel: marginal effect of future signal, $\beta^{Ey}_{y,t}$. The red curve: marginal effect created by estimated ASF with all signals. The blue curve: marginal effect created by ASF with only exposure of economic condition news.

Figure 10 presents how news exposure affects the marginal effects on future and past signals. It shows that news exposure only creates higher weights on future signals (from SPF) exactly when there is more news on economic status but not the weight change of past signals. According to Table 4, news exposure only accounts for 28% and 15% time-variation of weights on future and past signals, whereas economic conditions alone explain more than 50%. These suggest the explanation that attention-shift is majorly a result of more information available in recessions is unlikely to be true. On the other hand, economic condition signals without news exposure successfully recreate the key attention-shift pattern. This indicates economic condition signals explain a much bigger fraction of the time variation and are likely to be the main driving force for attention shift.

In Figure 11 I report the same exercise as in Figure 4 and 10 but with only signals on inflation, interest rate, and oil prices as input. The results show that information on prices alone cannot recreate the attention-shift pattern.

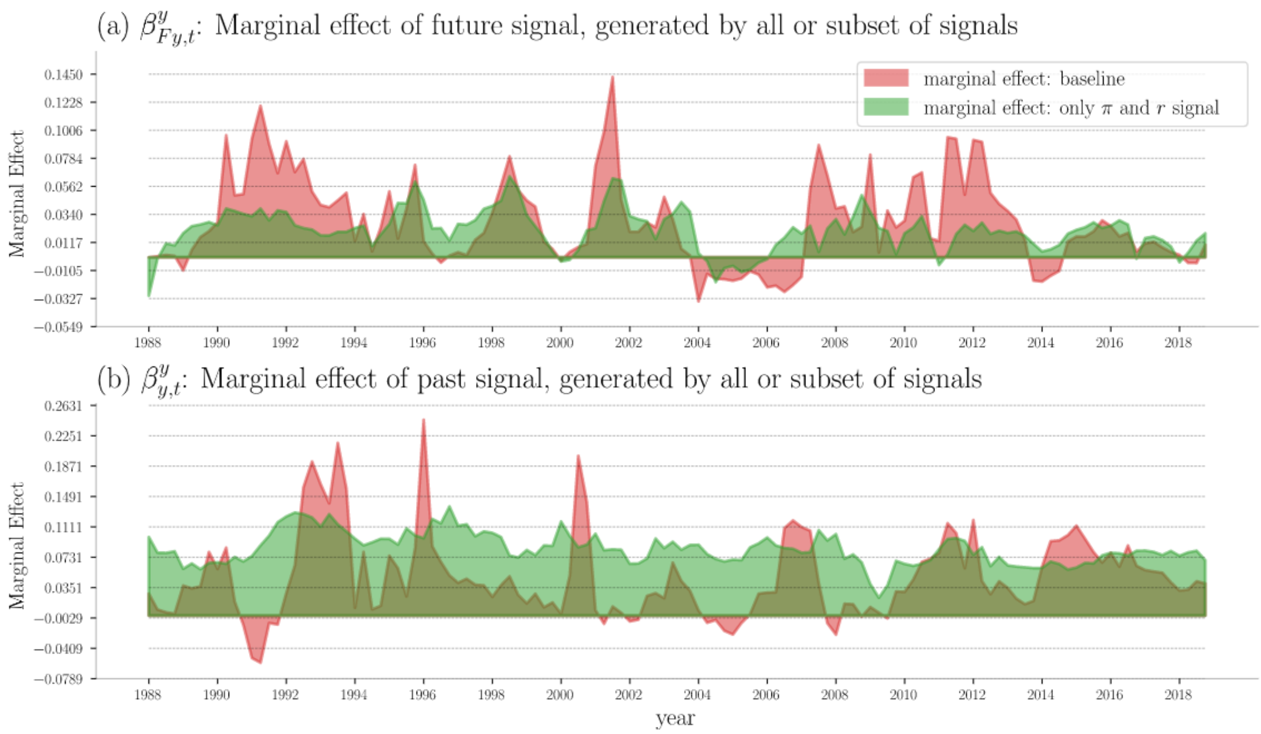


Figure 11: Time-varying marginal effect of past and future signal on real GDP growth. Top panel: marginal effect of future signal, $\beta^{Ey}_{Fy,t}$; bottom panel: marginal effect of future signal, $\beta^{Ey}_{y,t}$. The red curve: marginal effect created by estimated ASF with all signals. The blue curve: marginal effect created by ASF with only interest rate and inflation signals.

D.4 Variance Decomposition for Unemployment Expectation

In Table 9 I summarize the variance decomposition of time varying marginal effects of unemployment signal on unemployment expectations. It is consistent with what I find for expectation on economic condition. First the signals that explain most of the time-variation are those related to economic conditions. News exposure also explain a significant part of variation, especially for past signals. Finally these signals affect expectations through both accumulated states and covariates. Current signal usually plays a more important role in explaining the time-variation.

E Model Appdendix

Table 9: Variance Decomposition of Time-varying Marginal Effects: $E\Delta u$

Marginal Effect on Past Signal:		$\beta^{Eu}_{u,t}$				
Signal Type:		Economic Condition	Inflation	Interest rate	News	Total
	State $\theta_{i,t-1}$	28%	3%	6%	20%	57%
Channel:	Covariate $Z_{i,t}$	23%	2%	13%	5%	43%
	Total	52%	5%	18%	25%	
Marginal Effect on Future Signal:		$\beta^{Eu}_{Fu,t}$				
Signal Type:		Economic Condition	Inflation	Interest rate	News	Total
	State $\theta_{i,t-1}$	19%	6%	7%	4%	36%
Channel:	Covariate $Z_{i,t}$	36%	4%	9%	15%	64%
	Total	54%	10%	16%	19%	

E.1 Signals and Beliefs

At the beginning of time t, agent is endowed with some prior beliefs on states d_t and η_t, this reflects the latent states in empirical part. I denote the prior of foundamentals as:

$$\boldsymbol{X}_0 \equiv \begin{bmatrix} d_0 \\ \eta_0 \end{bmatrix} \sim N(\hat{\boldsymbol{X}}_0, \Sigma_0)$$

Where $\hat{\boldsymbol{X}}_0$ stands for prior mean of the states $\boldsymbol{X}_t$.[3]

The agent is Bayesian Learner and forms posterior beliefs using Kalman Filter. Agent updates his belief twice: first, he is exposed to a normal noisy signal z_0 about current state d_t. The variance of the noise is σ_z^2. The agent then updates her belief on $\boldsymbol{X}_t$. Because both prior and noise are normally distributed, the updated prior is also normal.

$$\boldsymbol{X}_{t|0} \equiv \begin{bmatrix} d_{t|0} \\ \eta_{t|0} \end{bmatrix} \sim N(\hat{\boldsymbol{X}}_{t|0}, \Sigma_{t|0})$$

I define $\boldsymbol{X}_{t|0}$ as conditional prior as it contains information about d_t. Specifically, its mean $\hat{\boldsymbol{X}}_{t|0}$ is a function of the unconditional prior mean and signal z_0, which contains information about d_t and noise. However, the agent has no control of the variance of this noise σ_z^2. It will

[3] In steady state one can think of the prior mean being at the long-run mean of each state, which is 0. When an agent observes a history of signals before time t, she may have a prior mean different from 0. This then can be thought of as a form of the "internal states" described in Section 4.2.4.

not be in the agent's choice set and will be treated as given when the agent solves the rational inattention problem later.

