



Munich Personal RePEc Archive

The Role of Bounded Rationality in Macro-Finance Affine Term-Structure Models

Yun, Tack and Kim, Jinsook and Ko, Eunmi

Seoul National University, Seoul National University, Seoul National
University

1 October 2012

Online at <https://mpra.ub.uni-muenchen.de/44212/>

MPRA Paper No. 44212, posted 05 Feb 2013 11:28 UTC

The Role of Bounded Rationality in Macro-Finance Affine Term-Structure Models

Jinsook Kim, Eunmi Ko, and Tack Yun

Department of Economics
Seoul National University

This Draft: October 2012

Abstract

Our goal in this paper is two-fold. First, we develop a class of term structure models that allow for the role of bounded rationality by incorporating either information-processing constraint or fear for mis-specification into affine term structure models. We indentify a set of sufficient conditions to generate the observational equivalence between affine term-structure models with rational inattention and a fear for model misspecification. The presence of bounded rationality creates a new additional factor that is not spanned by conventional factors such as level, slope, and curvature factors. Second, our empirical results indicate that substantial amounts of information capacity constraint and robustness preference for model misspecification are needed to explain the observed behavior of yields.

JEL classification: E43; E44; G11; G12

Keywords: Rational Inattention; Robustness; Affine Term Structure Models; No-Arbitrage

E-mail address: tackyun@snu.ac.kr. We appreciate helpful and valuable comments from participants of seminars at Hong Kong University, Seoul National University, Yonsei University, and the Joint Annual Conference of Korean and Korean-American Economic Associations. Tack Yun has been supported by SNU Institute for Research in Finance and Economics.

1 Introduction

In many recent works on macroeconomics and finance, researchers have emphasized the importance of informational frictions such as rational inattention and a preference for robustness to model mis-specification in understanding the actual behavior of key macroeconomic variables and financial indicators. In addition, macro-finance affine term-structure models have drawn a lot of attention from academic researchers and policy makers. In this paper, we combine affine term-structure models with recent works on rational inattention and a preference for robustness to model mis-specification, which can be used to understand the role of rational inattention or fear for model misspecification in actual data on yields and macroeconomic variables.

Our goal in this paper is two-fold. First, we develop a class of affine term structure models that allow for the role of bounded rationality by incorporating either information-processing constraint or fear for mis-specification into an otherwise canonical representation of the three-factor arbitrage-free affine model. We indentify a set of sufficient conditions to generate the observational equivalence between affine term-structure models with rational inattention and a fear for model misspecification. For each of these models, the presence of bounded rationality creates a new additional factor that is not spanned by conventional factors such as level, slope, and curvature factors.¹ Second, we explore the possibility that macro-finance affine models with such a bounded rationality factor help estimate the magnitude of rational inattention or fear for model misspecification.

We begin with a future-state information-tracking problem associated with the stochastic discount factor. Our characterization of information processing problems builds on the work of Sims (2011). We also incorporate the risk-sensitive approach of Hansen and Sargent (2011) into the identification of the unknown stochastic discount factor in an otherwise canonical affine term-structure model. As mentioned above, each of risk-sensitive and rational inattention approaches generates a new additional factor respectively, which is called a bounded rationality factor in our paper.

In particular, although these models have different rules governing intertemporal movements of their bounded rationality factors, the two models are observationally equivalent in terms of deflated bond prices (deflated by bounded rationality factors) when the affine structure is maintained. We thus exploit this equivalence result in order to construct an empirical measure of bounded rationality factors.

¹It has been emphasized in the literature that conventional three factors such as level, slope, and curvature factors help understand the observed behavior of the yield curve, as shown in Litterman and Scheinkman (1991), Dai and Singleton (2000), Ang and Piazzesi (2003) and so on.

Specifically, we adopt a two-step approach to estimate the degree of rational inattention and size of multiplier in the risk-sensitive approach. First, we construct an empirical measure of bounded rationality factors. Second, each model’s law of motion for its bounded rationality factor is then used to estimate the magnitude of rational inattention or the degree of fear for model mis-specification. Our empirical results indicate that substantial amounts of information capacity constraint and robustness for model misspecification are needed to explain the observed behavior of yields when we use affine term-structure models. In addition, out-of-sample forecasts of our models indicate that the incorporation of bounded rationality can enhance an affine term-structure model’s capability to forecast yields.

The empirical performance of our approach can be also evaluated in terms of the role of bounded rationality factors to help explain excess returns of bond holdings. A set of recent papers have emphasized the importance of additional factors in understanding the observed behavior of excess returns. For example, Cochrane and Piazzesi (2005) find that lagged forward rates contain information about future excess bond returns that is not in current forward rates. Duffee (2011) also argues the importance of a “hidden” factor in understanding the behavior of excess returns of long-term bonds.²

A plausible source of hidden factors suggested by Duffee is the news that raises risk premia and simultaneously leads investors to believe the Fed will soon cut short-term interest rates. In this case, the increase in risk premia induces an immediate increase in long-term bond yields, while the expected drop in short rates induces an immediate decrease in these yields. Our analysis suggests one more source for hidden factors: Inability to observe some current state variables that do not have direct influence on short-term interest rates but affect the market price of risk.

In particular, we present a couple of sufficient conditions for the inattention factors to be identical to the hidden factors. First, some factors that do not directly affect the short-term interest rate are subject to information-processing constraint under equivalent probability measures. Second, the market price of risk should respond to these inattention factors. This feature is in contrast with the case of rational inattention on future states. Once these two sufficient conditions are satisfied, the presence of rational inattention helps generate the linear structure of factor dynamics that is the same as the one used in Duffee, while it leads to a set of cross-equation restrictions on the parameters associated with inter-temporal movements of hidden factors.

²Recent affine term structure models with additional factors include Duffee (2002) and Joslin, Priebsch, and Singleton (2010), while Svensson (1995), Christensen, Diebold, and Rudebusch (2009) and many other researchers also have employed modified Nelson-Siegel models with additional factors.

Our framework might provide a different channel through which the inattention factor can create the identical effect of the hidden factor on yield curves and excess returns. Let us suppose that the central bank is not subject to any rational inattention constraint in observing factors when the central bank determines the short-term nominal interest rate, whereas financial-market participants are subject to a rational inattention constraint for one of the level, slope, and curvature factors. In this case, a shock to the information set of agents does not move the current time- t yield curve but affect the equilibrium dynamics of yields and thus expected excess returns of holding bonds.

The paper proceeds as follows. Section 2 present affine term structure models when a rational inattention constraint is imposed on the stochastic discount factor. In section 3, we discuss the implication of risk-sensitive approach for affine term structure models. In section 4, we exploit the observational equivalence result on bounded rationality factors when we estimate models. Section 5 contains our results on affine term structure models with rational inattention for current states. Section 6 concludes.

2 An Affine Term Structure Model with Rational Inattention on the Stochastic Discount Factor

Our main goal of this section is to present an affine term-structure model when a rational inattention constraint is imposed on the stochastic discount factor. We also discuss the case in which there is a signal extraction problem in the identification of the stochastic discount factor. The reason why we do this is that these two models produce comparable specifications of the stochastic discount factor.

Definition 2.1 (Target stochastic discount factor) *The history of states at period t is represented by $s^t = (s_0, \dots, s_t)$ and the state at period $t + 1$ is denoted by s_{t+1} . The transition probability between periods t and $t + 1$ is represented by $\pi(s_t, s_{t+1})$. The target stochastic discount factor $m(s^t, s_{t+1})$ is defined as the one that would have been effective under the assumption of rational expectations.*

The beliefs of agents are represented by the joint distribution of their perceived stochastic discount factor and the target stochastic discount factor which is defined as the one that would have been effective when the rational-expectations hypothesis holds. The construction of their beliefs is constrained by an upper bound on the relative entropy between the joint distribution and the product distribution of the two stochastic discount factors. This upper-

bound reflects the presence of capacity-constraint in processing the information regarding the target stochastic discount factor.

