# Inflation Expectations over the Life Cycle under Rational Inattention

**Jessica L. Schultz**

Professor Francesco Bianchi, Faculty Advisor

---

*Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.*

**Duke University**
Durham, North Carolina
April 2021

# Acknowledgements

I would first like to extend my deepest gratitude to my advisor and mentor, Dr. Francesco Bianchi. His constant encouragement of my intellectual questions and interests empowered me to embark on this three-semester project—and enjoy it every day. I am incredibly grateful for his help and sage advice. I would also like to thank Dr. Michelle Connolly for her mentorship and support over the last three years, both as a research mentor and as my thesis seminar instructor. In addition, I am grateful to Dr. Cosmin Ilut for his guidance and input in my theoretical model, and I am truly indebted to Dr. Connel Fullenkamp and Dr. Kyle Jurado, both of whom helped inspire my interest in inflation expectations and my future career in research. This thesis also owes special thanks to Beepul Bharti for his help with numerical programming, my dear friends Alexa Putka and Laudan Hatami for their patience and encouragement along every step of the way, and my friends and family for their unwavering support.

# Abstract

This paper explores how people track inflation over their lifetimes while facing tradeoffs between attention and certainty. It first employs a flexible modification of the Recursive Least Squares Learning approach from Malmendier and Nagel (MN) (2016) to find that households place weight on each inflation observation in a hump-shaped pattern over age when using past observations to set expectations about the future. This finding departs from MN, which models a strictly increasing weighting scheme with age. This paper then uses these findings to motivate a theory of Rational Inattention (RI) in inflation: as households age and accumulate wealth, their knowledge of the inflation rate becomes more important in their financial decisions—so they pay more attention to inflation. Consequently, as they decumulate wealth during their retirement, they have less reason to track inflation as accurately.

This paper subsequently formalizes this theory in a two-period RI model in which inflation-driven uncertainty in the interest rate between a working period and a retirement period can be reduced at a cost; this reduction in uncertainty occurs through observing an endogenously chosen signal that is correlated with the interest rate. It finds that as wealth increases before retirement, the optimal choice of signal precision increases as well. These findings help explain the hump-shaped weighting scheme for inflation observations in the empirical section, assuming changes in these weights over age are related in part to changes in household wealth. Ultimately, these findings suggest that monetary policy that focuses on long-term inflation stability or accounts for this heterogeneity may be most effective in anchoring consumer inflation expectations and increasing consumer welfare.



*JEL Classification:* E, E21, E31, E43, E70, C32, G51.
*Keywords:* Inflation Expectations, Rational Inattention, Life Cycle.

# 1 Introduction

How do people track inflation over the course of their lifetimes? A household's expectation of future inflation impacts its lifelong investment and consumption decisions, yet the amount a household invests may also determine how informed it must be about inflation. A young professional who has not yet started saving for retirement, for example, may have little interest in tracking the real rate of return on assets because her asset holdings are very low. Similarly, a retiree no longer interested in saving for the future may not be very concerned about the real interest rate. On the other hand, a 50-year old who holds many assets and is saving a large portion of her income has a high stake in the real rate of return, for it determines how much these assets will be worth in the future. This paper will empirically examine how agents learn about inflation over the course of a life cycle, and consequently, how they set expectations about future inflation. It will explore how this learning pattern mirrors asset accumulation over the life cycle and subsequently provide a theoretical model integrating inflation learning with a Rational Inattention model.

The cross-sectional distribution of inflation expectations provides insight into heterogeneity among households; moreover, the average of consumer inflation expectations over time is important because it reflects the state of the macroeconomy, as perceived by households/consumers; expectations that are stable over a long period of time suggest confidence in the monetary authority, whereas uncertainty or instability in these expectations can sometimes be attributed to general economic uncertainty.

Consumer expectations are important in part because monetary policy often depends on and reacts to them (Sargent, 1982). In fact, due to an increasingly convoluted financial system, efforts of a monetary authority to control credit and other monetary policy tools are less effective than policy actions to manage expectations. Furthermore, stickiness in inflation expectations is not factored into many existing macroeconomics models (Mankiw and Reis, 2006), though it is empirically observed.

Over the past few decades, the Federal Reserve has made an effort to "anchor," or pin down, expectations to their inflation targets in an effort to strengthen the efficacy of monetary policy tools. The strategy to anchor expectations began nearly 30 years ago as a key component of inflation targeting. At that time, inflation was relatively high (above 4%), but it had followed an even higher inflation regime where inflation levels reached double digits. Today, inflation is much lower, and often very close to zero. Nevertheless, in both cases of higher- or lower- than desirable inflation levels, it is useful for monetary policy purposes to understand how exactly people set their expectations, given how much expectations can affect actual inflation.

Furthermore, heterogeneity in consumer expectations are consequential because they impact differentials in household financial decisions and contract negotiations. A consumer who consistently underestimates inflation, for example, may unknowingly accept a real wage cut over time. Thus, examining both aggregate expectations and cross-sectional expectations can shed light on both the general economic state and the distribution of beliefs across consumers.

One effort to highlight such cross-sectional expectations is an analysis of how expectations change along the age dimension. The "learning from experience" approach to expectations from Malmendier and Nagel (2016) estimates that individuals place lower weights on all observations as they grow older. On the other hand, they still place the most emphasis on their most recent observations, regardless of their age. This conclusion is also intuitive; it suggests that younger individuals are more impressionable to the economy around them, whereas older individuals take more experiences from previous monetary regimes into account.

While evident that agents set their inflation expectations differently over the course of their lives, there is less discussion about why that is the case. Is there some behavioral or biological reason for these changes in learning? Or, are there other age-related factors that cause people to learn differently over the course of their lifetime? If the latter, are these age-driven changes in learning rational? This paper proposes that a household's financial decisions over the life cycle directly impact attentiveness to inflation, which is endogenously chosen. More, this attention level to a given observation of inflation is not only rational, but also corresponds to how much is learned from that particular observation.

# 2 Literature Review

Relatively recent macroeconomic models have incorporated deviations from rational expectations in their formulations. The Limited Attention literature, for example, argues that agents can be resistant to update or absorb information (Mankiw and Reis, 2002; Woodford, 2003), especially considering how far in proximity the macroeconomy is from an individual's day-to-day activities.

A similar constraint that allows for rationality but instead focuses the costs of obtaining information is the Rational Inattention (RI) theory. Here, agents may tolerate incorrect (but not behaviorally biased) estimates of economic information under rational behavior. Essentially, agents receive a sequence of noisy signals but have the opportunity to increase the precision of these signals at some cost. Thus, agents choose whether they prefer to gain more information at a cost, or whether they would rather tolerate higher uncertainty (Sims, 2003). The RI approach is appealing because it can be integrated with the existing (neoclassical) macroeconomic model by merely adding a capacity constraint (which limits the amount of information that can be processed by the agent) and an information-related choice variable used in the optimization. Furthermore, an RI explanation for economic aggregates is especially intuitive because paying attention to such distant economic variables is likely a lower priority to households than other more immediate decisions. Of course, the amount of attention paid to these economic aggregates may vary depending on overall economic stability or even the degree to which these variables affect an individual's wealth or finances. These choices are explored in more detail in Section 4.

Moreover, the Adaptive Expectations models use past observations (with some degree of discounting) to set future inflation expectations. One approach that somewhat resembles adaptive expectations is Recursive Least Squares (RLS) learning, which similarly uses past observations to form expectations; however, it instead takes a perceived law of motion (PLM) for an economic aggregate as given and "adaptively" estimates the *parameters* of the PLM, rather than the economic aggregate itself. The agent then sets expectations based on the estimated parameters from this RLS learning process.

The RLS approach to learning is a particularly studied case because its learning models converge to rational expectations (Evans and Honkapohja, 2001). Additionally, when the weights assigned to each observation from the past are equal, the RLS estimates becomes an Ordinary Least Squares estimation. When the RLS approach is employed under a constant updating parameter, defined as a "constant gain," more recent observations have higher weight, which supports the theory that economies are subject to structural change, so older observations become less relevant over time (Sargent, 1999).

Under the "learning from experience" approach employed in Malmendier and Nagel (MN) (2016), the RLS model is used to examine age-driven heterogeneity in expectations. Unlike in the constant gain approach, the learning-from-experience gain is not a constant parameter, but rather, a function inversely related to age. The estimation of this model still finds that older observations are less influential than newer observations in forming current expectations. However, the MN gain finds that higher weights are placed on older observations than the constant-gain literature would predict. In other words, individuals discount observations more slowly than the speed at which structural changes occur in the macroeconomy. Furthermore, the MN weighting scheme is derived from a parsimonious empirical specification that only allows weights placed on observations to peak in the current period or in the beginning of the life cycle. Thus, the model does not allow the most influential observations to occur in the middle of a lifetime.

A gain specification that allows the weighting function to peak during ones' middle age (40-60) could be particularly useful because household financial allocations often change over the course of their lifetimes, borrowing more in their early years, supplying more credit during their middle ages, and consuming more from savings in later years. A flexible parameterization can thus pick up possible effects on expectations from financial decisions, which are highly correlated with age.

Section 3 will suggest an alternative specification of the RLS gain in order to accommodate any such peak in the weighting function during the middle of an individual's life. Following this, Section 4 will present a theoretical model that uses RI as a possible explanation for age-driven heterogeneity in the RLS gain through changes in income over the life cycle.

# 3 The Learning Model

Under the RLS "learning" approach, agents set expectations about a given variable based on their beliefs about its law of motion. These beliefs are refined each period as they encounter new observations. The following sections introduce the learning model.