The second time the agent updates belief is after observing signal $\boldsymbol{Z}_t$. He forms a posterior belief about the fundamentals next period. As this is a two-period model, only belief on d_{t+1} is relevant. Again the agent forms belief using Bayes Rule:

$$\boldsymbol{X}_{t+1|t} \equiv \begin{bmatrix} d_{t+1|t} \\ \eta_{t+1|t} \end{bmatrix} \sim N(\hat{\boldsymbol{X}}_{t+1|t}, \Sigma_{t+1|t})$$

Where posterior mean $\hat{\boldsymbol{X}}_{t+1|t}$ and variance $\Sigma_{t+1|t}$ is defined as:

$$\hat{\boldsymbol{X}}_{t+1|t} \equiv \mathbb{E}[\boldsymbol{X}_{t+1}|\boldsymbol{Z}_t] = A\Big((I - KG)\hat{\boldsymbol{X}}_{t|0} + K\boldsymbol{Z_t}\Big) \tag{49}$$

$$\Sigma_{t+1|t} = A\Sigma_{t|0}A' - AKG\Sigma_{t|0}A' + \boldsymbol{Q} \tag{50}$$

And Kalman Gain is given by (51), where matrices A and G are given by exogenous parameters $\{\rho, \rho_\eta\}$ about the fundamentals.

$$K = \Sigma_{t|0}G'(G\Sigma_{t|0}G' + R)^{-1} \tag{51}$$

From (49)-(51), the choices of signal precision will affect both mean and variance of his posterior through the variance-covariance matrix on noise, R. Signals with lower variance are more accurate, and the agent will put higher weights on these signals. Each different choice of signal accuracy (represented by the variance-covariance matrix on noise, R) gives the agent a different information set. Given different information sets, the agent will form different posterior beliefs even if the signals realized are the same.

E.2 Two Special Information Set

At this point, it is worth describing several special information sets:

d_t Fully Observable: At time t, an agent has only perfect information about d_t and no information on η_t. This happens when $\sigma_{2,\xi} = 0$ and $\sigma_{1,\xi} \to \infty$. In this case agent will form adaptive expectation about return in the future: $E_t^A d_{t+1} = \rho d_t$.

Both fundamentals $\boldsymbol{X}_t$ Fully Observable: At time t, an agent has all the information about fundamentals at time t. Given the distribution of $\epsilon_{1,t+1}$, an agent with this information set can form a posterior belief on the distribution of d_{t+1} with mean being expressed as (23). This can be thought of as the Full Information Rational Expectation benchmark in this model as the forecasting error in this case will only be the unpredictable shock $\epsilon_{1,t+1}$.

An information set with an arbitrary variance-covariance matrix on noise, R, can be thought of as being in the middle of the two information sets described above. For each information set $\mathcal{I}_t$ given, the agent will solve her optimization problem (17) accordingly. Different information sets will then result in different choices, thus giving the agent different expected utility. In this sense, information has a value that can be evaluated with her expected utility. I will first illustrate the value of information in this model using different information sets described here.

E.3 State-dependent Value of Information

In this section, I explicitly compute the agent's expected utility conditional on different information sets given. I will illustrate that more information is valuable to agents as it increases their expected utility. Furthermore, the improvement of expected utility obtained by possessing more information depends on the current state of the economy, d_t.

I solve problem (18) under the two different information sets introduced before as well as the case with no extra information. Then I evaluate the agent's ex ante expected utility.

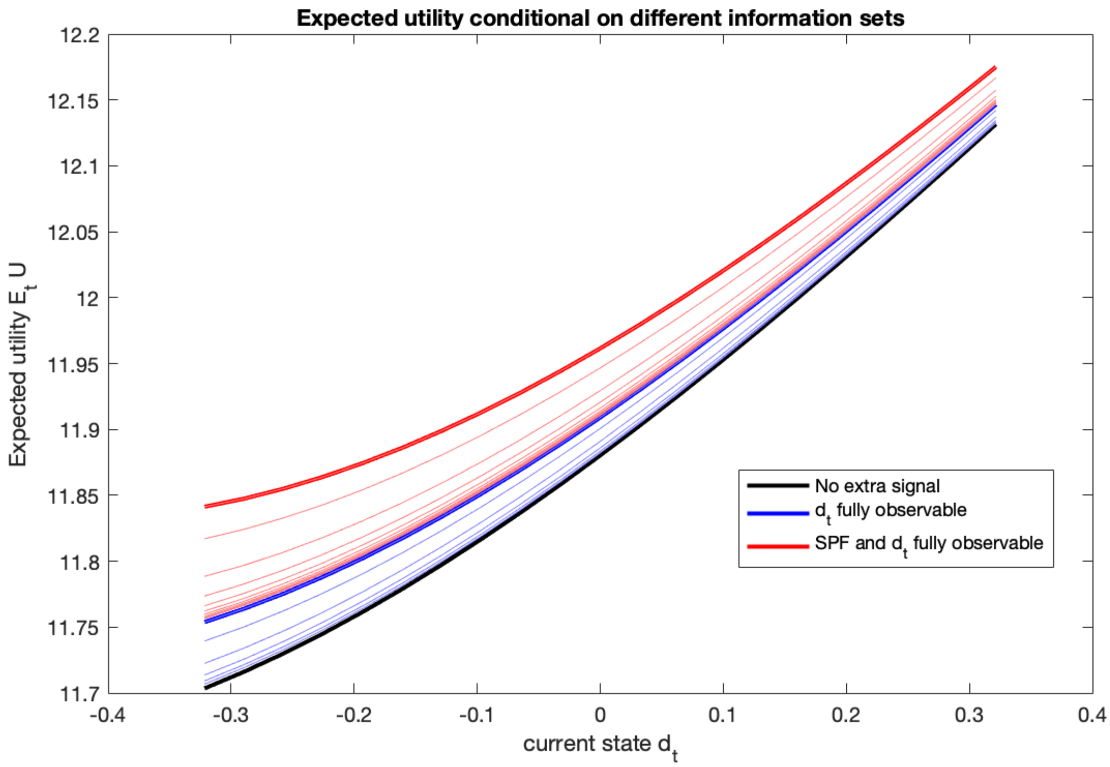
$$\mathbb{E}[u(e_t - s^*_{t+1}(\mathcal{I}_t)) + \beta u(r_{t+1}s^*_{t+1}(\mathcal{I}_t) + e_{t+1})|\mathcal{I}_0]$$