Assumption 2.1 (Imperfect information about the stochastic discount factor) *In the presence of the information-processing constraint, agents do not have perfect knowledge about the realized value at period $t + 1$ of the logarithm of the target stochastic discount factor $m(s^t, s_{t+1})$ given a state at period t . In addition, agents do not observe the realized value at period t of the (log) target stochastic discount factor $m(s^{t-1}, s_t)$ given a state at period $t - 1$.*

Given the first statement of Assumption 2.1, we will assume that agents choose a joint distribution of the target and perceived stochastic discount factors by minimizing their mean squared errors given the information-processing constraint. The second statement guarantees that agents cannot learn the true distribution of the target stochastic discount factor by observing realized values of the target stochastic discount factor over time.

In light of the first statement of Assumption 2.1, a constrained optimization problem for beliefs regarding the stochastic discount factor can be formulated as follows. In this optimization problem, m is the (log) target stochastic discount factor and \hat{m} is the actual (log) stochastic discount factor that are used by participants of financial markets. Following Sims (2011), agents choose a joint distribution to minimize the mean squared errors subject to an information-processing constraint.

$$\max_{f_t(m, \hat{m})} - \int \int f_t(m, \hat{m})(m - \hat{m})^2 dm d\hat{m} \quad (2.1)$$

subject to

$$\int \int f_t(m, \hat{m}) \log \frac{f_t(m, \hat{m})}{g_t(m)h_t(\hat{m})} dm d\hat{m} \leq \kappa \quad (2.2)$$

where $g_t(m)$ and $h_t(\hat{m})$ are defined as

$$g_t(m) = \int f_t(m, \hat{m}) d\hat{m}, \quad h_t(\hat{m}) = \int f_t(m, \hat{m}) dm. \quad (2.3)$$

We will present an application of Sims's problem specified above.³ In particular, we assume that both m and \hat{m} are normal random variables. In this case, when agents observe a noisy signal (for m) $z = m + \nu$ and the noise ν are a normal variable with means zero and its standard deviation σ_ν , the optimal forecast of m , defined as $\hat{m} = E[m|z]$, is given

³The solution of this non-linear optimization problem implies that the joint density function of m and \hat{m} is affected by the Lagrange multiplier associated with information-processing constraint (denoted by θ): $f_t(m, \hat{m}) = g_t(m)h_t(\hat{m}) \exp(-\frac{1}{\theta}(m - \hat{m})^2) / \int h_t(\hat{m}) \exp(-\frac{1}{\theta}(m - \hat{m})^2) d\hat{m}$.

by $\hat{m} = \tau z$ where $\tau = \sigma_m^2 / (\sigma_m^2 + \sigma_\nu^2)$ and σ_m^2 is the variance of m . The mutual information between m and \hat{m} is given by $I(m, \hat{m}) = - (1/2) \log(1 - \rho_{m, \hat{m}}^2)$ where $\rho_{m, \hat{m}}^2$ is the correlation coefficient between m and \hat{m} .⁴ Hence, the information-processing constraint can be written as

$$-\frac{1}{2} \log(1 - \tau) \leq \kappa \quad (2.4)$$

where κ is the upper bound on the capacity of processing information. Given that $\hat{m} = \tau(m + \nu)$ and $E[m\nu] = 0$, the mean squared error of m is $E[(m - \hat{m})^2] = (1 - \tau)^2 \sigma_m^2 + \tau^2 \sigma_\nu^2$. The rational inattention problem turns out to be

$$\min_{\tau} \{(1 - \tau)^2 \sigma_m^2 + \tau^2 \sigma_\nu^2 + \lambda(4\kappa + 2 \log(1 - \tau))\} \quad (2.5)$$

Definition 2.2 (Characterization of rational inattention problem) *Under the assumption that both m and \hat{m} are normal random variables, the rational inattention problem can be defined as*

$$\min_{\tau} \{(1 - \tau)^2 \sigma_m^2 + \tau^2 \sigma_\nu^2 + \lambda(4\kappa + 2 \log(1 - \tau))\} \quad (2.6)$$

where λ is the Lagrange multiplier for the information processing constraint.

The first-order condition to this minimization problem is

$$\lambda = (1 - \tau)(\tau \sigma_\nu^2 - (1 - \tau) \sigma_m^2). \quad (2.7)$$

As a result, when the information constraint is binding, the Lagrange multiplier λ is determined by the upper bound on the information-processing constraint:

$$\lambda = \exp(-2\kappa) ((1 - \exp(-2\kappa)) \sigma_\nu^2 - \exp(-2\kappa) \sigma_m^2). \quad (2.8)$$

Assumption 2.2 (Uniqueness of the target stochastic discount factor) *When the transition distribution of the state vector between periods t and $t + 1$ is denoted by $\pi(s_t, s_{t+1})$, this transition distribution is not affected by the presence of rational inattention. In addition, the logarithm of the target stochastic discount factor $m(s^t, s_{t+1})$ exists uniquely for each pair of states (s^t, s_{t+1}) .*

⁴The mutual information measures the amount of information updated in each period. The entropy of a normal $n \times 1$ random vector (denoted by f) is $H(f) = \frac{1}{2} \log((2\pi e)^n |K|)$ where K is the covariance matrix. The mutual information measures the reduction of uncertainty of X due to the knowledge of Y : $I(X, Y) = h(X) - h(X|Y)$ where $h(X)$ is computed at period $t - 1$ information set and $h(X|Y)$ is computed at period t information set. $h(X|Y) = \frac{1}{2} \log((2\pi e)^n |\Sigma_t|)$ and $h(X) = \frac{1}{2} \log((2\pi e)^n |M_t|)$. Hence, the mutual information is $I(X, Y) = \frac{1}{2} (\log(|\Sigma_t|) - \log(|M_t|))$.

The uniqueness of the target stochastic discount factor leads to a one-to-one correspondence between a future state s_{t+1} and the actual stochastic discount factor that holds for agents who are subject to the rational inattention constraint for future states. Hence, this one-to-one correspondence allows us to compute expected present-values of future payoffs by using the actual stochastic discount factor, given the assumption that agents know the true distribution of states in each period. Given these assumptions, no-arbitrage condition in the presence of rational inattention implies that the price at period t of a n -period bond ($= \hat{P}_{n,t}$) is

$$\hat{P}_{n,t}(s_t) = \int E[\exp(\hat{m}_{t,t+1})|m_{t,t+1}(s_{t+1})]\hat{P}_{n-1,t+1}(s_{t+1})d\pi(s_t, s_{t+1}) \quad (2.9)$$

where $\hat{m}_{t,t+1} = \hat{m}(s^t, s_{t+1})$ and $m_{t,t+1}(s_{t+1}) = m(s^t, s_{t+1})$. In the absence of rational inattention, the price at period t of a n -period bond (denoted by $P_{n,t}$) is

$$P_{n,t}(s_t) = \int M_{t,t+1}(s_{t+1})P_{n-1,t+1}(s_{t+1})d\pi(s_t, s_{t+1}) \quad (2.10)$$

where $M_{t,t+1}(s_{t+1}) = \exp(m_{t,t+1}(s_{t+1}))$.