## 3.1 Recursive Least Squares Model

First, let time be discrete $t = 0, 1, \ldots$ To examine a cross section of individuals, consider a person living at time $t$ in cohort $s$; then at time $t$, her age is $t - s$. Following Malmendier and Nagel (MN) (2016) and Orphanides and Williams (2003), suppose that individuals set inflation expectations according to the AR(1) Perceived Law of Motion (PLM):

$$\pi_{t+1} = \alpha + \phi \pi_t + \eta_{t+1}, \tag{1}$$

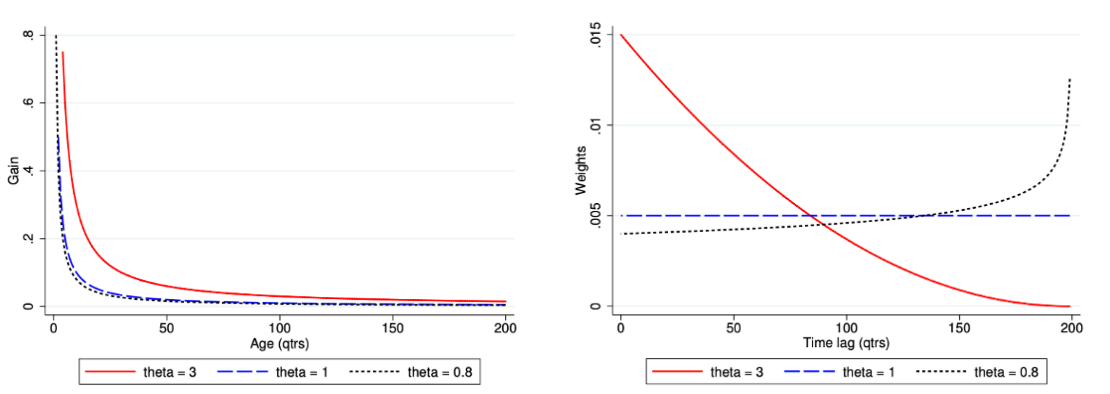
where $\eta_{t+1} \sim N(0, \sigma^2)$ is a noise term at the time of observation.

Under adaptive learning, individuals learn the values of $\alpha$ and $\phi$ simultaneously as they observe new inflation observations. At each new observation, they estimate $b \equiv (\alpha, \phi)'$ recursively by comparing their past expectations to the corresponding realized inflation levels of that period. The recursive expression of their estimation follows a Recursive Least Squares model:

$$b_{t,s} = b_{t-1,s} + \gamma R^{-1} x_{t-1}(\pi_t - b'_{t-1,s} x_{t-1}) \tag{2}$$

$$R_{t,s} = R_{t-1,s} + \gamma_{t,s}(x_{t-1}x'_{t-1} - R_{t-1,s}), \tag{3}$$

where $R_{t,s}$ is a 2 x 2 matrix of stacked coefficients. The expression $\pi_t - b'_{t-1,s}x_{t-1}$ is the difference between the actual inflation at time $t$ and the expected inflation for time $t$ using the learned parameters from time $0, 1, \ldots, t-1$ and the inflation level at time $t-1$. The parameter $\gamma_{t,s}$ is the "gain" for cohort s at time t. In other words, for a cohort $s$ at time




(a) Gain under MN specification for different values of theta

(b) Implied Weights under MN specification for different values of theta

Figure 1

$t+j$, where $j > 0$, the gain $\gamma_{t,s}$ represents the sensitivity of inflation expectations for time $t+j+1$ to an inflation surprise that happened at time $t$. The gain coefficient also has a mathematical relationship with the implied weights placed on observations over the lifetime. These weights are the degree to which each observation in an individual's history of inflation experiences can influence the present expectations. In other words, while the gain assigns a sensitivity in the present period, the history of gain values determine how much the past plays a role in current expectations. The direct relationship between the gain and weights can be found in the Online Appendix of MN (2016). Essentially, as the gain increases for any given observation, its weight also increases when determining its impact on all subsequent expectations.

## 3.2 The MN (Decreasing) Gain

Malmendier and Nagel (MN) (2016) parametrically specify the following decreasing gain:

$$\gamma_{t,s} = \frac{\theta}{t-s} \text{ for } 1 \leq t-s \leq T \tag{4}$$

where $T$ is the maximum age reached by the agent in her lifetime. The model also imposes the additional constraint that $\gamma_{t,s} \leq 1$ for all $t$, which follows from the assumption that agents do not overcorrect their expectations in response to a surprise. Age is in the denominator of this equation, indicating that as an individual grows older, their recent observations play a smaller and smaller role in informing their current expectations. This does not necessarily mean that the weights on recent observations are lower than older observations; this fact is determined by the value of $\theta$. This parameter, $\theta$, is a constant that determines the shape of the implied weighting function. Figure 1 can help illustrate how different values of $\theta$ can influence both the gain function and the weighting function. Figure 1a (from MN, 2016) show how different values of $\theta$ impact the gain, and 1b (from MN, 2016) show how these values impact the implied weights, where the horizontal axis is the number of time lags. MN estimates the value of $\theta$ to be 3.044, which behaves similarly to the functions in red in Figures 1a and 1b.

Furthermore, the MN case assumes that learning from experiences accounts for only part of inflation expectations. In other words, inflation expectations have a cohort-specific learning-from-experience component, $\tau_{t+1|t,s} \equiv b'_{t,s} x_t$ (interpreted as the inflation expectation at time $t+1$, learned from equation (2), given an observation at time $t$), as well as a common component, $f_t$, which is available to all cohorts at a given time. The common component "could be based on any kind of common information available to all individuals at time $t$, such as the opinion of professional forecasters or the representation of their opinions in the news media," or it "could capture a component that is based on all available historical data" (MN, 2016). If agents set their inflation expectations based in part on their experience and in part on this common component, they can be represented as follows:

$$\pi_{t+1|t,s} = \beta \tau_{t+1|t,s} + (1-\beta) f_t \tag{5}$$

where $0 \leq \beta \leq 1$, and $\beta$ represents an agent's "sensitivity" to observations from their experience. MN estimates this sensitivity parameter to be 0.672.

In summary, agents set their 1-year inflation expectation partly from commonly available information at the time, and more significantly, from their own inflation observations. The extent to which their expectations are updated depends on their age as well as how surprised they are by a recent inflation event.

### 3.3 A Non-parametric approach

While the results from the MN case indicate that people place greater weights on early observations than under the constant gain approach, the model is parsimoniously specified and consequently does not allow for the weights to peak in the middle of a lifetime, as evident by Figure 1b. However, due to financial changes over the course of the lifetime, it would be reasonable to expect a non-monotonic change in the weighting function over the life cycle. To investigate the possibility of a more nuanced function affecting learning, I employ a non-parametric approach to ascertain whether the MN (decreasing) gain scenario is truly appropriate, given historical data. Additionally, a non-parametric approach can help determine *qualitatively* whether adding additional parameters could be useful in mapping the gain function. In specifying the non-parametric model, I rewrite equation (4) as:

$$\gamma_{t,s} = \frac{\theta_{t-s}}{t-s} \text{ for } 1 \leq t-s \leq T, \tag{6}$$

where $\theta_{t-s} < t-s$ for all $t, s$. This is essentially the same expression as the MN gain, but with a key distinction: $\theta$ is allowed to vary over time. While this would be equivalent to merely estimating $\gamma_{t,s}$ nonparametrically, the gain in (6) nests nicely within the MN specification and thus allows for a more direct comparison with the MN results.

Mechanically, we can model their updating process using the RLS equations (2) and (3) as they learn the law of motion for inflation, while weighting each new observation in a way consistent with the gain specification in (6). We can examine the sensitivity of reported consumer expectations to the learning-from-experience expectations compared to other sources influencing inflation expectations by estimating the sensitivity parameter $\beta$ in equation (5). The next few sections will estimate this model.

## 3.4 Survey and Macroeconomic Data Sources

The data used in the estimation are taken directly from the MN replication files. The raw data are the 1-year forward consumer inflation expectations from the Michigan Survey of Consumer sentiments, which span the period 1953-2009. The survey data are expectations of individuals, but the dataset is then aggregated by cohort. This isolates heterogeneity in expectations to age-driven factors only. It also allows for a longitudinal analysis, in contrast to the raw data which are purely cross-sectional. Furthermore, only observations from ages 25 (the youngest age in the survey) to 75 are used. This is because observations after age 75 are sparse, so the aggregated data are much noisier and driven by individual responses than in the rest of the dataset. Following the MN approach, I assume the life cycle ends at age 75.

The actual inflation levels observed in the economy are also from the MN replication files, spanning from 1872-2009, and these observations are used in the recursive learning estimation. Other details about the dataset can be found in the data section of MN (2016) as well as their online appendix.

## 3.5 Joint Estimation of Parameters

Empirically, I estimate equation (5) following the MN approach with the following panel regression:

$$\tilde{\pi}_{t+1|t,s} = \beta \tau_{t+1|t,s} + \delta_t D_t + \epsilon_{t,s} \tag{7}$$

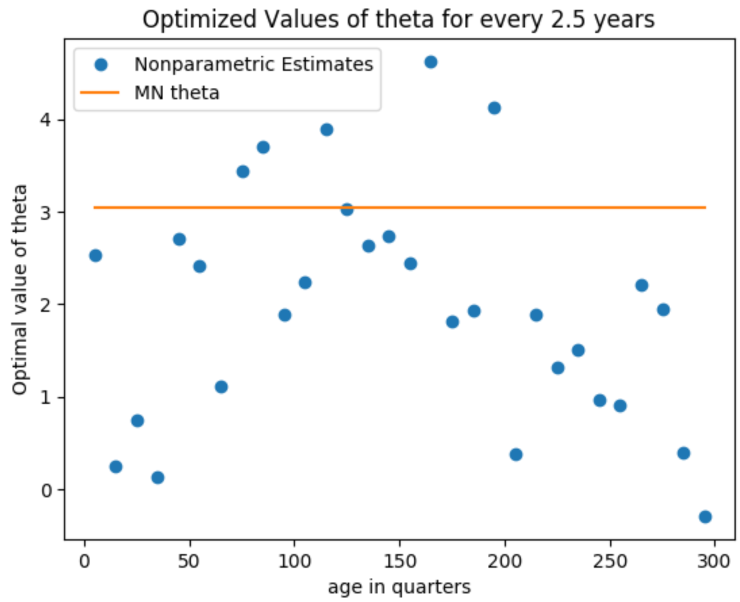



Figure 2

Here, the response variable $\tilde{\pi}_{t+1|t,s}$ is an aggregated 1-year survey expectation from the Michigan Survey of Consumer Sentiments, specific to a single cohort at a single time period. The indicator variable $D_t$ is included to allow for time fixed effects, which also absorbs the common component $f_t$ of expectations.