Recall the utility function takes the form $u(c_t) = c_t - bc_t^2$, and in the information set $\mathcal{I}_0$, it contains information about the current state d_t. The state-dependency seen later comes from the fact that the value of information changes as the mean of prior contained in $\mathcal{I}_0$ changes. The quadratic function form makes the point that the common mean-independent result of the rational inattention model is not due to linear quadratic preference per se, rather it's because of the quadratic approximation for the entire problem. However, the results are not restricted to such a utility function form. Recall that given different information set $\mathcal{I}_t$, the first order condition for problem (18) takes the form:

$$s^*_{t+1}(\mathcal{I}_t) = \frac{-1 + 2be_t + (\beta - 2b\beta e_{t+1})\mathbb{E}[r_{t+1}|\mathcal{I}_t]}{2b(1 + \beta\mathbb{E}[r^2_{t+1}|\mathcal{I}_t])} \tag{52}$$

For illustration purposes, I solve the model numerically using the following parametrization: $b = 1/40$, $\beta = 0.95$, $e_t = 10$ and $e_{t+1} = 5$. For the fundamentals I consider $\rho = 0.2$, $\rho_d = 0.9$, $\sigma_{1,\epsilon} = \sigma_{2,\epsilon} = 0.15$. The prior beliefs on states d_t and η_t are assumed to be mean zero with the stationary variance-covariance matrix obtained from the recursive Kalman Filter. The standard deviation of noise on passive signal is $\sigma_z = 0.22$. In Figure 12 I plot the ex ante expected utility conditional on various information sets, as functions of realized d_t. The thick black curve is expected utility when there's no more information other than the initial passive signal on d_t available to the agent. The thick blue curve is expected utility when d_t is fully observable and the thick red curve is when both SPF and d_t are fully observable (the FIRE benchmark)[4]. The curves between the thick lines depict the increase in expected utilities as the precision of the signal increases (or the variance of noise decreases).

Figure 12: Expected Utility under Different Information Set



Black thick line: Expected utility when no more information other than initial passive signal on d_t; blue thick line: expected utility when d_t becomes fully observable; red thick line: when both SPF and d_t fully observable. Blue thin lines are expected utilities when there are noise attached to extra signal on d_t, the more accurate the signal, the closer it gets to d_t fully observable case. Red thin lines are expected utilities when noise attached to signal on SPF, and d_t is fully observable. The more accurate the signal, the closer it gets to full information case.

There are two key messages from Figure 12. First, more information improves the agent's expected utility progressively: with a more accurate signal on d_t, the agent resolves the uncertainty about the current state and his utility increases at any given d_t from the black line to the blue line; and it continues to increase as the signal on SPF becomes more accurate, from blue curve to red curve. This is a typical result of informational models.

Secondly and more importantly, the value of information is decreasing in realized state d_t. This can be seen from the differences between expected utilities with different information sets. When realized state d_t is low and negative, getting the same amount of information will increase the agent's expected utility by more than the case when d_t is high. In other

[4] With the specific law of motion assumed in (19) - (21) together with definition of SPF (23), the case with only SPF fully observable will coincide with the FIRE case.

words, information is more valuable when the economic status is bad. This is a result different from that of standard rational inattention literature. The reason for such a difference is the non-linearity in the optimal saving/investment function.

The difference between expected utility comes from differences of optimal investment (52). The fact that optimal saving is a non-linear function of both posterior mean and posterior variance of state d_{t+1} makes the expected utility mean-dependent. To see this, we can utilize the assumption of the quadratic utility function, and re-write the expected utility as the following form:

$$\begin{aligned}
\mathbb{E}[U(s^*_{t+1}(\mathcal{I}_t))] &= \mathbb{E}[(e_t - s_{t+1}) - b(e_t - s_{t+1})^2 + \beta(e_{t+1} + r_{t+1}s_{t+1}) - \beta b(e_{t+1} + r_{t+1}s_{t+1})^2] \\
&= \mathbb{E}[-\underbrace{b(1+\beta r^2_{t+1})}_{\equiv\chi} s^2_{t+1} + (2be_t - 1 + \beta r_{t+1} - 2\beta b e_{t+1} r_{t+1}) s_{t+1} + (e_t - be^2_t + \beta e_{t+1} - \beta b e^2_{t+1})] \\
&= \mathbb{E}[-\chi(s^2_{t+1} - \underbrace{\frac{2be_t - 1 + \beta r_{t+1} - 2\beta b e_{t+1} r_{t+1}}{\chi}}_{\equiv 2\bar{s}_{t+1}} s_{t+1} + \frac{(2be_t - 1 + \beta r_{t+1} - 2\beta b e_{t+1} r_{t+1})^2}{4\chi^2}) \\
&+ \frac{(2be_t - 1 + \beta r_{t+1} - 2\beta b e_{t+1} r_{t+1})^2}{4\chi} + (e_t - be^2_t + \beta e_{t+1} - \beta b e^2_{t+1})] \\
&= -\mathbb{E}[\chi(s_{t+1} - \bar{s}_{t+1})^2] + \underbrace{\mathbb{E}[\frac{(2be_t - 1 + \beta r_{t+1} - 2\beta b e_{t+1} r_{t+1})^2}{4\chi} + (e_t - be^2_t + \beta e_{t+1} - \beta b e^2_{t+1})]}_{\equiv M}
\end{aligned}$$

Note in the above derivation all the expectations are conditional on initial information set $\mathcal{I}_0$. M has nothing to do with information set $\mathcal{I}_t$, thus the evaluating the expected utility under choice of $\mathcal{I}_t$ is equivalent to evaluating the quadratic loss term $\mathbb{E}[\chi(s^*_{t+1}(\mathcal{I}_t) - \bar{s}_{t+1})^2]$. This is a standard result from literature of Rational Inattention with linear quadratic preference. The key difference here is s^*_{t+1} is non-linear in fundamentals.[5] We can then write ante expected utility as a form of "quadratic loss":

$$\mathbb{E}[U(s^*_{t+1}(\mathcal{I}_t))|\mathcal{I}_0] = -\mathbb{E}[\chi(s^*_{t+1}(\mathcal{I}_t) - \bar{s}_{t+1})^2|\mathcal{I}_0] + M \tag{53}$$