Proposition 2.1 (Characterization of the stochastic discount factor under a signal extraction problem) *Suppose that assumptions 2.1 and 2.2 hold. The stochastic discount factor is then given by $E[\exp(\hat{m}_{t,t+1})|m_{t,t+1}] = \exp(\tau_t m_{t,t+1} + \tau_t^2 \sigma_v^2 / 2)$, where $\tau_t = \sigma_t(m_{t,t+1})^2 / (\sigma_t(m_{t,t+1})^2 + \sigma_\nu^2)$. The size of σ_ν^2 is constant and exogenously determined.*

In order to show that this proposition holds, we will use a version of signal extraction problem that permits the derivation of the functional form of the actual stochastic discount factor. Let us suppose that agents expect to observe a random variable z , while z can be interpreted as a noisy signal for the target stochastic discount factor: $z = m + \nu$. In this case, \hat{m} is the optimal forecast of m given the noisy signal, so that $\hat{m} = E[m|z]$. The solution to this signal extraction problem leads to $E[m|z] = \tau z$ and $\tau = \sigma_m^2 / (\sigma_m^2 + \sigma_\nu^2)$. The main difference from rational inattention is that the size of σ_ν^2 is exogenously determined in the model with signal extraction. Given that \hat{m} follows a normal distribution, we have $E[\exp(\hat{m})|m] = \exp(\tau m + \tau^2 \sigma_\nu^2 / 2)$ where $\tau = \sigma_m^2 / (\sigma_m^2 + \sigma_\nu^2)$. In sum, the actual stochastic discount factor between periods t and $t + 1$ can be written as

$$E[\exp(\hat{m}_{t,t+1})|m_{t,t+1}(s_{t+1})] = \exp(\tau_t m_{t,t+1}(s_{t+1}) + \frac{\tau_t^2 \sigma_\nu^2}{2}). \quad (2.11)$$

The logarithm of the actual stochastic discount factor is thus given by

$$\tilde{m}_{t,t+1} = \tau_t m_{t,t+1}(s_{t+1}) + \frac{\tau_t^2 \sigma_\nu^2}{2}. \quad (2.12)$$

The logarithm of the actual stochastic discount factor is conditionally affine in terms of the logarithm of the stochastic discount factor that would have been effective when rational-expectations hypothesis holds. The addition of a term such as $\tau_t^2 \sigma_v^2 / 2$ reflects the fact that the stochastic discount factor is a log-normal random variable.

Proposition 2.2 (Characterization of the stochastic discount factor under rational inattention) *Suppose that agents solve the rational inattention problem specified in definition 2.2. The resulting stochastic discount factor is given by $E[\exp(\hat{m}_{t,t+1}) | m_{t,t+1}] = \exp(\tau m_{t,t+1} + \tau^2 \sigma_v^2 / 2)$. Moreover, we should solve $\tau = \sigma_t(m_{t,t+1})^2 / (\sigma_t(m_{t,t+1})^2 + \sigma_{v,t}^2)$ and $\tau = 1 - \exp(-2\kappa)$ at the same time. As a result, $\sigma_{v,t}^2$ varies over time and is endogenously determined.*

The solution to our rational inattention problem (defined in definition 2.2) implies that $\tau = 1 - \exp(-2\kappa)$. It is important to note that τ is constant over time in the presence of rational inattention. Therefore, we have $\sigma_{v,t}^2 = \sigma_t(m_{t,t+1})^2 \exp(-2\kappa) / (1 - \exp(-2\kappa))$. As a result, the difference between rational inattention and signal extraction is reflected in whether the variance of noises is constant or time-varying.

Proposition 2.3 (Market price of risk) *Suppose that assumptions 2.1 and 2.2 hold. The market price of risk under rational inattention is*

$$\frac{\sigma_t(\hat{M}_{t,t+1})}{E_t[\hat{M}_{t,t+1}]} = \sqrt{\exp(\tau_t^2 \sigma_t(M_{t,t+1})^2) - 1} \quad (2.13)$$

where $\hat{M}_{t,t+1}$ is the stochastic discount factor under rational inattention and $M_{t,t+1}$ is the target stochastic discount factor. In particular, we have $\tau_t = 1 - \exp(-2\kappa)$ in the case of rational inattention and $\tau_t = \sigma_t(m_{t,t+1})^2 / (\sigma_t(m_{t,t+1})^2 + \sigma_v^2)$ in the case of signal extraction.

We move onto the impact of rational inattention on the market price of risk. The market price of risk is defined as the ratio of the standard deviation of the stochastic discount factor (conditional on current-period's information) to its conditional expectation. The market price of risk is $\sigma_t(M_{t,t+1}) / E_t[M_{t,t+1}]$ in the absence of any informational friction. In order to see the impact of rational inattention on the market price of risk, we compare the market price of risk under rational inattention with the one under rational expectations:

$$\frac{\sigma_t(\hat{M}_{t,t+1})}{E_t[\hat{M}_{t,t+1}]} = \frac{\sigma_t(M_{t,t+1}^{\tau_t})}{E_t[M_{t,t+1}^{\tau_t}]} \quad (2.14)$$

where $\hat{M}_{t,t+1} = E[\exp(\hat{m}_{t,t+1})|m_{t,t+1}]$. Under the assumption of conditional log-normality, the market price of risk under rational expectations is

$$\frac{\sigma_t(M_{t,t+1})}{E_t[M_{t,t+1}]} = \sqrt{\exp(\sigma_t(M_{t,t+1})^2) - 1} \quad (2.15)$$

while the market price of risk under rational inattention is

$$\frac{\sigma_t(\hat{M}_{t,t+1})}{E_t[\hat{M}_{t,t+1}]} = \sqrt{\exp(\tau_t^2 \sigma_t(M_{t,t+1})^2) - 1}. \quad (2.16)$$

An implication of this equation is that the market price of risk is decreased with the introduction of rational inattention when $\tau_t (> 0)$ is less than one. The key reason why we have this result is that even in the presence of rational inattention, the stochastic discount factor depends only on true states, which means that expectation errors of agents due to rational inattention are not reflected in the pricing kernel of assets. Hence, we implicitly assume that the pricing kernel does not depend on man-made artificial uncertainties even when agents are subject to rational inattention. In this case, the logarithm of the stochastic discount factor under rational inattention is proportional to that under rational expectations while the proportionality constant is less than one reflecting the presence of information processing constraint. As a result, the market price of risk is decreased with the introduction of rational inattention as shown above.

We will see whether or not the introduction of rational inattention increases the number of factors that affect yields and prices of pure discount bonds. In order to see this, we will assume that under rational expectations, a standard affine term-structure model holds.

Assumption 2.3 (Specification of the target stochastic discount factor) *Suppose that under rational expectations, a standard affine term-structure model holds. In this model, the target stochastic discount factor is specified as*

$$M_{t,t+1} = \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1}), \quad (2.17)$$

while the dynamics of factors (denoted by X_t) are described by the following equation:

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t. \quad (2.18)$$

An implication of assumption 2.3 is that the market price of risk is not affected by the presence of rational inattention. In fact, we need this assumption regarding the price of risk, in order to maintain the affine structure of the model in the presence of rational inattention.

Proposition 2.4 (Characterization of belief factor) *Suppose that assumptions 2.1 - 2.3 hold. The bounded rationality factor in affine term-structure models with rational inattention for the stochastic discount factor evolves over time according to the following equation:*

$$\psi_{t+1} = \psi_t - (1 - \tau_t)m_{t,t+1} + \tau_t^2\sigma_v^2/2. \quad (2.19)$$

where $\tau_t = 1 - \exp(-2\kappa)$ in the case of rational inattention and $\tau_t = \sigma_t(m_{t,t+1})^2/(\sigma_t(m_{t,t+1})^2 + \sigma_v^2)$ in the case of signal extraction.

In order to show that this proposition holds, we impose no-arbitrage condition for prices of bonds that are traded by agents who are subject to rational inattention or a signal extraction problem. Specifically, no-arbitrage conditions for bond prices can be rewritten as follows:

$$\tilde{P}_{n,t} = E_t[M_{t,t+1}\tilde{P}_{n-1,t+1}] \quad \tilde{P}_{n,t} = \Psi_t\hat{P}_{n,t} \quad (2.20)$$

where $\tilde{P}_{n,t}$ denotes a deflated bond price of $\hat{P}_{n,t}$ under rational inattention (deflated by Ψ_t) and the evolution of Ψ_t can be described by the following equation:

$$\Psi_{t+1} = \Psi_t \exp(-(1 - \tau_t)m_{t,t+1} + \tau_t^2\sigma_v^2/2). \quad (2.21)$$

We also emphasize that this Euler equation for bond prices is the same as the one that would have been effective under rational expectations. We thus see that the affine structure of logarithms of bond prices and yields holds for the set of bond prices in each period denoted by $\{\tilde{P}_{n,t}\}_{n=0}^T$. Given these two assumptions, logarithms of bond prices under rational inattention and signal extraction are linear functions of state variables:

$$\hat{p}_{n,t} = \bar{a}_n + \bar{b}'_n X_t - \psi_t, \quad (2.22)$$

where \bar{a}_n and \bar{b}_n are determined by the following difference equations:

$$\bar{b}'_{n+1} = \bar{b}'_n(\phi - \Sigma\lambda_1) + \bar{b}'_1, \quad \bar{a}_{n+1} = \bar{a}_n + \bar{b}'_n(c - \Sigma\lambda_0) + \frac{1}{2}\bar{b}'_n\Sigma\Sigma'\bar{b}_n + \bar{a}_1 \quad (2.23)$$

for $n = 1, 2, \dots, T$.