The time fixed effects in the panel regression absorb any trends that affect all observations over time, allowing us to assume that age-related weighting trends are independent of time. Moreover, by exploiting time variation across different cohorts, we can isolate weighting trends that are independent of the cohort. The parameters of interest are the gain parameters in (5), which do not appear directly in (7), but rather, influence the RLS-predicted expectation $\tau_{t+1|t,s}$. Furthermore, because the RLS-predicted expectations depend on a choice of gain parameters, the parameters in (7) must be estimated simultaneously with the gain parameters. This requires a two-stage nonlinear approach. Algorithmically, an inner optimization performs the RLS estimation for $b_{t,s}$ as a function of the gain parameters from (5), while an outer optimization finds the optimal gain parameters and the sensitivity to learning $\beta$ through the panel regression from (7).

Using this two-stage optimization approach for the nonparametric gain, I jointly estimate a sequence $\{\tilde{\theta}_i\}_{i=1}^{T/10}$ of the gain parameters from (6), where $\theta_i$i represents the optimal value of $\theta_{t-s}$ over a 10-quarter period. Figure 2 charts these optimal values as a function of age.

From Figure 2, the non-parametric approach indicates that the true values for $\{\theta_{t-s}\}$ have some bell-shaped pattern with respect to age. This suggests that there is a gradual structural shift in learning over the life cycle, where the initial structure of learning resembles the MN case, but it evolves such that recent observations may no longer be more important than earlier ones. To help visualize this phenomenon, I specify a more parsimonious function for $\theta_{t-s}$ that adds two additional parameters to the MN case but conserves many degrees of freedom compared to the nonparametric approach.

Nonetheless, there is a high level of noise in the optimal values, which indicates that adding two gain parameters, albeit bringing about an improvement in fit and likely removing some bias, may not result in a statistically significant improvement. The subsequent section employs a parametric approach to estimate the gain parameters in (4) and examine the implications of this "life cycle gain" specification.

### 3.6 A Life Cycle Gain

Much like the nonparametric scenario, the following specification maintains the decreasing-gain denominator to keep it analogous to the MN case. This life cycle gain fits a function for $\theta_{t-s}$ to the data such that weights on more recent observations are higher than past observations until a certain point in the life cycle. After this point, a structural change occurs and weights on more recent observations begin to decrease with age. Because the gain function is flexible, if the bell-shaped function of $\theta_{t-s}$ in the nonparametric estimation seen Figure 2 is less desirable than the baseline MN case, the parameter values will adopt values resembling the MN results.

The life cycle gain is specified as follows:

$$\gamma_{t,s} = \frac{\alpha e^{-(\nu(t-s-\eta))^2}}{t-s} \text{ for } 1 \leq t-s \leq T \tag{8}$$

Under this specification, three parameters ($\alpha$, $\nu$, and $\eta$) are used in the gain, so it is flexible compared to the more parsimonious one-parameter approach in MN).

### 3.7 Estimation

Following the same 2-stage optimization algorithm detailed in the nonparametric case, we can estimate the gain parameters in (8) and recover the weighting scheme. The results are listed in Table 1.

These estimations indicate that the weights assigned to older observations are still lower than more recent observations, until a person is in their early 50s. The nonzero value of $\nu$ indicates that the weighting of observations peak at some age, then start declining. Figure 3 shows the implied weights from the estimated gain parameters at various ages.

While the weighting trends for ages 10-40 years show similar behavior to the strictly decreasing gain scenario in MN, the weights for ages 50-70 years show how this weighting

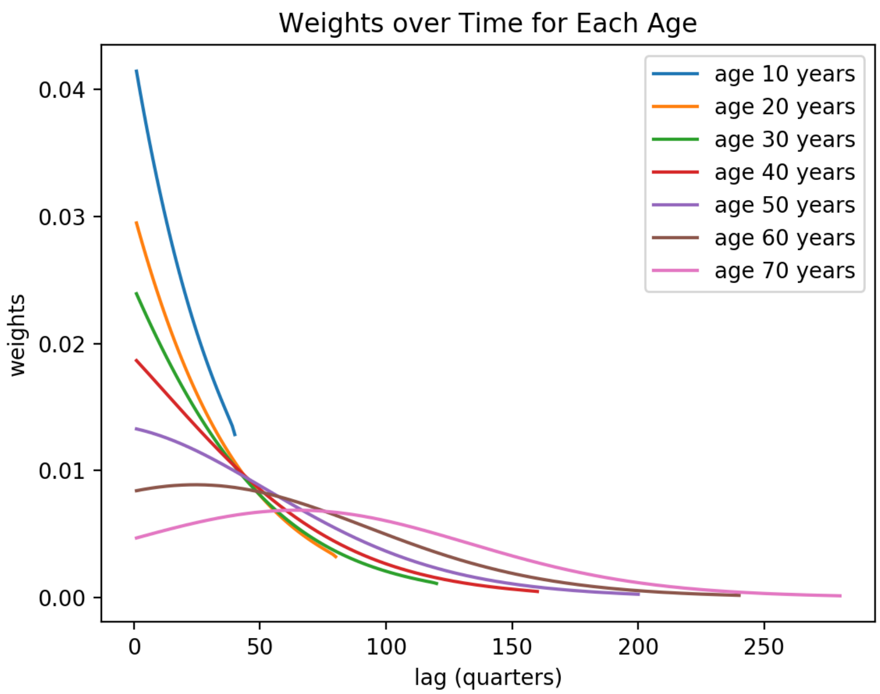



Figure 3: Implied weights of past inflation observations over time lags, separated by age. The leftmost points are the most recent, whereas the rightmost points are the first observations of a person's life.

| Estimation of Parameter Values | | |
|---|---|---|
| Parameter | Estimate | Standard Error |
| $\alpha$ | 3.025 | 0.724 |
| $\nu$ | 0.0073 | 0.0022 |
| $\eta$ | 157.4 | 4.310 |
| $\beta$ | 0.713 | 0.085 |

Table 1: Estimation of gain parameters and sensitivity $\beta$ from 2-stage optimization.

scheme shifts into a more complicated hump-shaped pattern as agents place lower and lower weights on later-life events as they age past 50 years.

These results indicate that there is a more complicated life cycle pattern of inflation learning, most notably with a structural break in learning in the early 50s.

## 3.8 Implications of a Life-Cycle Gain

These findings suggest that there is a structural shift in the way expectations are set over the life cycle; first, people increase their attention to events as they age, then during their 50s, they gradually decrease their attention. Interestingly, the shape of the weighting function resembles that of wealth accumulation over the life cycle. A possible explanation for this behavior is that consumers are rationally inattentive; Rational Inattention (RI) in inflation would suggest that individuals should pay more attention to inflation when their loss from incorrectly estimating it overshadows the cost of obtaining information. Under this framework, we should similarly see a peak in the weighting of inflation when individuals are most engaged in saving and investing, which would likely occur during the middle of the life cycle. Empirical studies have similarly found that knowledge about the economy and precision in expectations is strongly correlated with wealth, further evidence of this line of reasoning. One such study found that wealthier households rely more on their own private wealth accumulation and consequently accumulate more financial knowledge endogenously, whereas households with less wealth have lower overall financial knowledge. This endogenous accumulation of financial knowledge over the life cycle parallels the trend of wealth accumulation and savings. (Lusardi, Michaud, and Mitchell, 2017).

Investigating an empirical model that parses out the individual effects of wealth and age on the RLS gain would be challenging due to the fact that existing datasets are not longitudinal, so individuals cannot be followed over time. Similarly, aggregating data by both age and asset ownership may also be difficult due to both the high correlation in the data between these two variables as well as the lack of detailed cross-sectional variation in asset holdings within the Michigan Survey data.

On the other hand, a theoretical model for the RLS gain using RI theory could prove a useful thought experiment by providing both a microfoundation and an intuitive explanation for age-driven heterogeneity in learning. In the following section, I present an RI model where uncertainty in the inflation rate can be reduced at some cost. Under the assumption that some of these variations in implied weights over the life cycle are a result of the natural accumulation of wealth before retirement, the gain is in fact related to an individual's wealth. Furthermore, a RI model that finds differences in attention to inflation by wealth level could help explain the results of this life-cycle gain.

# 4 A Rational Inattention Approach

In this section, I present a two-period model of Rational Inattention (RI) to illustrate how planning for retirement can influence the allocation of attention toward inflation. Following economic modeling conventions, the agents in this model are "households" rather than the "individuals" from the RLS Learning model in Section 3. However, the two are analogous and can be interpreted similarly.

The model builds on RI theory from Sims (2003) and applies the optimization approach of Maćkowiak and Wiederholt (2009) to the consumer problem. The model further provides a partial equilibrium solution where households take the interest rate and any other prices as given, then make their consumption and attention decisions accordingly.

In the model, the household has income in the first period and is retired in the second. The finite horizon of a two-period model can help intuitively explain the life-cycle gain because it allows us to see simply how differences in income before retirement endogenously affect attention. Further, this model will assume that households have no motivation to bequeath their wealth to their children, and instead, want to consume as much of their wealth as possible before the end of their life cycle. While not entirely realistic, this assumption can help isolate the part of the consumption-saving decision that is attributed to future uncertainty and income rather than preferences for the well-being of future generations. The subsequent sections will outline the household's problem in more detail.