The variable $\bar{s}_{t+1}$ is given by (54). It stands for the optimal investment under perfect foresight when the agent observes d_{t+1} perfectly.

$$\bar{s}_{t+1} = \frac{-1 + 2be_t + \beta r_{t+1} - 2b\beta r_{t+1} e_{t+1}}{2b(1+\beta r^2_{t+1})} \tag{54}$$

The transformed utility function (53) is usually referred to as a quadratic loss function in rational inattention models, intuitively agent will seek to minimize the expected loss between optimal choice under limited information set $\mathcal{I}_t$ and optimal choice under Full Information Rational Expectation.[6] From (53) it is obvious that if the optimal choice of s is linear in state

[5]In standard rational inattention models, the action will be linear in fundamentals thus optimal choice of signal will not depend on prior mean of fundamentals. For example, see Maćkowiak *et al.* (2018) or Kamdar (2019).

[6]It is worth noting that M is not involved in choosing the optimal information structure $\mathcal{I}_t$ as it is only related to the actual distribution of r_{t+1}.

r_{t+1}, the expected utility only depends on the posterior variance of r_{t+1} given information set $\mathcal{I}_t$. It is not related to the posterior mean of states or realized state at time t.

Using the transformed expected utility (53), I can explore reasons for the value of information decreasing in d_t. To see this, consider the cases with or without full information from SPF. Conditional on the realization of a specific d_t, without information from SPF agent faces uncertainty from both η_t and $\epsilon_{1,t+1}$ being unobservable. With information from SPF uncertainty from η_t is resolved. Because in both cases agents have no information on $\epsilon_{1,t+1}$, the utility improvement comes solely from knowledge on η_t. For simplicity, I consider an extreme case when $\epsilon_{1,t+1} = 0$. Then $\bar{s}_{t+1}$ can be seen as the optimal saving choice when SPF is available. The utility loss of the agent not having information from SPF can then be evaluated by differences between optimal savings with or without SPF observable, weighted by the agent's subjective belief.

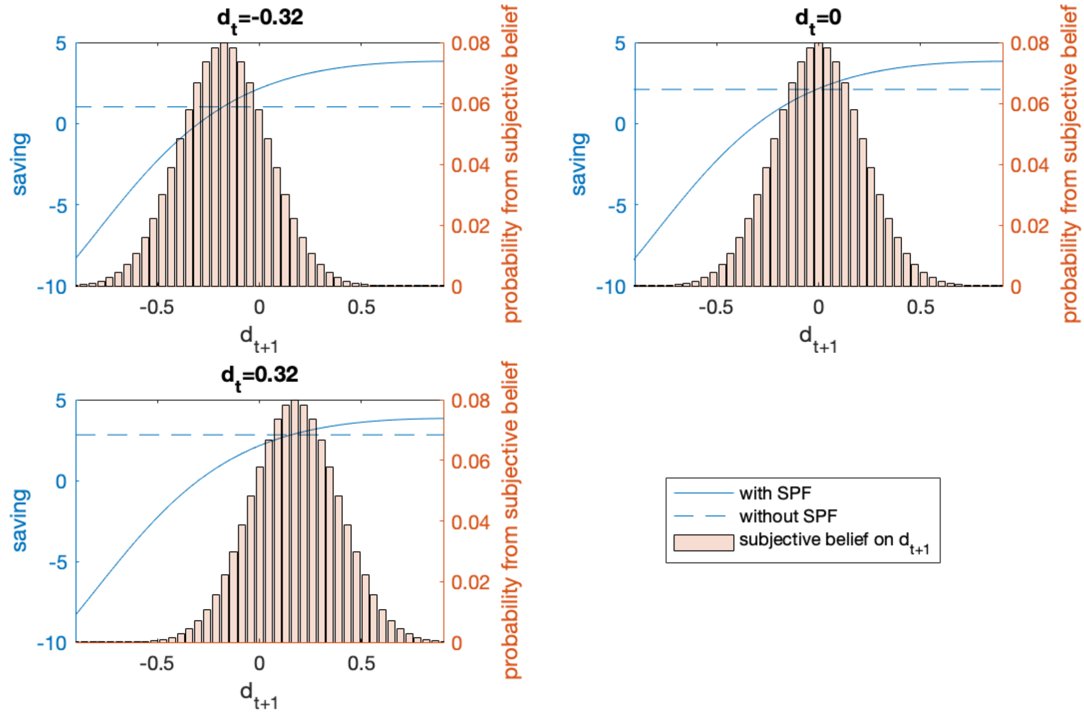
In Figure 13 I depict the optimal saving choices at three realized values of d_t: when the current state is bad ($d_t = -0.32$), neutral ($d_t = 0$) and good ($d_t = 0.32$). In each case, I plot the optimal saving choice as a function of future state d_{t+1}. The dotted line is the optimal saving that the agent chooses when he only observes the initial signal on d_t. It is a flat line because the agent's choice does not depend on d_{t+1} (or realization of η_t) when SPF is not observable. The solid line is the optimal saving choice when SPF is observable to the agent. This line is a function of d_{t+1} because under the assumption $\epsilon_{1,t+1} = 0$, when SPF is observable then η_t and d_{t+1} are fully observed. An important feature is then this function is increasing and concave in d_{t+1}. This is because the higher the return d_{t+1} is, the more agent wants to save. The concavity comes from the fact that the substitution effect becomes weaker as the return on asset increases and is finally dominated by the income effect.[7]

Now for agents without information from SPF, the solid line is not feasible. For a given realized d_t, the agent will evaluate her utility loss of not having information on η_t following (53). This is done by measuring the distance between optimal saving choices with and without information from SPF and computing the expected value of (the square of) this distance using their posterior belief on d_{t+1} (η_t). In Figure 13 this belief is shown with a bar plot. When realized d_t is higher, the belief of distribution on d_{t+1} is centered at a higher mean. Because of the non-linearity of the optimal saving choice, the average distance between saving choices with and without information from SPF is higher when d_t is low. This gives rise to the fact that value of information from SPF is decreasing in d_t.

With the simple structure presented above, I show the key pattern my model generates: the value of information decreases in the state of the economy. The agent is willing to pay higher costs to acquire information as the state of the economy gets worse. This gives the key mechanism to create the time-varying marginal effect and non-linearity I documented

[7]Interestingly, if one would instead assume a riskless asset with a risky endowment in $t+1$, the optimal saving curve under full information will be linear and the value of information won't depend on the current state anymore.