In particular, the bounded rationality factor is conditionally linear in terms of shocks to state variables:

$$\psi_{t+1} = \psi_t - (1 - \tau_t)m_{t,t+1} + \tau_t^2\sigma_v^2/2. \quad (2.24)$$

It follows from this equation that $\psi_{t+1} = \psi_t$ when $\tau_t = 1$ and $\sigma_v^2 = 0$ that means the absence of any constraint on information-processing capacity. In this case, bond prices become identical to those under the assumption of rational expectations.

Proposition 2.5 (Impact of the inattention belief factor on excess returns) *The market price of risk in period t is given by*

$$\lambda_t = \lambda_0 + \lambda_1 X_t. \quad (2.25)$$

Suppose that assumptions 2.1 - 2.3 hold. The logarithm of the excess return from holding bonds with maturity n between periods t and $t + 1$ is defined as $rx_{t,t+1}^{(n)} = \hat{p}_{n-1,t+1} - \hat{p}_{n,t} + \hat{p}_{1,t}$. The logarithm of excess return can be written as

$$rx_{t,t+1}^{(n)} = \bar{b}'_{n-1} \Sigma (\lambda_0 + \lambda_1 X_t) - \frac{1}{2} \bar{b}'_n \Sigma \Sigma' \bar{b}_n + \bar{b}'_{n-1} \Sigma \epsilon_{t+1} - \psi_{t+1} \quad (2.26)$$

where $c = 0$.

The key message of this proposition is that the presence of rational inattention for future states forces excess returns of holding bonds to be influenced by the inattention belief factor on top of conventional factors.

3 An Affine Term Structure Model with Risk-Sensitive Approach

We will discuss the difference between risk-sensitive and rational inattention approaches in terms of affine term-structure models. In this section, we still assume that, although agents do not know the true stochastic discount factor, they have a set of probability evaluations about their approximating models for the stochastic discount factor. In this regard, we take the risk-sensitive approach developed in Hansen and Sargent (2011), while the following assumption is still needed for the existence of conditional expectations of agents.

Assumption 3.1 (Uniqueness of the target stochastic discount factor) *When the transition distribution of the state vector between periods t and $t + 1$ is denoted by $\pi(s_t, s_{t+1})$, this transition distribution is not affected by the presence of rational inattention. In addition, the target stochastic discount factor $M(s^t, s_{t+1})$ exists uniquely for each pair of states (s^t, s_{t+1}) .*

Proposition 3.1 (Characterization of the stochastic discount factor) *Suppose that assumption 3.1 holds. We also assume that agents have fear for the possibility of misspecifying their approximating model as in the framework of Hansen and Sargent (2011). Specifically, agents choose a probability distribution of their approximating model (for the stochastic discount factor) by solving the following problem:*

$$\min_{K_{t,t+1}(s_{t+1}) \geq 0} \int K_{t,t+1}(s_{t+1}) [M_{t,t+1}(s_{t+1}) + \theta \log K_{t,t+1}(s_{t+1})] \pi(s_t, s_{t+1}) ds_{t+1} \quad (3.1)$$

subject to

$$\int K_{t,t+1}(s_{t+1})\pi(s_t, s_{t+1})ds_{t+1} = 1. \quad (3.2)$$

The solution to this problem for $K_{t,t+1}(s_{t+1})$ is given by

$$K_{t,t+1}(s_{t+1}) = \frac{\exp(-M_{t,t+1}(s_{t+1})/\theta)}{\int \exp(-M_{t,t+1}(s_{t+1})/\theta)\pi(s_t, s_{t+1})ds_{t+1}}. \quad (3.3)$$

Our proposition is an direct application of the work of Hansen-Sargent (2011) for the risk-sensitive operator that is applied to a (value) function $V(x)$ of a random vector x with density $f(x)$. In their model, the risk-sensitivity operator is defined in terms of the indirect utility function $T[V]$ that emerges from:

$$T[V] = \min_{K(x) \geq 0} \int K(x)[V(x) + \theta \log K(x)]f(x)dx \quad (3.4)$$

subject to

$$\int K(x)f(x)dx = 1. \quad (3.5)$$

The solution to this problem for $K(x)$ is given by

$$K(x) = \frac{\exp(-V(x)/\theta)}{\int \exp(-V(x)/\theta)f(x)dx}. \quad (3.6)$$

At this point, it is worthwhile to discuss the interpretation of θ . In order to do this, we assume that the data are actually generated by a nearby model of an approximating model denoted by $\hat{\pi}(s_t, s_{t+1})$, while $\pi(s_t, s_{t+1})$ is an approximating model for the conditional distribution of future states. In addition, there is an upper bound on the conditional relative entropy between these two distributions:

$$I_t(\pi, \hat{\pi}) = \int \hat{\pi}(s_t, s_{t+1}) \log \frac{\hat{\pi}(s_t, s_{t+1})}{\pi(s_t, s_{t+1})} ds_{t+1} \geq \eta$$

A Lagrangian of the minimization problem for the belief distortion is then given by

$$\min_{K_{t,t+1}(s_{t+1}) \geq 0} \int K_{t,t+1}(s_{t+1})M_{t,t+1}(s_{t+1})\pi(s_t, s_{t+1})ds_{t+1} + \theta(I_t(\pi, \hat{\pi}) - \eta)$$

where θ is the Lagrange multiplier and $K_{t,t+1}(s_{t+1})$ is the conditional likelihood ratio:

$$K_{t,t+1}(s_{t+1}) = \frac{\hat{\pi}(s_t, s_{t+1})}{\pi(s_t, s_{t+1})}.$$

Hence, the parameter θ can be interpreted as the Lagrange multiplier for the belief distortion of agents.

The main implication of this proposition is that $K_{t,t+1}(s_{t+1})M_{t,t+1}(s_{t+1})$ is the actual stochastic discount factor when agents have fear for model misspecification and their target approximating model is $M_{t,t+1}(s_{t+1})$, given that conditional expectations of agents are computed by using the transition probability distribution between states at periods t and $t + 1$ denoted by $\pi(s_t, s_{t+1})$. In the absence of arbitrage, therefore, the nominal price at period t of pure nominal discounted (riskless) bonds whose maturity is period $t + n$ (denoted by $P_{n,t}^h$) satisfies the following relation:

$$P_{n,t}^h = E_t[K_{t,t+1}M_{t,t+1}P_{n-1,t+1}^h] \quad (3.7)$$

for $n = 1, 2, \dots, T$ and where $P_{0,t}^h = 1$ for all t .

In the same way as is done in the previous section, we define a new variable Ψ_t^h satisfying the following relation:

$$\Psi_{t+1}^h = K_{t,t+1}\Psi_t^h. \quad (3.8)$$

By using this newly defined variable, we rewrite the no-arbitrage relation specified above as follows:

$$\tilde{P}_{n,t}^h = E_t[M_{t,t+1}\tilde{P}_{n-1,t+1}^h] \quad (3.9)$$

where $\tilde{P}_{n,t}^h$ denotes a deflated bond price of $\hat{P}_{n,t}^h$ under the risk-sensitive approach (deflated by Φ_t).