## 4.1 Utility and Income

Like in standard macroeconomic models, the household makes all of its choices in order to maximize its expected lifetime utility. In this section, I allow the utility to be modeled by a power utility function, such that for some level of consumption $c$, the utility associated with that value is:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where $\gamma$ is a parameter such that $\gamma > 0$. Under this utility function, the household is risk averse, so it will be willing to pay to reduce uncertainty about the future. This type of behavior in response to risk aversion is known as precautionary savings. Essentially, uncertainty causes households to save more than they would under certainty in order to hedge against lower-than-expected returns in the future. The household's expected lifetime utility is the sum of its utility from consumption in each period, with the discount factor between the two periods denoted by $\beta$. For the two-period model, this can be expressed as:

$$\mathbb{E}(U) \equiv \mathbb{E}\big[u(c_0) + \beta u(c_1)\big],$$

where $c_0$ is the consumption level during the working period ($t = 0$), and $c_1$ is the consumption level during the retirement period ($t = 1$). The household earns income in period 0, which I denote $Y_0$, and it can be thought of as endowment or labor income. In the two-period model, this is essentially wealth from income earned before retirement. Additionally, the income can be consumed in the current period or saved for retirement. If the income is saved, it grows at a rate $R$ between the working period (period 0) and the retirement period (period 1); in other words, $R$ is the total rate of return.

Further, I specify that the real rate of return between time $t = 0$ and $t = 1$ is stochastic. However, the *nominal* rate of return between these two periods is known. Thus, uncertainty in the real rate can be entirely attributed to the lack of knowledge about inflation between the two periods at $t = 0$. As a result, I allow households to employ RI to reduce this inflation-driven uncertainty by paying more attention to indicators of the inflation rate.

## 4.2 Signal Structure

Suppose there is some inflation-driven uncertainty in the interest rate such that the rate of return $R$ has the following distribution:

$$R \sim \text{Lognormal}(\mu, \sigma^2),$$

Alternatively, this can be represented as:

$$R = e^r, \text{ for } r \sim N(\mu, \sigma^2)$$

In this representation, $\log(R) = r$ can be thought of as the net rate of return (like the interest rate) between the working and retirement periods. If we assume that $\mu$ is merely some function of the nominal interest rate, then $\sigma^2$ is uncertainty directly attributed to inflation.

Now suppose that there is some signal $S$ (a random variable with realization $s$) that covaries with $r$ such that:

$$S = r + \epsilon,$$

where $\epsilon \sim N(0, \sigma_\epsilon^2)$ is some observation noise. Then, we know the distribution of $S$ conditional on $r$ is:

$$S|r \sim N(\mu, \sigma^2 + \sigma_\epsilon^2).$$

The precision of a distribution is merely the inverse of its variance. Thus, let $\rho_r$ and $\rho_\epsilon$ denote the precision for the distributions of $r$ and $\epsilon$ respectively. Then

$$\rho_r = \frac{1}{\sigma^2} \text{ and } \rho_\epsilon = \frac{1}{\sigma_\epsilon^2}.$$

The value $\rho_\epsilon$ is henceforth referred to as the *"signal precision"*. Using a standard Bayesian mean squared error estimator for normally distributed variables, we can find that the conditional distribution (the posterior distribution) $r|S$ of the net rate of return $r$, given a signal realization $s$, is characterized by the conditional precision,

$$\rho \equiv \rho_r + \rho_\epsilon$$

and the conditional mean,

$$\begin{aligned}\hat{\mu} &\equiv \frac{\rho_r}{\rho}\mu + \frac{\rho_\epsilon}{\rho}s \\ &= \frac{\rho_r\mu + \rho_\epsilon s}{\rho},\end{aligned}$$

where $s$ is some realization of signal $S$ that can be observed before the real rate of return is known.

Putting this all together, we can now describe the distribution of $r|S$, the probability of the net rate of return being a certain value, conditional on seeing some signal realization $S = s$. The distribution is:

$$\hat{r} \equiv r|S \sim N\Big(\hat{\mu}, \frac{1}{\rho}\Big),$$

and letting $\hat{\sigma}^2$ represent $1/\rho$, we can express this as:

$$\hat{r} \sim N(\hat{\mu}, \hat{\sigma}^2)$$

Now we know the distribution for the net rate of return, and consequently, the real rate of return upon observing a signal. However, In the RI problem, the precision of the conditional distribution $\rho$, and implicitly, the precision of the signal $\rho_\epsilon$, are chosen optimally.

How can we measure the tradeoff between paying more attention to find a more precise conditional distribution (a better signal), and paying less attention and consequently facing more uncertainty? To do this, I introduce a capacity constraint that can be relaxed at some cost.

## 4.3 Capacity and Budget Constraints

To quantify how much information is gained from choosing a particular conditional distribution, I use the mutual information between the unconditional and conditional distributions of $r$. Letting $H$ denote the entropy function of a random variable, we have the information flow equation:

$$H(r) - H(r|S).$$

Here, an increase in the precision of $r|S$ results in a lower entropy, which increases the differential entropy expressed above. The greater the difference, the more information is gained from the signal.

Under RI, households have a constraint on their information capacity, denoted $\kappa$, where the information flow must always satisfy:

$$H(r) - H(r|S) \leq \kappa.$$

Now, we can see that a better (more precise) signal can be found up to a certain extent, depending on the value of this capacity constraint. Because both $r$ and $r|S$ are normally distributed, the values of the entropy functions are given by $H(r) = \frac{1}{2}\log(2\pi e\sigma^2)$ and $H(r|S) = \frac{1}{2}\log(2\pi e/\rho)$. Additionally, under optimality, the household will fully utilize its information capacity, so the information constraint becomes:

$$\tfrac{1}{2}\log(\sigma^2\rho) = \kappa. \tag{9}$$

Finally, I allow the information constraint $\kappa$ to be chosen optimally based on the choice of $\rho$. So, a household may choose to relax its constraint by increasing $\kappa$, and consequently incur some cost. Similarly, a household may choose to tighten its constraint by decreasing $\kappa$, and consequently incur some payoff from the attention it "saved" by not putting as much effort into tracking inflation.

I further specify that this attention cost will enter into the budget constraint, such that for every unit increase in capacity, the household must pay the fixed fee $\delta$, which is the cost of capacity in terms of the consumption good. Since the conditional variance enters into the attention cost as the inverse of the conditional precision, the marginal attention cost is hyperbolically increasing as the *variance* approaches 0. In other words, it becomes more and more costly to reduce uncertainty as your lower the noise of your signal. This phenomenon is illustrated in Figure 4.

Intuitively, we can think of paying attention to macroeconomic variables (like inflation) as some taxing endeavour that limits the amount of time spent earning income. Thus, the more you increase your capacity, the less money you have left to spend or save. Alternatively, we can think of the cost from the information capacity as some fee for allowing a professional advisor to manage your budget; a higher information capacity represents a more expensive advisor who is better at tracking the signal and telling you how much to save.

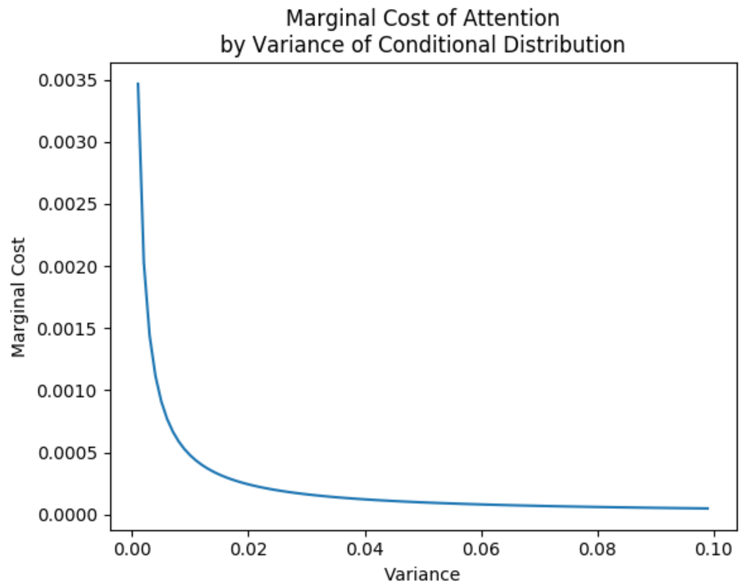



Figure 4: Marginal Cost of Attention when $\delta = 0.01$, a parameter value for the capacity cost used throughout the simulations in this paper.

Putting some of these components together, we have the following budget constraint for periods 0 and 1, respectively:

$$c_0 + B + \delta\kappa = Y_0$$

$$c_1 = RB,$$

where $B$ is the amount saved during the working period for retirement. Note that there is no income in period 1 because it represents the retirement period, and the information cost is only incurred during period 0, before the real rate of return is known. Combining equations, the 2-period budget constraint becomes:

$$c_1 = R(Y_0 - \delta\kappa - c_0). \tag{10}$$

## 4.4 The Household's Problem

Essentially, we want to model the household as choosing how precisely it wants to track inflation. However, to make its choice of precision, it must weigh how much is gained from a more precise estimate of inflation against the cost of obtaining the more precise estimate. In this model, the precision of its tracking is analogous to the conditional distribution of the signal, the amount gained is the increase in lifetime utility from a more informed consumption vs saving rule, and the cost is the capacity cost.

Furthermore, in the background of this Life Cycle RI model, the household first chooses how much it would consume in the working period and how much it would save for the retirement period, given any conditional distribution as well as any signal realization. Its choice for this consumption rule is based on its objective to maximize expected lifetime utility, subject to its budget constraints. Knowing its consumption rule as a function of any conditional distribution, it then chooses the conditional distribution whose associated consumption and savings choices are optimal. Finally, upon observing a signal realization, it chooses its consumption level for that time period, given the optimal consumption rule it previously found.