Figure 13: Optimal Saving under Full Information and Limited Information



Solid line: optimal saving choice under full information: when both d_t and SPF are fully observable. Dash line: optimal policy when SPF signal is not available. Bar plot: agent's subjective belief on future state d_{t+1}, when SPF is not observable. Top left panel is when current state is very bad ($d_t = -0.32$), top right panel is when $d_t = 0$ and bottom left panel is when current state very good, $d_t = 0.32$.

with RNN. Because when agents can choose the precision of signals (thus information set) optimally, they will make different choices during bad and ordinary times and this will result in different weights on these signals.

E.4 Derivation of Information Cost

In this subsection, I derive the information cost measured by entropy in (25) following Mackowiak and Wiederholt (2009). Recall the state-space representation of fundamentals are:

$$\boldsymbol{X}_{t+1} = A\boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{Q})$$

The initial noisy signal z_0 and chosen signals $\boldsymbol{Z}_t$ are given by:

$$z_0 = d_t + \xi_0 = G_0\boldsymbol{X}_t + \xi_0, \quad G_0 = [1 \quad 0]$$

$$\boldsymbol{Z}_t = G\boldsymbol{X}_t + \boldsymbol{\xi}_t, \quad \boldsymbol{\xi}_t \sim N(\boldsymbol{0}, R)$$

First notice all the random variables $\boldsymbol{X}_t$, z_0, $\boldsymbol{Z}_t$ are normally distributed. The information set $\mathcal{I}_0$ only contains a noisy Gaussian signal z_0, the entropy of $\boldsymbol{X}_{t+1}$ given $\mathcal{I}_0$ is then:

$$\mathcal{H}(\boldsymbol{X}_{t+1}|\mathcal{I}_0) = \mathcal{H}(\boldsymbol{X}_{t+1}|z_0) = \frac{1}{2}log_2[(2\pi e)^2 det\Sigma_{t+1|0}] \tag{55}$$

Where $\Sigma_{t+1|0}$ denotes the variance-covariance matrix of $\boldsymbol{X}_{t+1}$ given z_0. The prior variance covariance matrix of $\boldsymbol{X}_t$ is denoted as Σ_0, then the conditional variance-covariance matrix $\Sigma_{t+1|0}$ is given by:

$$\Sigma_{t+1|0} = A\Sigma_{t|0}A' + \boldsymbol{Q} \tag{56}$$

Where:

$$\Sigma_{t|0} = \Sigma_0 - \Sigma_0 G_0'(G_0\Sigma_0 G_0' + \sigma_z^2)^{-1}G_0\Sigma_0 \tag{57}$$

It is obvious $\Sigma_{t|0}$ by construction the posterior variance-covariance matrix for hidden states $\boldsymbol{X}_t$ after observing z_0 derived from Kalman Filter.

Then recall $\mathcal{I}_t = \mathcal{I}_0 \cup \{\boldsymbol{Z}_t\}$, similar as above we have:

$$\mathcal{H}(\boldsymbol{X}_{t+1}|\mathcal{I}_t) = \mathcal{H}(\boldsymbol{X}_{t+1}|z_0, \boldsymbol{Z}_t) = \frac{1}{2}log_2[(2\pi e)^2 det\Sigma_{t+1|t}] \tag{58}$$

Where:

$$\Sigma_{t+1|t} = A\Sigma_{t|0}A' - A\Sigma_{t|0}G'(G\Sigma_{t|0}G' + R)^{-1}G\Sigma_{t|0}A' + \boldsymbol{Q} \tag{59}$$

Again by construction, the $\Sigma_{t+1|t}$ is the posterior variance-covariance matrix for $\boldsymbol{X}_{t+1}$ after observing $\{z_0, \boldsymbol{Z}_t\}$ derived from Kalman Filter. Moreover, it is obvious the uncertainty after observing $\boldsymbol{Z}_t$ is reduced compared to the uncertainty after only observing z_0.

Now information cost is obtained by measuring uncertainty reduction induced by extra information, using (58) and (55) we have the information cost in (25):

$$\mathcal{H}(\boldsymbol{X}_{t+1}|\mathcal{I}_0) - \mathcal{H}(\boldsymbol{X}_{t+1}|\mathcal{I}_t) = \frac{1}{2}log_2(\frac{det\Sigma_{t+1|0}}{det\Sigma_{t+1|t}})$$

E.5 Derivation of $\mathbb{E}[r_{t+1}|\mathcal{I}_t]$ and $\mathbb{E}[r_{t+1}^2|\mathcal{I}_t]$

From (49):

$$\begin{pmatrix} \mathbb{E}[d_{t+1}|\mathcal{I}_t] \\ \mathbb{E}[\eta_{t+1}|\mathcal{I}_t] \end{pmatrix} \equiv \hat{\boldsymbol{X}}_{t+1|t} = A\Big((I - KG)\hat{\boldsymbol{X}}_{t|0} + K\boldsymbol{Z_t}\Big)$$

Where $\hat{\boldsymbol{X}}_{t|0} = \mathbb{E}[\boldsymbol{X}_t|\mathcal{I}_0]$ is the mean of belief on $\boldsymbol{X}_t$ after observing passive signal z_0. The prior before observing z_0 is denoted as $\boldsymbol{X}_0 \sim N(\hat{\boldsymbol{X}}_0, \Sigma_0)$ from Section E.1. Now denote the Kalman Gain for observing z_0 as K_0, we can write:

$$\hat{\boldsymbol{X}}_{t|0} = (I - K_0G_0)\hat{\boldsymbol{X}}_0 + K_0 z_0 \tag{60}$$

Where $G_0 = \iota = [10]$ as defined in Appendix E.4 and K_0 is given by:

$$K_0 = \Sigma_0 G_0'(G_0 \Sigma_0 G_0' + \sigma_z^2)^{-1} \tag{61}$$

Combine (49), (60) and (61) we have:

$$\begin{aligned}\mathbb{E}[r_{t+1}|\mathcal{I}_t] &= 1 + \iota A\Big((I - KG)\big((I - K_0 G_0)\hat{\boldsymbol{X}}_0 + K_0 z_0\big) + K Z_t\Big) \\ &= 1 + \iota A\Big((I - KG)\big((I - \Sigma_0 G_0'(G_0\Sigma_0 G_0' + \sigma_z^2)^{-1} G_0)\hat{\boldsymbol{X}}_0 + \Sigma_0 G_0'(G_0 \Sigma_0 G_0' + \sigma_z^2)^{-1} z_0\big) + K Z_t\Big)\end{aligned} \tag{62}$$