Assumption 3.2 (Specification of the target approximating stochastic discount factor) *Suppose that a standard affine term-structure model holds in the absence of agents' fear for model misspecification. In addition, the target stochastic discount factor is specified as*

$$M_{t,t+1} = \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\epsilon_{t+1}), \quad (3.10)$$

while the dynamics of factors (denoted by X_t) are described by the following equation:

$$X_t = c + \phi X_{t-1} + \Sigma\epsilon_t. \quad (3.11)$$

An implication of assumption 3.2 is that the market price of risk is not affected by the presence of fear of model mis-specification. In fact, we need this assumption in order to maintain the affine structure of the model with risk-sensitive approach. Given assumption 3.2, the logarithm of the bond price under the risk-sensitive approach is a linear function of state variables:

$$p_{n,t}^h = \bar{a}_n + \bar{b}_n'X_t - \psi_t^h \quad (3.12)$$

where ψ_t^h is the logarithm of Ψ_t^h and $p_{n,t}^h$ is the logarithm of $P_{n,t}^h$.

Proposition 3.2 (Characterization of the risk-sensitive factor) *Suppose that assumptions 3.1 and 3.2 hold. In affine term-structure models with the risk sensitive approach described above, the risk-sensitive factor evolves over time according to the following equation:*

$$\psi_{t+1}^h = \psi_t^h + (\lambda'_t \epsilon_{t+1})/\theta - \frac{1}{2}(\lambda'_t \lambda_t)/\theta^2. \quad (3.13)$$

In order to show that this proposition holds, we point out that, given the result of proposition 3.1 together with assumption 3.2, the belief factor $K_{t,t+1}(s_{t+1})$ can be written as

$$K_{t,t+1}(s_{t+1}) = \frac{\exp(\lambda'_t \epsilon_{t+1}/\theta)}{\int \exp(\lambda'_t \epsilon_{t+1}/\theta) \pi(s_t, s_{t+1}) ds_{t+1}}. \quad (3.14)$$

Hence, taking logarithm to both sides of this equation leads to

$$\log K_{t,t+1}(s_{t+1}) = \lambda'_t \epsilon_{t+1}/\theta - \frac{1}{2}(\lambda'_t \lambda_t)/\theta^2 \quad (3.15)$$

where the second term in the right hand side of this equation reflects the fact that ϵ_{t+1} is a normal random variable. Substituting this equation into the law of motion for ϕ_t specified above leads to the conclusion of proposition 3.2.

Proposition 3.3 (Impact of the risk-sensitive factor on excess returns) *The market price of risk in period t is given by*

$$\lambda_t = \lambda_0 + \lambda_1 X_t. \quad (3.16)$$

Suppose that assumptions 3.1 and 3.2 hold. The logarithm of excess return can be written as

$$rx_{t,t+1}^{(n)} = \bar{b}'_{n-1} \Sigma (\lambda_0 + \lambda_1 X_t) - \frac{1}{2} \bar{b}'_n \Sigma \Sigma' \bar{b}_n + \bar{b}'_{n-1} \Sigma \epsilon_{t+1} - \psi_{t+1}^h \quad (3.17)$$

where $c = 0$.

The key message of this proposition is that excess returns of holding bonds are directly affected by the risk-sensitive factor.

4 Observational Equivalence and Empirical Measures of Bounded Rationality Factors: Stochastic Discount Factor

The key theoretic result of this section is that the bounded rationality factor from the affine model with rational inattention is observationally equivalent to that of the model with risk-sensitive approach. In this section, we will incorporate this equivalence result into our empirical works in order to estimate the size of rational inattention and the magnitude of agents' fear for model misspecification by using actual U.S. data on yields.

Before going further, we summarize our equivalence result between models with rational inattention and risk-sensitive approaches. Specifically, to the extent which the approximating model in the risk sensitive approach is identical to the target model under rational inattention, the risk-sensitive and rational inattention factors are the same: $\psi_t = \psi_t^h$.

Proposition 4.1 (Equivalence Result) *Suppose that the approximating model in the risk sensitive approach is identical to the target model under rational inattention. Given this assumption, the risk-sensitive and rational inattention factors are the same:*

$$\psi_t = \psi_t^h. \quad (4.1)$$

As a result, up to the deflated bond prices, the two models have the same set of equilibrium conditions for bond prices in each maturity. The key reason why we have this result is that, when the approximating model in the risk sensitive approach is identical to the target model under rational inattention, deflated bond prices in the two models (equation (3.9) of page 12 and equation (2.20) of page 9, respectively) are subject to the same stochastic discount factor.

It would be now worthwhile to discuss the difference between rational inattention and risk-sensitive approaches. In this regard, we note that the risk-sensitive approach delivers the following equation:

$$\psi_{t+1}^h = \psi_t^h + (\lambda_t' \epsilon_{t+1})/\theta - \frac{1}{2}(\lambda_t' \lambda_t)/\theta^2. \quad (4.2)$$

As noted in the previous section, the presence of rational inattention implies that its inattention factor is conditionally linear in terms of shocks to state variables:

$$\psi_{t+1} = \psi_t - (1 - \tau_t)\lambda_t' \epsilon_{t+1} - (1 - \tau_t)E_t[m_{t,t+1}] + \tau_t^2 \sigma_v^2/2 \quad (4.3)$$

where $\tau_t = 1 - \exp(-2\kappa)$ in the case of rational inattention and $\tau_t = \sigma_t(m_{t,t+1})^2/(\sigma_t(m_{t,t+1})^2 + \sigma_\nu^2)$ in the case of signal extraction. Hence, while each of them contains a predictable component, the size of the predictable component in inattention factor (denoted by ψ_t) depends on the degree of rational inattention, while that of the risk-sensitive approach (denoted by ψ_t^h) relies on the shadow value of the relative entropy constraint.

In particular, we adopt a two-step approach to estimate the degree of rational inattention and size of multiplier in the risk-sensitive approach. First, we construct an empirical measure of bounded rationality factors. Second, each model's law of motion for its bounded rationality factor is then used to estimate the magnitude of rational inattention or the degree of fear for model mis-specification. Our empirical results indicate that substantial amounts of

information capacity constraint and robustness for model misspecification are needed to explain the observed behavior of yields when we use affine term-structure models. In addition, out-of-sample forecasts of our models indicate that the incorporation of bounded rationality can enhance an affine term-structure model’s capability to forecast yields.

In our estimation, we assume that a subset of yields are observed without measurement errors as is done in Chen and Scott (1993) and Ang and Piazzesi (2003). Specifically, maturities of yields without measurement errors are one-month, six-month, one-year, and five-year, while maturities of yields with measurement errors are three-month and three-year. In particular, this assumption facilitates our analysis to estimate models with rational inattention constraints for the stochastic discount factor in the presence of time-varying intercept terms that are correlated with lagged innovations to state variables.

Our data set for yields is the same as the one used in Piazzesi (2010), while it covers a shorter period than the one used in Ang and Piazzesi (2003). Specifically, the sample covers data on zero coupon bond yields of maturities 1, 3, 12, 36, and 60 from January 1964 to December 2003. The bond yields (12, 36, 60 months) are from the Fama CRSP zero coupon files, while the shorter maturity rates (1 and 3 months) are from the Fama CRSP Treasury Bill files. All bond yields are continuously compounded. The observed macro variables in our data set consist of the PCE inflation rate (PCEPI) and the civilian unemployment rate (UNRATE) that can be downloaded from the FRED at the Federal Reserve Bank of St. Louis. We rely on observed macro variables without using the principal component analysis. In addition, since agents are assumed to be not subject to the information-processing constraint for observing macro variables, the information-processing constraint in our estimation is applied to only yield data, not to macro data.⁵

Having described our data set, we will discuss our estimation procedure employed in this section. We begin with the case where a group of yields can be observed without measurement errors. Specifically, when we observe N yields, K_1 is the number of yields that are observed without measurement errors and K_2 is the number of yields that are observed with measurement errors where $K_1 + K_2 = N$. As well known in the literature, this assumption is useful in making one-to-one correspondence between latent factors and some observed yields. For example, when Y_{1t} is the vector of yields that can be observed without measurement errors, the affine term-structure model of rational inattention constraint for future states implies that the set of yields without measurement errors can be written as $Y_{1t} = A_1 + B_1 X_t + C_1 \psi_t$ where Y_{1t} contains yields whose maturities are six-month, one-year, and

⁵Ang and Piazzesi relied on the principal component analysis to obtain two macro latent factors: One is associated with the aggregate price indices and the other covers some measures of aggregate real activities.

five-year. The short-rate equation is $r_t = \delta_0 + \delta_1 X_t + \psi_t$. Combining these two equations, the vector of yields without measurement errors (net of the short rate) can be written as

$$\tilde{Y}_{1t} = \tilde{A}_1 + \tilde{B}_1 X_t \quad (4.4)$$

where $\tilde{Y}_{1t} = Y_{1t} - C_1 r_t$, $\tilde{A}_1 = A_1 - C_1 \delta_0$, and $\tilde{B}_1 = B_1 - C_1 \delta_1$.