The household chooses its consumption rule (how much to consume vs. save), as well as its conditional precision $\rho$ such that it solves the following objective function:

$$\max_{\rho} \mathbb{E}\big[u(c_0^*) + \beta(c_1^*)|\rho\big] \tag{11}$$

subject to:

$$c_0^* = \arg\max_{c_0} \mathbb{E}\big[u(c_0) + \beta u(c_1^*)|\rho, s\big], \tag{12}$$

$$c_1^* = R(Y_0 - \delta\kappa - c_0), \tag{13}$$

and the budget and information constraints in (10) and (9), respectively. This follows the constrained optimization setup from Maćkowiak and Wiederholt (2009). The above problem can be interpreted in the following chronology:

1. Given a conditional precision $\rho$, and a signal realization $s$, the household determines its consumption rule subject to the budget and information constraints, or solves (12) given (13). Thus, the value $c_0^*$ is a function of the conditional precision and signal. Furthermore, the value of $c_1^*$ is determined based on the optimal decision for $c_0^*$.

2. Knowing this consumption rule as a function of any signal realization, the household then chooses the conditional precision $\rho$ that maximizes its expected lifetime utility in (11).

To solve the problem, I proceed following the above chronology. First, I solve the first step as a signal extraction problem, fixing the conditional precision. Second, I solve the second step as a RI problem, with precision as a choice variable.

## Step 1: A Signal Extraction Problem

In this section, I first solve for the values of $c_0^*$ and $c_1^*$ in equations (12) and (13), which are the optimal consumption decisions given some conditional precision and the signal realization.

These variables are thus functions of the conditional precision and signal, and the latter is a function of the realized rate of return. Equivalently,

$$c_0 \equiv c_0(s, \rho) \text{ and } c_1 \equiv c_1(s, \rho, R).$$

Additionally, I will calculate the expectations:

$$\mathbb{E}(c_0|\rho), \mathbb{E}(B|\rho), \text{ and } \mathbb{E}(c_1|\rho),$$

at time $t = 0$, where $B$ is the amount of savings carried over from time $t = 0$ to time $t = 1$. In other words, for a given level of precision of the conditional distribution, we want to know the expected consumption and saving levels, summing over all possible signal realizations. These expectations describe the consumption rule under the signal extraction problem.

**Solving the Household's Problem:**

We want to find the value of $c_0$ that solves:

$$\max_{c_0} \mathbb{E}\Big[u(c_0) + \beta u\big(R(Y_0 - \delta\kappa - c_0)\big)|\rho, s\Big]$$

$$\text{subject to:}$$

$$\tfrac{1}{2}\log(\sigma^2\rho) = \kappa \tag{14}$$

The optimality condition under power utility is

$$c_0^{-\gamma} = \beta\mathbb{E}\big[R^{1-\gamma}(Y_0 - c_0 - \delta\kappa)^{-\gamma}|\rho, s\big],$$

which can be simplified to

$$c_0 = \big[\beta\mathbb{E}(R^{1-\gamma}|\rho, s)\big]^{-1/\gamma}(Y_0 - c_0 - \delta\kappa).$$

Further details deriving this condition are included in the Appendix. In order to simplify the expectation of this function of $R$, we can use the following proposition:

**Proposition 1.** *(Proof in Appendix) Suppose $R$ is given by $R = e^r$, with $r$ normally distributed and $\widehat{R}$ is given by $\widehat{R} = e^{r|S}$, with $r|S$ distribution $r|S \sim N(\hat{\mu}, \hat{\sigma}^2)$. Then,*

$$\mathbb{E}(R^{1-\gamma}|\rho, s) = \mathbb{E}\big[(\widehat{R})^{1-\gamma}\big] = e^{(1-\gamma)\hat{\mu} + (1-\gamma^2)(\hat{\sigma}^2/2)}$$

This is a result from the exponential function when conditioning on information that affects the posterior distribution of the exponent. Using the above proposition to simplify our optimality condition, we have an expression without the expectation operator that is a direct function of $s$ and $\rho$.

$$c_0 = (e^{(1-\gamma)\hat{\mu}+(1-\gamma)^2(\hat{\sigma}^2/2)})^{-1/\gamma}(Y_0 - c_0 - \delta\kappa).$$

To simplify notation, let

$$Q \equiv e^{(1-\gamma)\hat{\mu}+(1-\gamma)^2(\hat{\sigma}^2/2)}. \tag{15}$$

Then $c_0$ can be rewritten as

$$c_0 = (\beta Q)^{-1/\gamma}(Y_0 - c_0 - \delta\kappa),$$

and we have our optimal choices for $c_0$ and $c_1$:

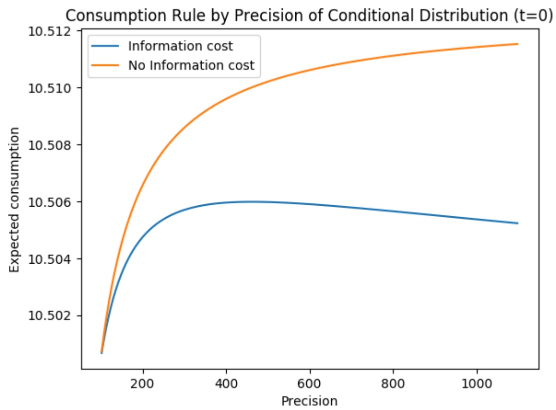
$$c_0 = \frac{(\beta Q)^{-1/\gamma}}{1+(\beta Q)^{-1/\gamma}}(Y_0 - \delta\kappa) \tag{16}$$

$$c_1 = R(Y_0 - c_0 - \delta\kappa) = R\left(1 - \frac{(\beta Q)^{-1/\gamma}}{1+(\beta Q)^{-1/\gamma}}\right)(Y_0 - \delta\kappa). \tag{17}$$

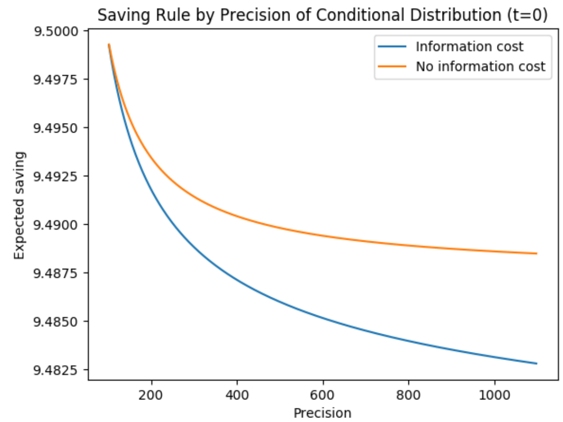
The fraction on the RHS of equation (16) in terms of $\beta$ and $Q$ is decreasing with conditional variance (or increasing with signal precision) and increasing with the signal. This means that for a given signal realization, as we increase the precision, consumption during the working period increases as well. We can think of the increasing precision giving more weight to the signal realization—a high signal realization under a high precision level signifies that there is a good chance the interest rate will in fact be high. As the precision lowers for that same signal realization, it becomes more and more likely that the high realization was just due to noise in the signal itself. Conversely, as the signal realization becomes higher and higher for a fixed precision, consumption in the working period increases because it is more likely that that the real rate of return will be high.

Thus, we can already see the effect of a household's precautionary saving motive, even before summing over all possible signal realizations. As the estimate of future inflation becomes more and more precise, uncertainty over future income is reduced. Since the household is risk averse, it engages in precautionary saving under uncertainty, and this effect diminishes with greater levels of precision. Thus, households don't have to worry as much about not saving enough when they are more confident about the future interest rate, so they can spend more money during their working period.

On the other hand, since we know that the capacity $\kappa$ is an increasing function of the precision, as the precision increases more and more, the capacity costs starts to lower the net income of a household. This means that as precision increases, eventually consumption




(a) Expected consumption with and without an information cost at time $t = 0$.




(b) Expected Saving with and without an information cost at time $t = 0$.

Figure 5: Expected consumption vs. saving when income/wealth in first period is $Y_0 = 20$, $\beta = 0.9, \gamma = 2$, $\delta = 0.01$, $\mu = 0.1$, and $\sigma = 0.1$. The expectations with an information cost use $\delta = 0.01$.

during the working period will decrease from a pure income effect as a result of money lost on increasing the information capacity.

These two effects work in opposite directions, and the subsequent section will build the numerical groundwork for the optimal choices of consumption over precision when summing over all signal realizations, or when taking into account the *expected* signal realization.

### Calculating Expectations

To find the expected consumption rule from the "signal extraction" problem, I calculate $\mathbb{E}(c_0|\rho)$ and $\mathbb{E}(c_1|\rho)$, which are expectations that integrate over all signal realizations, given a conditional precision. This can be interpreted as what the household expects to consume depending on the signal choice, but before actually seeing the signal realization. The expectation $\mathbb{E}(c_0|\rho)$ can be written as:

$$\mathbb{E}(c_0|\rho) = \int_{-\infty}^{\infty} c_0(s, \rho) f(s|\mu, \rho)\, ds, \tag{18}$$

where $f(s|\mu, \rho)$ is the normal density function for realizations of $s$, given the mean of signal realizations, $\mu$, and the precision of the conditional distribution, $\rho$. In other words, the function $f$ reflects the density of $\Pr(S|r)$, given the parameters in the distribution of $S|r$. Similarly, $c_0$ is expressed as a function of $s$ and $\rho$, which is based on the result in equation (16). This expectation can be calculated numerically, and the result, given various levels of the conditional precision, is illustrated in Figure 5a.