Before I show the derivation of $\mathbb{E}[r_{t+1}^2|\mathcal{I}_t]$, it's useful to consider what is $Var(d_{t+1}|\mathcal{I}_t)$. It is the first element of posterior variance covariance matrix $\Sigma_{t+1|t}$, which is given by (59). So $Var(d_{t+1}|\mathcal{I}_t)$ can be written as:

$$\begin{aligned}Var(d_{t+1}|\mathcal{I}_t) &= \iota \Sigma_{t+1|t} \iota' \\ &= \iota\big(A\Sigma_{t|0}A' - A\Sigma_{t|0}G'(G\Sigma_{t|0}G' + R)^{-1}G\Sigma_{t|0}A' + \boldsymbol{Q}\big)\iota'\end{aligned} \tag{63}$$

Now we can derive $\mathbb{E}[r_{t+1}^2|\mathcal{I}_t]$:

$$\begin{aligned}\mathbb{E}[r_{t+1}^2|\mathcal{I}_t] &= Var(r_{t+1}|\mathcal{I}_t) + (\mathbb{E}[r_{t+1}|\mathcal{I}_t])^2 \\ &= Var(d_{t+1}|\mathcal{I}_t) + (\mathbb{E}[r_{t+1}|\mathcal{I}_t])^2 \qquad (64) \\ &= \iota\big(A\Sigma_{t|0}A' - A\Sigma_{t|0}G'(G\Sigma_{t|0}G' + R)^{-1}G\Sigma_{t|0}A' + \boldsymbol{Q}\big)\iota' + (\mathbb{E}[r_{t+1}|\mathcal{I}_t])^2 \qquad (65)\end{aligned}$$

In the above equation, $\Sigma_{t|0}$ is given by (57), which contains σ_z^2 and prior variance Σ_0. $\mathbb{E}[r_{t+1}|\mathcal{I}_t]$ is given by (62), which depends on prior mean $\hat{\boldsymbol{X}}_0$, precision (variance) of the signal R and passive signal z_0. From (62) and (65), it is clear that the optimal saving choice is a non-linear function of all these variables related to the information friction.

Now to see how the *ex-post* weights on signal Z_t depend on variances of signals R, denote the weight on SPF signal as w_{spf} and weight on signal about current state as w_t, from (62) we have:

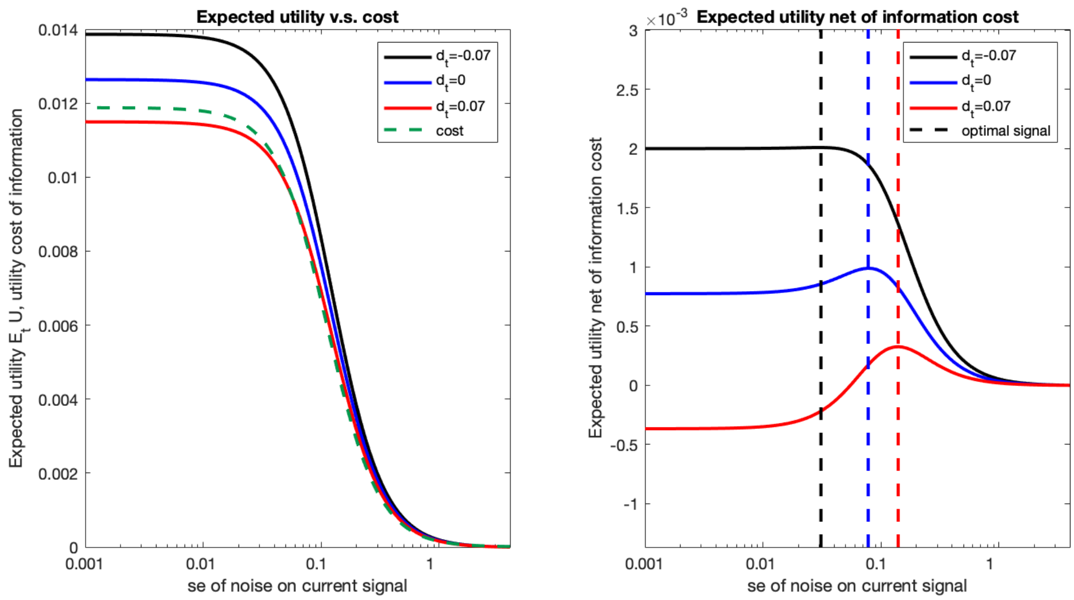
$$\begin{aligned}\begin{pmatrix} w_{spf} \\ w_t \end{pmatrix} &= (\iota' \quad \iota')AK \\ &= (\iota' \quad \iota')A\Sigma_{t|0}G'(G\Sigma_{t|0}G' + R)^{-1}\end{aligned} \tag{66}$$

From we see first for given Σ_{t+0}, G and A, a lower variance of noise on signal (contained in R) leads to higher weights put on corresponding signal. Moreover, as $\Sigma_{t|0}$ is affected by σ_z^2, the weights on signals also change with σ_z^2. This will be verified in Section E.7.1.

E.6 State-dependent Optimal Signals

Now turn to the rational inattention problem (26)-(28). I first show that the trade-off between the benefit and cost of acquiring information changes with the current state d_t. In Figure 14, I present the expected utility, information cost and the objective function in (26) when the current state d_t is negative, at mean zero and positive. The left panel describes how the expected utility changes as the standard error of the current signal ($\sigma_{2,\xi}$) changes for the three cases of d_t. Because at a different level of d_t, the expected utility for the same signal precision will be different, I normalize it by the utility at $\sigma_{2,\xi} = \infty$, which corresponds to the case when the agent acquires no extra signal on the current state. It is obvious in all three cases of d_t, the higher(lower) the precision(standard error) of the signal, the higher the expected utility comparing to the no information case. I then present the information cost for all three cases and show that the information costs are the same across different levels of d_t. This is because, in (30), the passive signal z_0 contains information about the current state d_t thus making the expected utility depending on it. Whereas in the information cost (32), the evaluation of posterior variance is mean-independent, which means only the variance of the passive signal matters in accessing the information cost so that the cost will not change as d_t changes. The key message from the left panel is that both the cost and the benefit of information increase with the precision of the current signal. Meanwhile for the same level of signal precision, the higher the current state d_t, the lower the benefit from that signal.