We now exploit the assumption that matrix \tilde{B}_1 is invertible. The state vector X_t can be expressed in terms of yields without measurement errors: $X_t = \tilde{B}_1^{-1}(\tilde{Y}_{1t} - \tilde{A}_1)$. Substituting this equation into the law of motion for the state vector, we can derive a VAR(1) representation for Y_{1t} as follows:

$$\tilde{Y}_{1t+1} = \hat{A}_1 + \hat{\phi} \tilde{Y}_{1t} + \tilde{B}_1 \Sigma \epsilon_{t+1} \quad (4.5)$$

where \hat{A}_1 and $\hat{\phi}$ are defined as

$$\hat{A}_1 = (I - \hat{\phi})\tilde{A}_1 + \tilde{B}_1 c \quad \hat{\phi} = \tilde{B}_1 \phi \tilde{B}_1^{-1}. \quad (4.6)$$

For the set of yields with measurement errors, three-month and three-year maturities bond yield, we have the following representation:

$$\tilde{Y}_{2t} = \hat{A}_2 + \tilde{B}_2 \tilde{B}_1^{-1} \tilde{Y}_{1t} + \eta_t \quad (4.7)$$

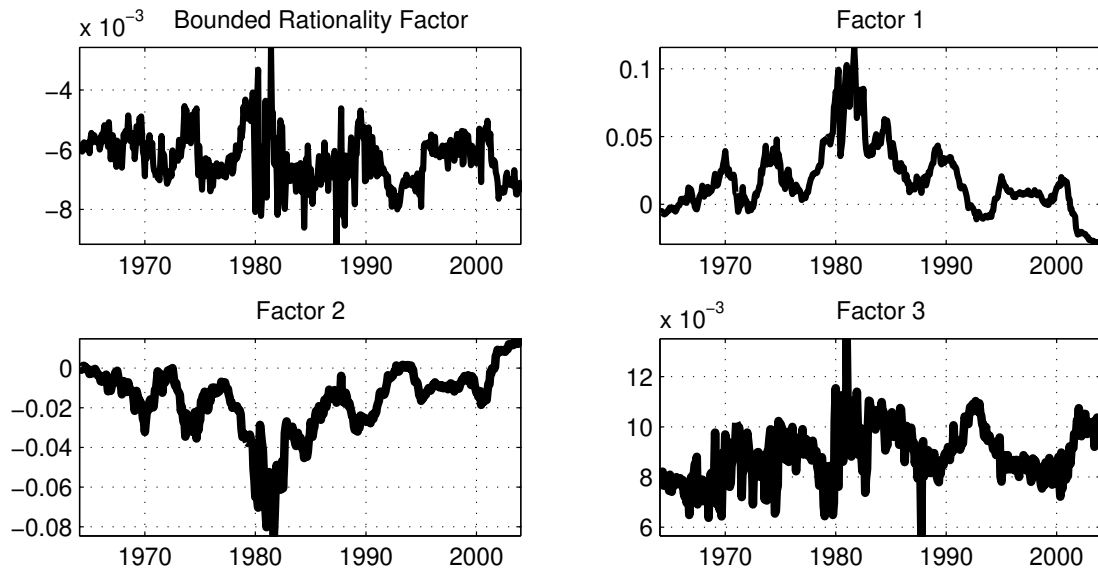
where η_t is a vector of measurement errors, \hat{A}_2 and \tilde{B}_2 are defined as

$$\hat{A}_2 = \tilde{A}_2 - \tilde{B}_2 \tilde{B}_1^{-1} \tilde{A}_1 \quad \tilde{A}_2 = A_2 - C_1 \delta_0 \quad \tilde{B}_2 = B_2 - C_2 \delta_1. \quad (4.8)$$

Given this representation, we adopt the *consistent* two-step approach suggested by Hamilton and Wu (2011): First, one can use either OLS or maximum likelihood estimation to obtain estimates of the reduced-form parameters. Second, structural parameters are obtained by using the minimum-chi-squared estimation. In this case, a chi-squared statistic measures the distance between the estimates of the reduced-form parameters and the values implied by the structural parameters, while it is minimized via numerical optimization. It is shown in Hamilton and Wu (2011) that this two-step procedure is asymptotically equivalent to ML estimation but greatly reduces the computational problems. Our estimates of structural parameters are reported in the appendix.

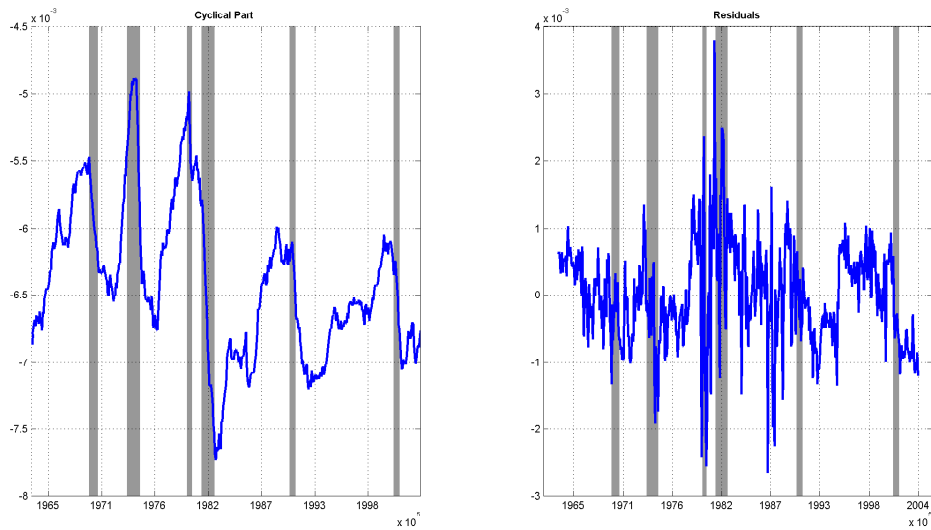
In Figure 1, we plot the estimated bounded rationality factor in the upper left panel. The other factors are contained in the rest of panels. In this figure, the data for only yields (hereafter “Yields Only”) are used in the model’s estimation. In order to see the cyclical

Figure 1: Estimated Bounded Rationality Factors



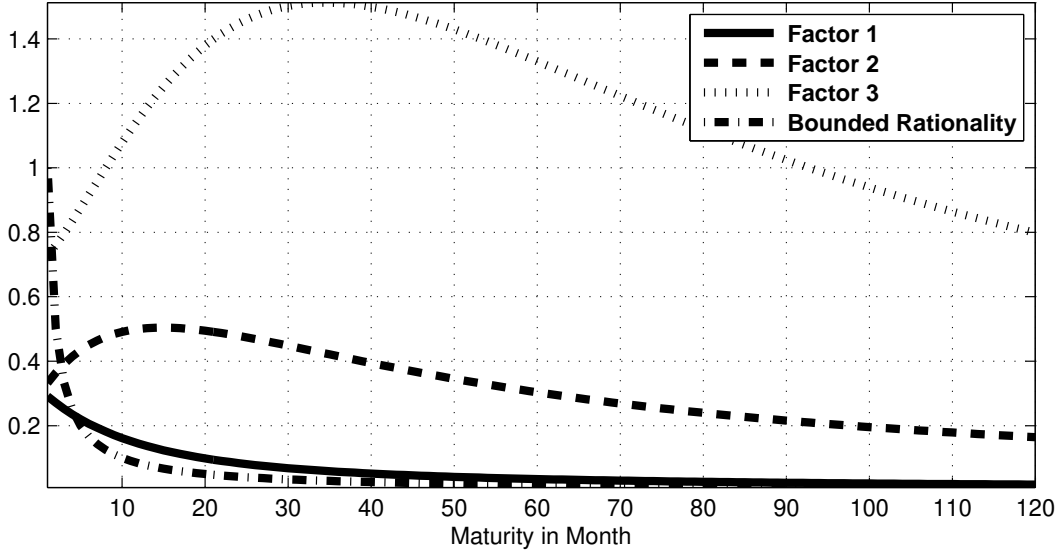
Note: The upper left panel represents the bounded rationality factor for future states. The other panels show the rest of factors.