Similarly, we can calculate the expected saving for retirement, given a precision level before seeing the signal realization.

$$\mathbb{E}(B|\rho) = Y_0 - \delta\kappa - \mathbb{E}(c_0|\rho),$$

where the equation for $\mathbb{E}(c_0|\rho)$ is given in (18). This expectation can similarly be calculated numerically, and the result, given various levels of precision is graphed in Figure 5b.

These results also highlight how the attention cost $\delta$ impacts the consumption and saving rules. Without an information cost, higher levels of precision result in greater levels of consumption in the working period ($t = 0$). This is because regardless of the signal realization, the household's expected consumption involves precautionary saving, so consumption in the first period decreases with uncertainty. When adding a cost to paying attention, however, consumption in the working period begins to decline after a certain point because the cost of paying attention eventually outweighs the increased payoff from certainty. This cost effectively reduces the income that can be consumed.

When looking at expected saving, both scenarios with and without an attention cost involve monotonically decreasing expected levels of saving as precision increases. Both involve a precautionary saving motive, but in the case with the attention cost, increased precision reduces net income, which reduces saving. In fact, the difference between the expectations with and without an attention cost grows with precision because the attention cost affects saving more and more as the capacity increases.

The calculation for $\mathbb{E}(c_1|\rho)$ uses the budget constraint, and it can be found using

$$\mathbb{E}(c_1|\rho) = \mathbb{E}\big[R(Y_0 - \delta\kappa - c_0)|\rho\big] = \mathbb{E}(R|\rho)(Y_0 - \delta\kappa) - \mathbb{E}(Rc_0|\rho),$$

where the last equality holds since $\kappa$ is measurable with respect to $\rho$, and $Y_0$ is a constant. We can rewrite the following terms using the law of iterated expectations:

$$\mathbb{E}(R|\rho) = \mathbb{E}\big[\mathbb{E}(R|\rho, s)|\rho\big] \text{ and } \mathbb{E}(Rc_0|\rho) = \mathbb{E}\big[\mathbb{E}(Rc_0|\rho, s)|\rho\big]$$

In the second expectation, since $c_0 \equiv c_0(s, \rho)$ is measurable with respect to $\rho, s$, it can be pulled out of the expectation, so we have

$$\mathbb{E}\big[\mathbb{E}(R|\rho, s)|\rho\big] = \mathbb{E}(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}|\rho) \text{ and } \mathbb{E}\big[c_0\mathbb{E}(R|\rho, s)|\rho\big] = \mathbb{E}(c_0 e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}|\rho),$$

by Proposition 1. Now, we can numerically integrate these terms over all possible signal realizations, so period-1 consumption becomes:

$$\mathbb{E}(c_1|\rho) = (Y_0 - \delta\kappa)\int_{-\infty}^{\infty} e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} f(s|\mu, \rho)\, ds - \int_{-\infty}^{\infty} c_0 e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} f(s|\mu, \rho)\, ds, \tag{19}$$

where $c_0$ is given in (16). Or, in expectation form, the expectation is

$$\mathbb{E}(c_1|\rho) = (Y_0 - \delta\kappa)\mathbb{E}(e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}|\rho) - \mathbb{E}(c_0 e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}|\rho),$$

where $f$ is the same density function defined previously.

This formulation suggests that expected consumption during the retirement period depends on the expected future value of income left over after consuming and paying the attention cost in period $t = 0$.

The following section expands on these findings by exploring what happens when the agent is allowed to choose the precision of its conditional distribution optimally.

## Step 2: A Rational Inattention Problem

In this section, the household is able to choose its signal based on the conditional precision that maximizes its lifetime utility. I break the problem into parts: first, I calculate the expected utility given a precision level, then, I optimize over all possible choices of precision numerically.

### Finding Expected Utility

We can use a similar approach as in the previous section and calculate the expected lifetime utility, given a level of signal precision. In other words, we want to calculate:

$$\mathbb{E}(U|\rho) \equiv \mathbb{E}\big[u(c_0) + \beta(c_1)|\rho\big]$$

such that $c_1$ and $c_0$ solve the households problem in (11). Plugging in the CRRA utility function, this becomes:

$$\mathbb{E}(U|\rho) = \frac{1}{1-\gamma}\mathbb{E}(c_0^{1-\gamma}|\rho) + \frac{\beta}{1-\gamma}\mathbb{E}\big[R^{1-\gamma}(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big]. \tag{20}$$

The last expectation can be simplified using the law of iterated expectations like before and plugging in the value of $Q$, defined in (15), so it becomes

$$\mathbb{E}\big[R^{1-\gamma}(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big] = \mathbb{E}\big[Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big].$$

This expectation can be calculated using $c_0$ as defined in (16), so

$$\mathbb{E}\big[Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big] = \int_{-\infty}^{\infty} Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma} f(s|\mu_0, \rho)\, ds.$$

The first term in equation (20) can be calculated numerically as a function of $c_0$, so putting these together, we have:

$$\mathbb{E}(U|\rho) = \frac{1}{1-\gamma}\int_{-\infty}^{\infty} c_0^{1-\gamma} f(s|\mu_0, \rho)\, ds + \frac{\beta}{1-\gamma}\int_{-\infty}^{\infty} Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma} f(s|\mu_0, \rho)\, ds, \tag{21}$$

or in expectation form,

$$\mathbb{E}(U|\rho) = \frac{1}{1-\gamma}\mathbb{E}(c_0^{1-\gamma}|\rho) + \frac{\beta}{1-\gamma}\mathbb{E}\big[Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big]. \tag{22}$$

Not only can this be calculated numerically, but we can also notice that $\mathbb{E}(U|\rho)$ is merely a function of model parameters and the conditional precision, $\rho$. As a result, we can use a single-variable optimization to find the optimal choice of $\rho$ to solve our RI problem.

**Optimization**

Putting these all together, the Rational Inattention optimization solves the objective function (11), subject to the consumption rule given the signal realizations given by (12) and (13), as well as the original budget and information constraints in (10) and (9). Written again below for convenience, it is:

$$\max_{\rho} \mathbb{E}\big[u(c_0^*) + \beta(c_1^*)|\rho\big]$$

$$\text{subject to:}$$

$$c_0^* = \arg\max_{c_0} \mathbb{E}\big[u(c_0) + \beta u(c_1^*)|\rho, s\big] \tag{23}$$

$$c_1^* = R - (Y_0 - \delta\kappa - c_0)$$

$$\tfrac{1}{2}\log(\sigma^2\rho) = \kappa$$

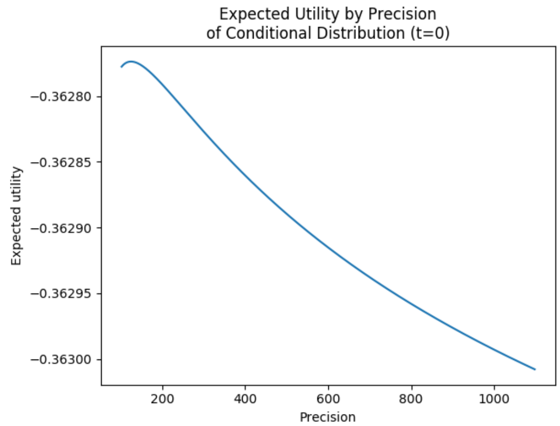
Where equation (23) is the added constraint which reflects households knowing their optimal consumption rule as a function of the signal realizations.

Since the calculation for $\mathbb{E}(U|\rho)$ has already been completed, we can calculate $\rho^*$, the optimal precision of the conditional distribution, numerically by maximizing the value of (22). In other words, the solution for the choice of conditional precision follows:

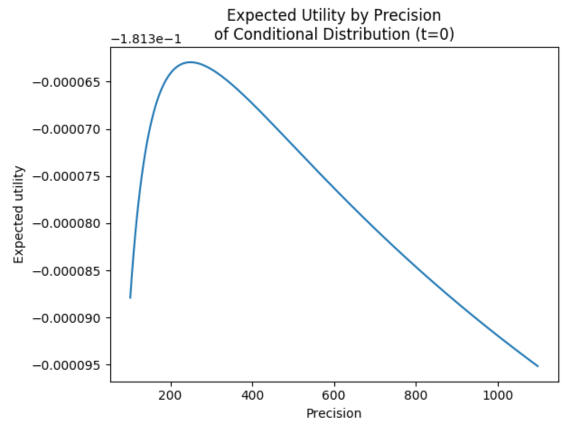
$$\rho^* = \arg\max_{\rho}\Big\{\frac{1}{1-\gamma}\mathbb{E}(c_0^{1-\gamma}|\rho) + \frac{\beta}{1-\gamma}\mathbb{E}\big[Q(Y_0 - c_0 - \delta\kappa)^{1-\gamma}|\rho\big]\Big\} \tag{24}$$

While equation (24) is not an analytical solution, we can computationally simulate how the optimal precision changes over variations in the model specifications.