Figure 14: Information Benefit, Information Cost and Households' Objective: Function of Current Signal

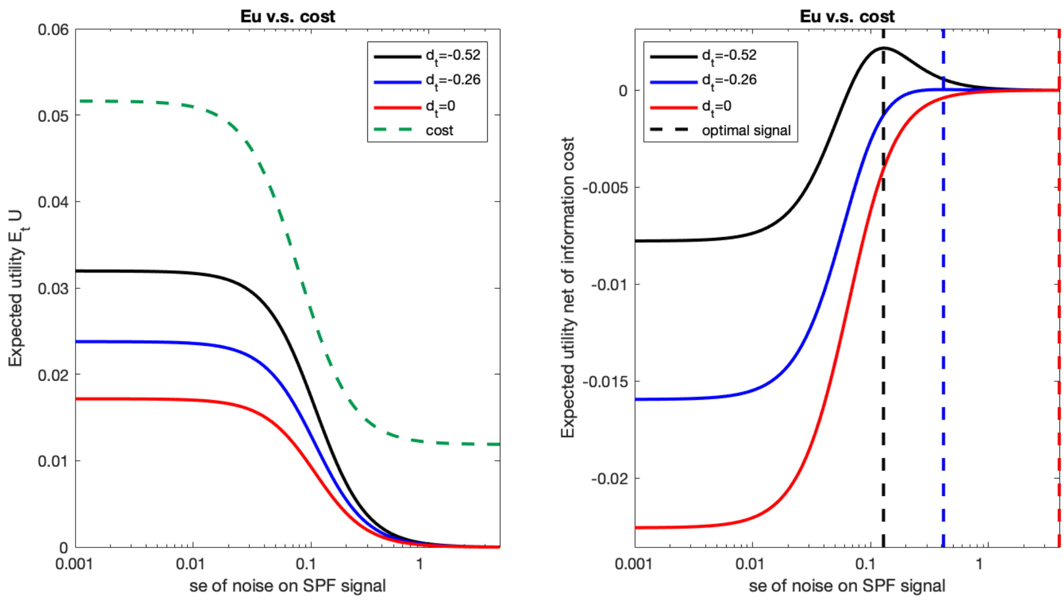


Left panel: the information benefit is evaluated by expected utility and plotted with solid lines, information cost is evaluated by entropy cost and plotted with dashed line. Right panel: objective function is obtained by information benefit minus cost and plotted with solid lines. The figure considers three different cases: current state is high with $d_t = 0.07$, current state is at its mean $d_t = 0$ and current state is low at $d_t = -0.07$. Horizontal axis is standard error of noise on *current signal*, higher s.e. leads to lower weight. Vertical dashed line in the right panel labels optimal s.e. for *current signal* in three scenarios.

The agent's objective function considers both the cost and benefit of acquiring information. The right panel of Figure 14 then presents the objective function under the same three cases of d_t. A utility-maximizing agent will choose the current signal with a standard error that maximizes her objective function. These choices under different states are represented by the dashed line. The right panel then shows that when the current state is worse, the agent will choose a higher precision and a lower standard error for the current signal.

These patterns hold true for precision on signals about SPF as well. In Figure 15, I again show a similar graph as in Figure 14 but for signals on SPF. The major difference between this figure and Figure 14 is that expected utility is computed assuming $\sigma_{2,\xi} = 0.001$, which means the agent chose a very precise current signal.[8] All the objects plotted in Figure 15 are then functions of the precision on SPF signal, $\sigma_{1,\xi}$. Similar to that in Figure 14, we see that both the benefit and the cost increase when the agent acquires more information on SPF. Meanwhile, the optimal precision of the SPF signal decreases with the current state d_t.

Figure 15: Information Benefit, Information Cost and Households' Objective: Function of SPF Signal



Left panel: the information benefit is evaluated by expected utility and plotted with solid lines, information cost is evaluated by entropy cost and plotted with dashed line. Right panel: objective function is obtained by information benefit minus cost and plotted with solid lines. The figure considers three different cases: current state $d_t = 0, -0.26$ and -0.52. For $d_t > 0$ the agent will always choose precision that leads to weight zero because here I plot all the objects under $\sigma_{2,\xi} = 0.001$. Horizontal axis is standard error of noise on *future (SPF) signal*, higher s.e. leads to lower weight. Vertical dashed line in the right panel labels optimal s.e. for *future (SPF) signal* in three scenarios.

[8] However, changing the level of $\sigma_{2,\xi}$, in this case, will not change the results qualitatively.

E.7 Comparative Statistics for Rational Inattention Model

I now examine the impacts of changing model parameters σ_z and prior mean $\hat{\boldsymbol{X}}_0$, on the optimal signal choices.

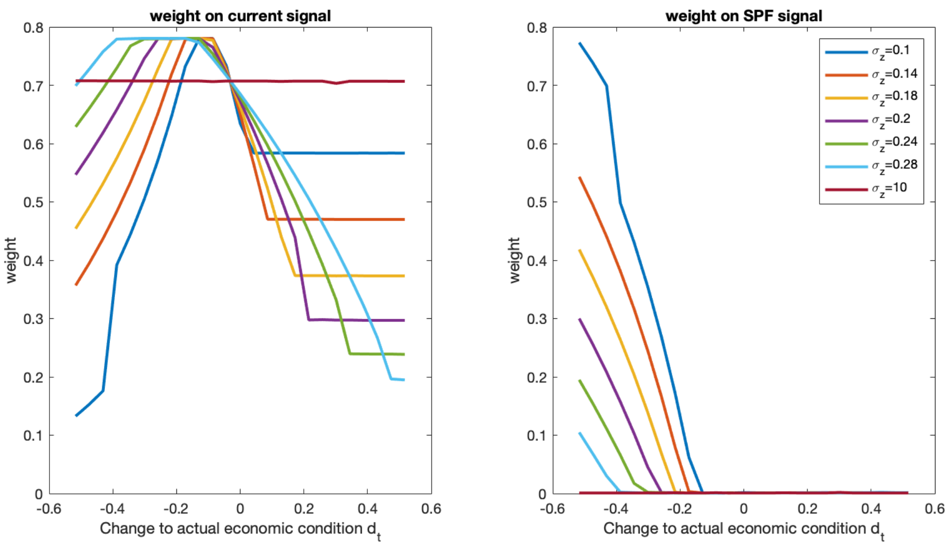
E.7.1 Impact of Passive Signal Precision σ_z

An important parameter in the information friction proposed in this chapter is the precision of the passive signal that the agent is exposed to. This passive signal z_0 will contain information about the current state d_t, thus making the agent's optimal choice of precision dependent on this state. The precision of this signal then determines the prior variances that the agent considers to evaluate the benefit and cost for more information. Through this channel, it will also affect the agent's optimal choices on signal precisions.