Figure 2: Cyclical Behavior of the Bounded Rationality Factor



Note: The left panel represents the cyclical part of the bounded rationality factor for future states and the right panel contains residuals of the regression.

Figure 3: Factor Loadings



Note: This figure depicts the factor loadings of yields. The solid line corresponds to factor loadings of factor 1, the dashed line shows factor loadings of factor 2, and the dotted line is the factor loadings of factor 3. The dash-dot line represents the factor loadings of bounded rationality factor.

property of the bounded rationality factor, we have done a regression of this factor on inflation and unemployment:

$$\hat{\psi}_t = 0.0010 + 0.0166 \pi_t - 0.0314 u_t \quad (4.9)$$

(0.0001) (0.0014) (0.0022)

where $\hat{\psi}_t$ is our measure of the bounded rationality factor, π_t denotes the inflation rate, and u_t is the unemployment rate. The estimates of coefficients are significant at the level of 0.01 and the adjusted R^2 of this regression is 0.3370. Hence, our regression results imply that the bounded rationality factor is affected by macro-economic conditions.

Figure 2 displays the cyclical behavior of our measure of the bounded rationality factor, while NBER recession dates are shaded. The left panel represents the cyclical part of the bounded rationality factor that can be constructed by using the regression shown above, and the right panel contains residuals of the regression. Specifically, the cyclical part is defined as the part of our measure that is explained by the aggregate inflation and unemployment rate. In particular, we can see from this figure that our measure of the bounded rationality factor tends to drop down rapidly during recent recessions.

Figure 3 depicts the factor loadings of yields on factors. The solid line corresponds to

factor loadings of factor 1, the dashed line contains factor loadings of factor 2, and the dotted line is the factor loadings of factor 3. The dash-dot line represents the factor loadings of bounded rationality factor. Most of all, the solid line (factor 1) is almost flat over the whole maturity horizon so that this factor can be interpreted as the level factor in our model. The factor loadings of the bounded rationality factor (the dash-dot line) decline dramatically and then remain flat to the long end of the yield curve. Although factor 3 looks like the curvature factor in a reverse manner, this factor has larger effects on yields than the other factors with more influences on bond yields of relatively long maturities. Moreover, factor 2 has significant impacts on yields whose maturities are lower than around two-year. In sum, our model has less apparent distinction among conventional latent factors as level, slope, and curvature factors than prototypical three-factor affine models.

One might wonder how much would be the required magnitude of rational inattention when we want to match the actual data on yields by using our model. In order to address this issue, we estimate the value of parameter τ by using the following conditional likelihood of the inattention factor:

$$f(\hat{\psi}_{t+1}|\hat{\psi}_t, \hat{\lambda}_t, \tau) = -\frac{1}{2\hat{\lambda}'_t\hat{\lambda}_t} \|\hat{\psi}_{t+1} - \hat{\psi}_t - (1-\tau)(r_t + \frac{1}{2}\hat{\lambda}'_t\hat{\lambda}_t) - \tau^2\sigma_{v,t}^2/2\|^2 - \frac{1}{2}\log(2\pi) \quad (4.10)$$

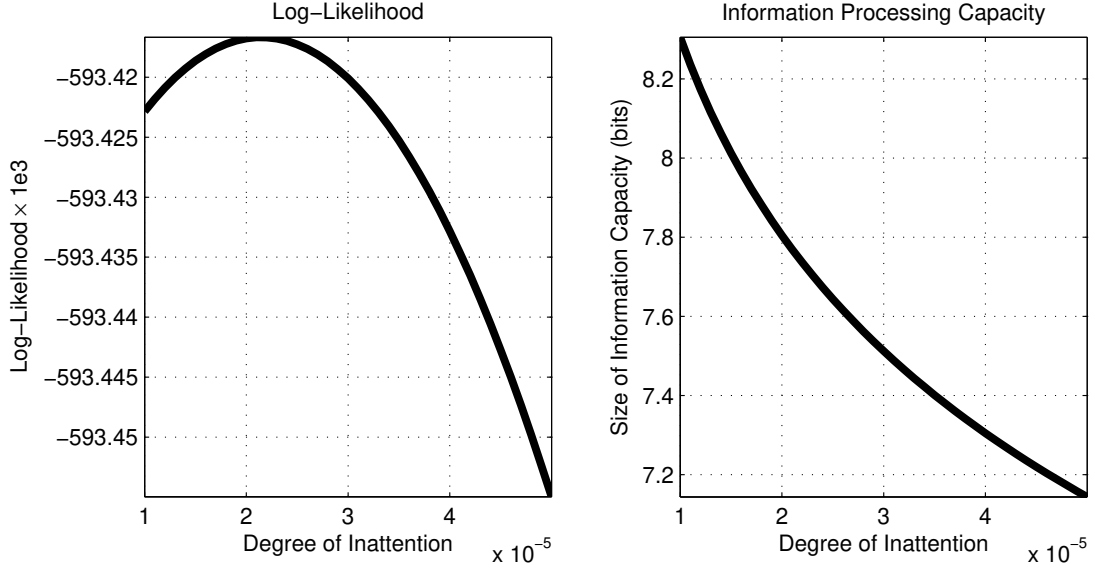
where $\hat{\lambda}_t$ is our measure of the market price of risk and $f(\hat{\psi}_{t+1}|\hat{\psi}_t, \hat{\lambda}_t, \tau)$ denotes the conditional log-likelihood of $\hat{\psi}_{t+1}$ given values of $\hat{\psi}_t$ and $\hat{\lambda}_t$. The estimated size of rational inattention is defined as follows:

$$\tau^* = \arg \max_{\tau} \sum_{t=0}^{T-1} f(\hat{\psi}_{t+1}|\hat{\psi}_t, \hat{\lambda}_t, \tau). \quad (4.11)$$

In Figure 4, we plot the likelihood function of $\hat{\psi}_t$ as a function of τ . In this figure, we can find that the log-likelihood is maximized at a small value of $(1-\tau)$ that is close to zero (2.2×10^{-5}). However, a relatively large value of τ does not necessarily mean that the magnitude of the capacity-constraint on information-processing is insignificant. In order to understand this argument, it is useful to see the relationship between the size of τ and the upper bound on the capacity of information processing as shown in the right panel. The estimate of τ implies that $\kappa = 7.7696$, while the case of pure rational expectations corresponds to $\tau = \infty$.