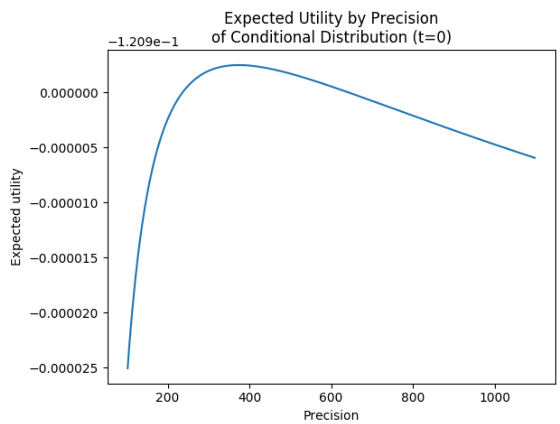
In Figure 6, we can see that as the precision of the conditional distribution increases for each income level, there is an initial increase in utility followed by a decrease. For the lower income levels, this decrease sets in at a lower precision level and is steeper. This is because for any precision level, the attention cost is a larger fraction of income for lower income levels, so the cost of attention increases more rapidly in relative terms. However, in all the graphs, there is some degree of increasing returns from higher precision up to a



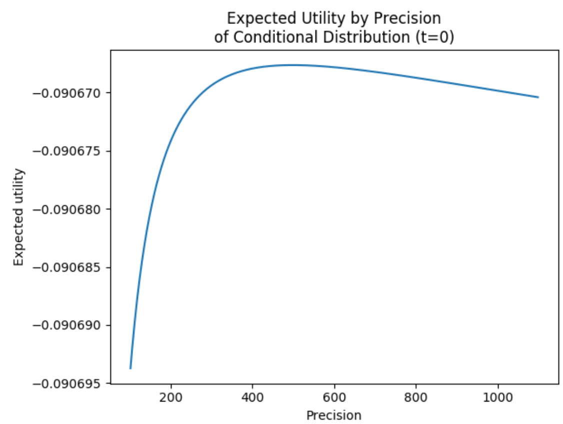

(a) Income = 10



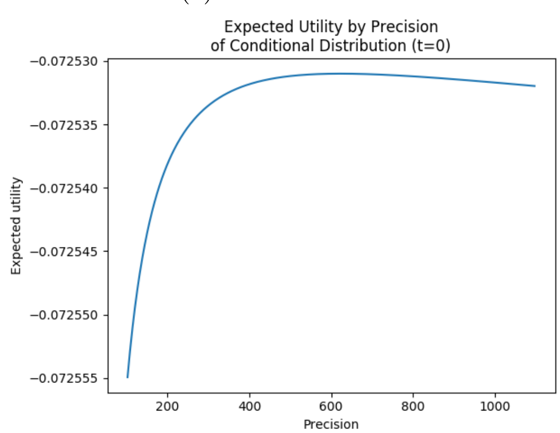

(b) Income = 20



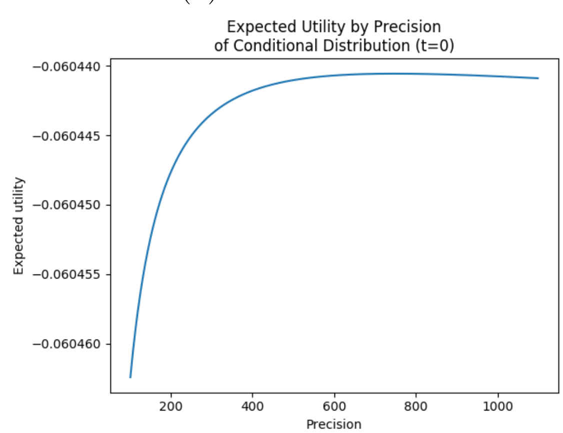

(c) Income = 30




(d) Income = 40

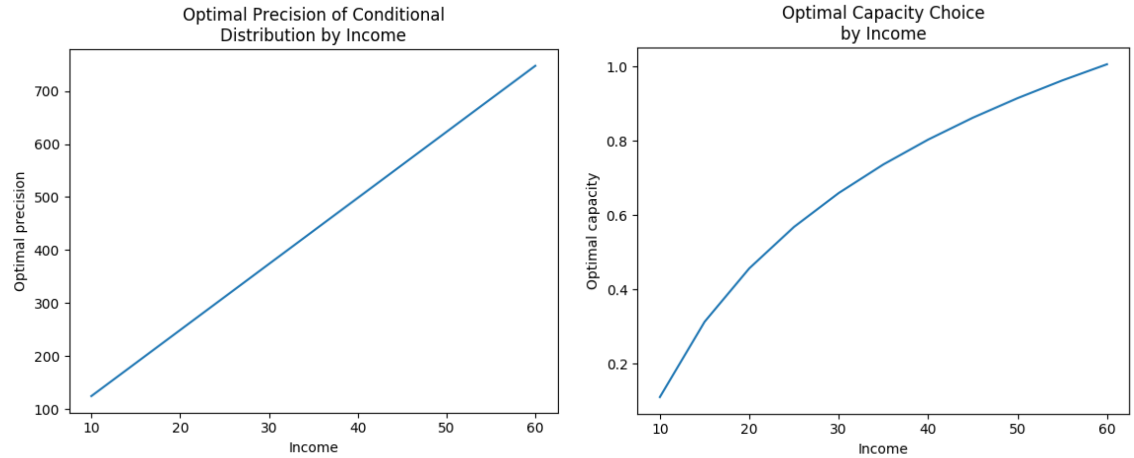



(e) Income = 50




(f) Income = 60

Figure 6: Expected Lifetime Utility over signal precision by income/wealth level when $\beta = 0.9, \gamma = 2$, $\delta = 0.01$, $\mu = 0.1$, and $\sigma = 0.1$.




(a) Optimal choice of precision of conditional distribution $\rho$ by income.

(b) Optimal Choice of Capacity $\kappa$ by Income

Figure 7: Optimal choice of precision of conditional distribution and capacity constraint, respectively. These simulations use the following parameters specifications: $\beta = 0.9$, $\gamma = 2$, $\delta = 0.01$, $\mu = 0.1$, and $\sigma = 0.1$.

certain point. This is due to a benefit from precautionary savings when certainty is increased (until the attention cost overcomes it). In fact, we can see the utility price of uncertainty through precautionary saving behavior. Households are motivated largely by consumption smoothing, so precautionary saving adversely impacts the utility from maintaining relatively similar consumption levels across periods. However, under high uncertainty, precautionary saving is better than not saving enough for a risk averse household. In effect, by increasing the precision of the signal, the household is reducing risk and smoothing consumption, allowing them to increase their lifetime utility.

Furthermore, Figure 7 shows the optimal values of conditional precision $\rho^*$ and capacity $\kappa^* \equiv \frac{1}{2}\log(\sigma^2\rho^*)$ when solving equation (24) numerically, given variations in working-period income. Income is nearly linearly related to the choice of $\rho$, and capacity is an increasing function of income. This affirms the notion that households will choose optimally to pay more attention to inflation if they have more pre-retirement income.

We can similarly solve (24) numerically for variations in noise in the underlying interest rate distribution. When the mean of the rate of return $R$ remains centered, Figure 8 shows how the implied *signal precision* $\rho_\epsilon$ increases with unconditional interest rate noise (where $\rho_\epsilon = \rho - \rho_r$). This means as the fundamental distribution of the rate of return becomes more noisy, households choose a higher capacity level (and pay a higher attention cost) in order to reduce this uncertainty.

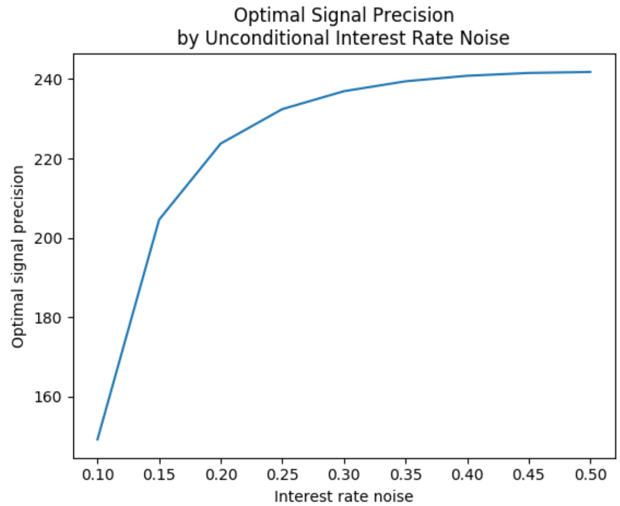



Figure 8: Optimal choice of signal precision $\rho_\epsilon$ (implied from optimal precision of conditional distribution) as unconditional variance $\sigma^2$ increases. The following parameters are used: $\mu = \mu^* - \frac{1}{2}\sigma^2$, where $\mu^* = 0.1$ is a constant, $\beta = 0.9$, $\gamma = 2$, $\delta = 0.01$, and $Y_0 = 20$.

# 5 Discussion

The RI model in Section 4 suggested that as income before retirement increases, optimal precision of signal—which can be interpreted as the attention paid to inflation—increases as well. Assuming the hump-shaped weighting pattern from the RLS life-cycle gain in Section 3 is somewhat influenced by accumulation of wealth over the life cycle, we can see that a RI approach can help to explain the results of the life-cycle RLS model through changes in wealth.

Suppose in the RI model that the signal is the only information that individuals ultimately acquire about inflation. In other words, once the consumption decision is made upon a signal realization, the consumer focuses her attention to the next period rather than checking to see how close the signal realization was to the actual rate after the fact. Then, the accumulated knowledge this individual has about inflation is entirely based on the history of her signal observations, rather than observations of actual inflation. Of course, since there is some covariance between the signal and the actual inflation rate (depending on the chosen level of precision), her observation of the signal may be very close to the actual inflation level. Additionally, since this individual knows (and chose) the exact precision of the conditional distribution, she understands the probability that her signal observation is very close to actual inflation at that time. Then it is reasonable to suggest this consumer would place more stock in those signals with higher precision levels when recalling the history of actual inflation over her lifetime.

Drawing a comparison between this phenomenon and the RLS Learning approach, we can see how some individual may choose a higher RLS gain when setting expectations about the future during periods when her chosen RI precision was high. In other words, we can conjecture some indirect mapping from the capacity $\kappa$ from the RI model to the gain $\gamma$ from the RLS model. As the RI capacity $\kappa$ increases with income during the working period, the individual has more precise estimates of current inflation, so her RLS sensitivity $\gamma$ to current expectations would rationally increases. Similarly, during the retirement period, the two-period RI model chooses $\kappa = 0$ because the savings from the working period are consumed regardless of the inflation rate. In the RLS gain, there are $T$ periods in the life cycle (rather than just two), so the gain $\gamma$ does not immediately drop to $\gamma = 0$ during retirement. However, as $t - s \to T$, or as the individual retires, the RLS implied weights begin to decrease, as seen from the results of the Life Cycle gain in Figure 3.