In Figure 16 I show again the optimal weights on current and SPF signals as functions of current state d_t, with different values of σ_z. The left panel shows the results for weights on current signals. First, notice higher precision on the passive signal (lower σ_z) leads to higher weight on the current signal to start with. This shows up in the figure as a higher weight in the flat area before the agent puts excess weight on the current signal. Intuitively, this means the agent already has a better understanding of the current d_t before choosing the extra signals on the current state and SPF. This leads to the fact that in the right panel, the agent with the lowest σ_z will start to pay attention to the SPF signal at a relatively higher state, because the information cost for choosing that precision level is relatively lower to her. An extreme example will be that when $\sigma_z = 0$, which implies that the agent has perfect information on d_t. In this case, we will see her only choosing an extra signal on SPF starting from a relatively high value of d_t.

Another interesting aspect in Figure 16 is that when the quality of the passive signal is very low so that σ_z is quite high, the agent's optimal choices of signal precisions will not depend on the current state d_t anymore. This is because the passive signal z_0 contains almost no information about d_t before the agent chooses her information set. As a result, the agent will not be able to choose different precisions according to the realization of d_t. This result is also shown in Figure 16 as the case for $\sigma_z = 0$. Moreover, in this case, the agent will not necessarily choose a very noisy signal about the current state. The optimal precisions will depend on the prior mean of the agent, which is $\hat{\boldsymbol{X}}_0 = \boldsymbol{0}$ as in the baseline results. Such a pattern then has an important implication: the weights on signals will not only depend on the realized current state of the economy d_t, it will also depend on the prior mean that the agent carried on across time. I will illustrate how the optimal weights change with the prior mean in the next subsection.

Figure 16: Weights on Signals: Noise on Passive Signal σ_z



Left panel: model implied weights on *current signal* as function of actual economic condition d_t. Right panel: model implied weights on *SPF signal*. Each set of weights corresponds to a different standard error of noise in the passive signal σ_z. The baseline results come from $\sigma_z = 0.18$.

E.7.2 Impact of "Internal State": Prior Mean

Intuitively, when an agent chooses the information set she uses to form expectations, her ex-ante belief about the future state should matter. This can be seen directly from (30): when the agent thinks about future state d_{t+1} before she chooses information set that will generate $\boldsymbol{Z}_t$, her effective prior mean should be:

$$\hat{\boldsymbol{X}}_{t|0} = (I - K_0 G_0)\hat{\boldsymbol{X}}_0 + K_0 z_0 \tag{67}$$

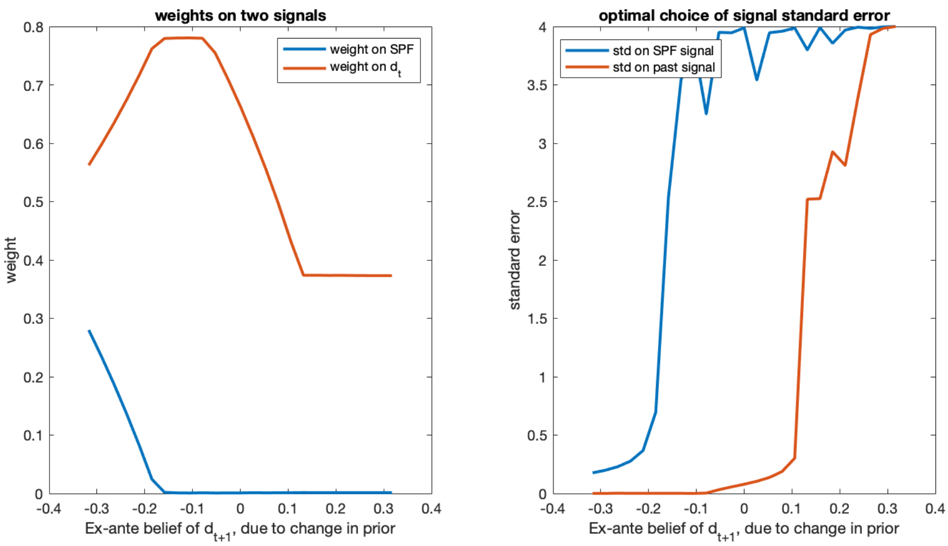
From previous sections, I have shown that the current state of economy d_t will affect her choice of optimal precision through the passive signal z_0. For the same reason, the optimal precision on signals should depend on the prior mean $\hat{\boldsymbol{X}}_0$ as well.

The illustration of the impact of prior mean involves several parts. First I want the change of optimal precision to come solely from the differences of $\hat{\boldsymbol{X}}_0$, so I will keep z_0 at a fixed value. Secondly, the reason why the prior mean will affect the optimal precision choice is that the agent will use the information set $\{\mathcal{I}_0\} = \{\boldsymbol{X}_{t|0}\}$ to evaluate her expected utility. The prior mean vector $\hat{\boldsymbol{X}}_0$ will affect this information set thus affecting the agent's expected benefit for any precision level. As discussed in section E.3, when the prior makes the agent believe on average the future state will be worse, she will choose a signal will higher precision. A straightforward way to illustrate this point is to depict the optimal weights and standard errors of the signals as functions of the implied posterior mean on d_{t+1} using the ex-ante information set $\mathcal{I}_0$. For simplicity, I call this "ex-ante belief on d_{t+1}", defined as:

$$\mathbb{E}[d_{t+1}|\mathcal{I}_0] = \iota A\Big((I - K_0 G_0)\hat{\boldsymbol{X}}_0 + K_0 z_0\Big) \tag{68}$$

In Figure 17, I show the optimal weights and standard errors as function of $\mathbb{E}[d_{t+1}|\mathcal{I}_0]$, while fixing $z_0 = 0$ and $d_t = 0$. This means the variation of the ex-ante belief comes solely from the differences in $\hat{\boldsymbol{X}}_0$.

Figure 17: Weights and Standard Error of Signals: Functions of Prior Beliefs



Left panel: model implied weights on current and future signal as functions of prior mean beliefs on the future state. Right panel: model implied standard error of the noises attached to current and future signal. The blue curves are for future (SPF) signal and red curves are for current signal.

Figure 17 shows that the optimal choices of weights (left panel) and precision (in terms of standard error, right panel) indeed depend on the agent's prior belief. In particular, when the prior belief leads to on average a bad state in the future, the agent will first pay more attention to the current signal, then shift to SPF signal as the implied state getting worse. This piece of evidence is also consistent with my empirical finding. As the prior mean is accumulated from the history of signals and usually not observable in the data, it can then be thought of as a proxy of the "internal state" in my empirical section. As discussed in Section 4.2.4, both the current state of the economy and the internal state accumulated from the past signals play a role in creating the state-dependent marginal effects of signals.