To recapitulate, the RI capacity choice and the RLS gain can be thought of as analogous variables that reflect how much confidence individuals have in their inflation estimates. Under this analogy, we can see that as income increases, people increase their capacity choice (Figure 7b), consequently observing more precise signals of inflation. Similarly, in the RLS gain, as people age and thus have more opportunity to accumulate wealth, they increase both the gain and implied weights on their inflation observations. If we assume that variations over age in the RLS implied weighting scheme are in part due to the associated variations in wealth over the life cycle, we can see how a RI approach can help explain the motivations behind this age-driven heterogeneity.

Moreover, from the results of Section 4, we also know that the greater the fundamental interest rate noise, the more attention is paid to inflation (Figure 8). In other words, when the interest rate noise is naturally high, individuals choose a more precise signal to reduce this uncertainty. Considering this result in the context of monetary policy, we can see that more volatility in inflation could in fact make people pay more attention. Since attention is costly under the RI model, the greater the optimal precision (holding all other inputs constant), the lower the consumer welfare. Thus, this RI model also provides intuition for the welfare loss from volatility in inflation over time.

Section 3 shows that there is age-driven heterogeneity in expectations, and Section 4 shows how this could be in part attributed to wealth- or income- driven heterogeneity. When considering monetary policy for the aim of anchoring consumer inflation expectations, this heterogeneity in attention and consequently, expectations, suggests there may not be a simple or broad-based approach.

Additional research in this area is needed to determine causality between the RI response to changes in income and the RLS gain. If longitudinal data following individuals with different income and wealth profiles over the life cycle were to become available, further empirical studies could control for the effect of age on wealth to isolate any direct effects of wealth on the RLS gain. Further theoretical research could also be helpful—by adding a third period in the RI model, for example, different income profiles could be examined to determine exactly how attention varies with different life experiences. Alternatively, further modeling could add another dimension to the RI model by examining the tradeoff between attention to human capital vs. attention to financial capital. If in fact such a tradeoff occurs and there is a capacity constraint, we may see that young people pay less attention to inflation, even if they are very wealthy, in order to invest more time in their human capital. Finally, additional policy research may be prudent in determining how age- and wealth-driven dispersion impacts the effectiveness of inflation expectations as a monetary policy tool.

# References


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

Lusardi, A., Michaud, P., Mitchell, O. S. (2017). Optimal Financial Knowledge and Wealth Inequality. *Journal of Political Economy, 125*(2), 431-477. https://doi.org/10.1086/690950

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

Malmendier, U. Nagel, S. (2016). Learning from Inflation Experiences, *The Quarterly Journal of Economics, 131*(1), 53–87, https://doi.org/10.1093/qje/qjv037

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

Mankiw, G. N. Reis, R. (2006). Pervasive Stickiness. *American Economic Review, 96*(2), 164-169.

Orphanides, A. Williams J. C. (2003). Imperfect Knowledge, Inflation Expectations, and Monetary Policy. *NBER Working Paper Series 9884.* https://doi.org/10.3386/w9884

Sargent, T. J. (1982). The Ends of Four Big Inflations. In Hall, R. E. (Ed.), *Inflation: Causes and Effects.* (pp. 41-98). University of Chicago Press.

Sargent, T. J. (1999). *The Conquest of American Inflation.* Princeton University Press.

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

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

# A Proof of Proposition 1

**Proposition 1.** If $R_0$ is given by $R_0 = e^{Z_0}$, with $Z_0$ normally distributed, and for some signal $s_0$ that covaries with $Z_0$, the conditional distribution is given by $Z_0|s_0 \sim N(\hat{\mu}, \hat{\sigma}^2)$, then:

$$\mathbb{E}(R_0^{1-\gamma}|s_0) = e^{(1-\gamma)\hat{\mu}+(1-\gamma^2)(\hat{\sigma}^2/2)}$$

*Proof.* First we want to show that $\mathbb{E}(e^z|\rho, s_0) = \mathbb{E}(e^{\hat{z}})$. By definition of an expectation,

$$\mathbb{E}(e^z|\rho, _0) = \int_{-\infty}^{\infty} e^z \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}((z-\hat{\mu}_t)/\hat{\sigma}^2)^2}\, dz$$

Expanding out the squared component of the exponent we have

$$= \int_{-\infty}^{\infty} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-(1/2\hat{\sigma}^2)(z^2-2z\hat{\mu}_t+\hat{\mu}_t^2-2z\hat{\sigma}^2)}\, dz$$

Now, multiply and divide by $e^{\hat{\mu}+\hat{\sigma}^2/2}$, which is a constant.

$$= e^{\hat{\mu}+\hat{\sigma}^2/2} \int_{-\infty}^{\infty} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-(1/2\hat{\sigma}^2)(z^2-2z\hat{\mu}_t+\hat{\mu}_t^2-2z\hat{\sigma}^2+2\hat{\mu}\hat{\sigma}^2+\hat{\sigma}^4)}\, dz$$

Simplifying this further, we get

$$= e^{\hat{\mu}+\hat{\sigma}^2/2} \int_{-\infty}^{\infty} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-(1/2\hat{\sigma}^2)(z^2-2z(\hat{\mu}_t+\hat{\sigma}^2)+(\hat{\mu}_t+\hat{\sigma}^2)^2)}\, dz$$

$$= e^{\hat{\mu}+\hat{\sigma}^2/2} \int_{-\infty}^{\infty} \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-(1/2\hat{\sigma}^2)(z-(\hat{\mu}_t+\hat{\sigma}^2))^2}\, dz$$

Now, we see that the expression in the integral is the density function for some normal random variable with distribution $N(\hat{\mu}_t + \hat{\sigma}^2, \hat{\sigma}^2)$, so it integrates to one. The expression then becomes:

$$\mathbb{E}(e^z|\rho, _0) = e^{\hat{\mu}+\hat{\sigma}^2/2}.$$

By properties of normal random variables, we know that for some linear function of $\widehat{Z}$:

$$f(\widehat{Z}) = a\widehat{Z} + b,$$

we have:

$$\mathbb{E}(e^{f(Z)}) = e^{a\hat{\mu}+a^2\hat{\sigma}^2/2+b}.$$

And so we can use this to find:

$$\mathbb{E}(e^{(1-\gamma)\hat{Z}}) = e^{(1-\gamma)\hat{\mu}+(1-\gamma)^2(\hat{\sigma}^2/2)}.$$

□

# B Details of the Signal Extraction Model

**Solving the Household's Problem for $c_0$ and $c_1$:**

We want to find the values of $c_0$, $c_1$ that solve:

$$\max_{c_0,c_1} \mathbb{E}\big[u(c_0) + \beta u(c_1)|\rho, s\big] \tag{25}$$

subject to:

$$c_1 = R(Y_0 - \delta\kappa - c_0) \tag{26}$$

$$\tfrac{1}{2}\log(\sigma^2\rho) = \kappa \tag{27}$$

Equivalently, we can combine equations (25) and (26) so that our objective function becomes:

$$\max_{c_0} \mathbb{E}\Big[u(c_0) + \beta u\big(R(Y_0 - \delta\kappa - c_0)\big)|\rho, s\Big].$$

We can find the following first order condition:

$$\frac{\partial}{\partial c_0}\mathbb{E}\big[u(c_0)|\rho, s\big] - \frac{\partial}{\partial c_0}\mathbb{E}\Big[\beta u\big(R(Y_0 - \delta\kappa - c_0)\big)|\rho, s\Big] = 0.$$

By "nice" properties of $u(c)$, we can switch the order of derivatives and expectations such that this becomes:

$$u'(c_0) = -\beta\mathbb{E}\Big[u'\big(R(Y_0 - \delta\kappa - c_0)\big)|\rho, s\Big].$$

Now, we can substitute in the quadratic utility function,

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}.$$

Using this utility function, the optimality condition becomes:

$$\begin{aligned} c_0^{-\gamma} &= \beta\mathbb{E}\Big[\big(R(Y_0 - c_0 - \delta\kappa)\big)^{-\gamma}R|\rho, s\Big] \\ &= \beta\mathbb{E}\big[R^{1-\gamma}(Y_0 - c_0 - \delta\kappa)^{-\gamma}|\rho, s\big]. \end{aligned}$$

Then, $Y_0, c_0, \kappa$ can be pulled out of the expectation and moved to the other side since they are measurable with respect to $\rho, s$.

$$\Big(\frac{c_0}{Y_0 - c_0 - \delta\kappa}\Big)^{-\gamma} = \beta\mathbb{E}(R^{1-\gamma}|\rho, s)$$

Finally, we get,

$$c_0 = \big[\beta\mathbb{E}(R^{1-\gamma}|\rho, s)\big]^{-1/\gamma}(Y_0 - c_0 - \delta\kappa).$$

Using Proposition 1, we know that

$$Q \equiv e^{(1-\gamma)\hat{\mu}+(1-\gamma^2)(\hat{\sigma}^2/2)} = \mathbb{E}(R^{1-\gamma}|\rho, s),$$

so we can simplify $c_0$ to

$$c_0 = (\beta Q)^{-1/\gamma}(Y_0 - c_0 - \delta\kappa),$$

which yields our optimality conditions:

$$c_0 = \frac{(\beta Q)^{-1/\gamma}}{1 + (\beta Q)^{-1/\gamma}}(Y_0 - \delta\kappa)$$

$$c_1 = R(Y_0 - c_0 - \delta\kappa) = R\left(1 - \frac{(\beta Q)^{-1/\gamma}}{1 + (\beta Q)^{-1/\gamma}}\right)(Y_0 - \delta\kappa)$$

The solution to the other problems are stated in full detail in the RI model section.