We now move onto the discussion of what would be the value of multiplier when we want to match the actual data on yields by using our model with risk-sensitive approach. In order to address this issue, we estimate the value of parameter θ by using the following conditional

Figure 4: Estimated Size of Rational Inattention



Note: The left panel shows the relation between the degree of rational inattention ($1 - \tau$) and log-likelihood value. The right panel depicts the relation between the degree of rational inattention ($1 - \tau$) and the information processing capacity κ (bits).

likelihood of the inattention factor:

$$f(\hat{\psi}_{t+1}^h | \hat{\psi}_t^h, \hat{\lambda}_t, \theta) = -\frac{\theta}{2\hat{\lambda}_t' \hat{\lambda}_t} \|\hat{\psi}_{t+1}^h - \hat{\psi}_t^h + \frac{1}{2}(\hat{\lambda}_t' \hat{\lambda}_t) / \theta^2\|^2 - \frac{1}{2} \log(2\pi) \quad (4.12)$$

where $f(\hat{\psi}_{t+1}^h | \hat{\psi}_t^h, \hat{\lambda}_t, \theta)$ denotes the conditional log-likelihood of $\hat{\psi}_{t+1}^h$ given a value of $\hat{\psi}_t^h$.⁶ The estimate of multiplier is defined as follows:

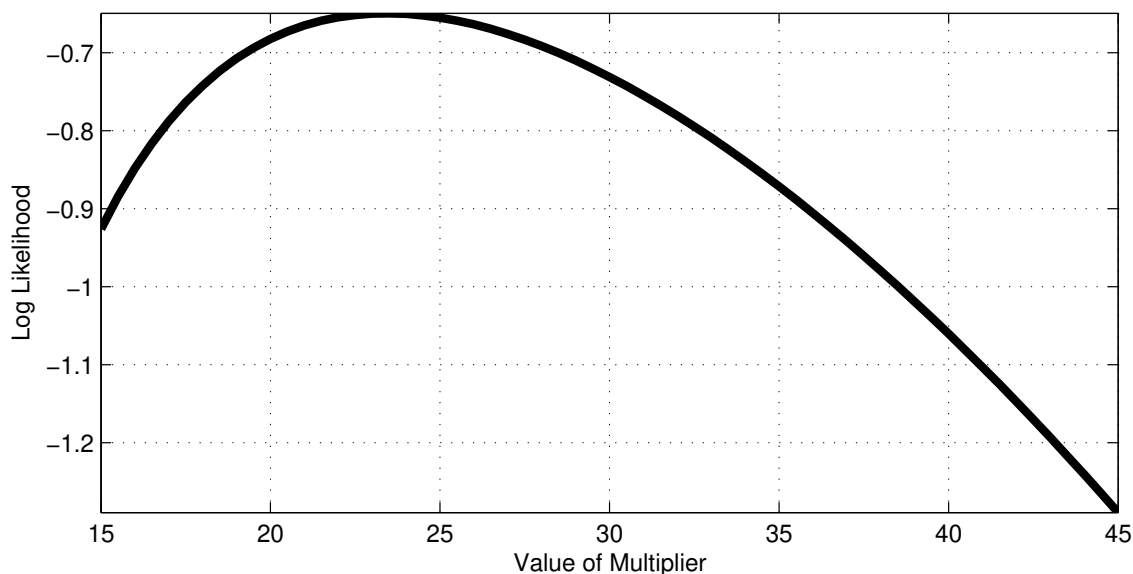
$$\theta^* = \arg \max_{\theta} \sum_{t=0}^{T-1} f(\hat{\psi}_{t+1}^h | \hat{\psi}_t^h, \hat{\lambda}_t, \theta). \quad (4.13)$$

In Figure 5, we plot the likelihood function of $\hat{\psi}_t^h$ as a function of θ . In this figure, we can find that the log-likelihood is maximized at $\theta = 23.5$. In order to see how far our model is away from the case of rational expectations, we note that the case of pure rational expectations corresponds to $\theta = \infty$.

In order to show the empirical performance of the model, we report (out-of-sample) forecast errors that are obtained from the estimated model. Specifically, we use our estimated model to forecast over the last 60 months (the out-of-sample) of our sample and record the

⁶Here, note that $\hat{\psi}_t^h = \hat{\psi}_t$, because of the equivalence result described in Proposition 4.1.

Figure 5: Estimate of Multiplier in Risk-Sensitive Approach



Note: This figure shows the relation between the value of multiplier θ and log-likelihood value. The likelihood value is maximized at $\theta = 23.5$, while the case of pure rational expectations corresponds to $\theta = \infty$.

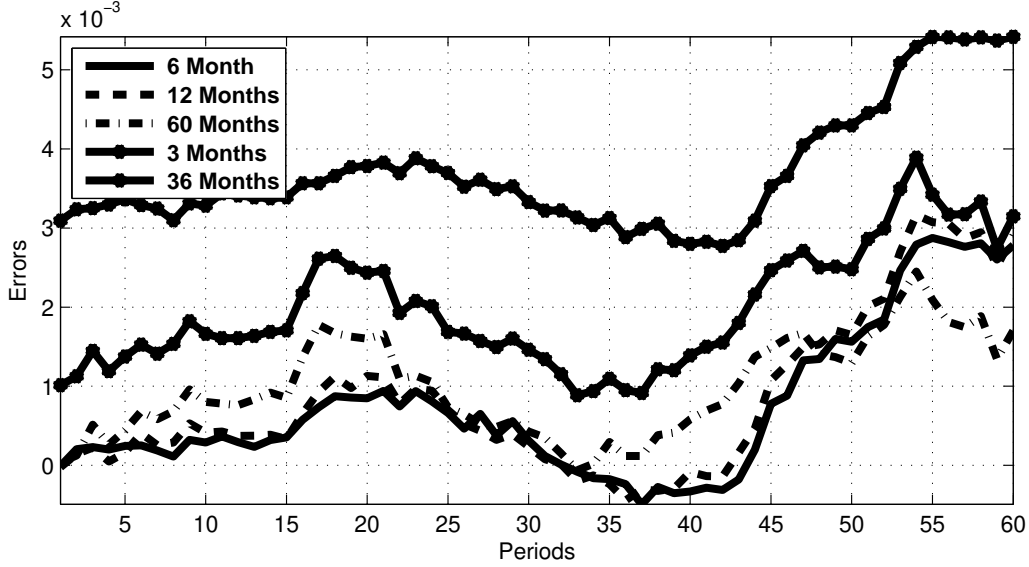
forecast error that is defined as the difference between forecast and actual values. The forecast errors tend to rise as the length of forecasting time period has increased as shown in Figure 6. The absolute size of each forecast error is less than 10^{-2} for each maturity of yields. In addition, RMSEs of yields whose maturities are six-month, one-year, five-year (without measurement errors), three-month, and five-year (with measurement errors) are 0.0093, 0.0101, 0.0090, 0.0059, and 0.0337, respectively. As a result, the forecasting power of our model is remarkably good.⁷

5 An Affine Term Structure Model with Rational Inattention for Current States

We will show that inattention factors can arise in the presence of rational inattention on current states. We then move onto the discussion of their impacts on prices of bonds. An important point of this section is that the inattention factors help explain the expected excess

⁷We also estimate the case in which macroeconomic variables such as PCE inflation rate and unemployment rate are included and find that estimation results for the size of rational inattention and the Lagrange multiplier of model misspecification are similar with those shown in this section, which are contained in the appendix.

Figure 6: Prediction Errors (Out of Sample)



Note: This figure displays (out-of-sample) forecast errors.

returns of long-term bonds, while they do not affect the current yield curve. We thus present some sufficient conditions for the inattention factors to be identical to the hidden factors emphasized in Duffee (2011).

In order to incorporate current information-tracking problems into the state vector of affine-term structure models, we assume that both risk-neutral and risk-averse agents do not have perfect information about the current realization of state variables because they are subject to a sequence of information-processing constraints. In addition, risk-neutral agents create the linear dynamics of factors under the risk-neutral measure and risk-averse agents create the linear dynamics of factors under the physical measure. The linear dynamics of factors under the physical measure corresponds to the data-generating process.

Assumption 5.1 (Factor dynamics under the risk-neutral measure) *Risk-neutral agents create the linear dynamics of factors under the risk-neutral measure. The law of motion for the state vector X_t under the risk-neutral measure is*

$$X_t = c^q + \phi^q X_{t-1} + \Sigma \epsilon_t^q \quad (5.1)$$

where X_t represents the state vector (a $n \times 1$ random vector) under the risk-neutral measure and ϵ_t follows a normal distribution $\mathcal{N}(0, I)$.

Assumption 5.2 (Factor dynamics under the physical measure) *Risk-averse agents create the linear dynamics of factors under the physical measure. The law of motion for the state vector X_t under the physical measure is*

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t. \quad (5.2)$$

The absence of arbitrage leads to tight connections between parameters in the two linear evolution equations of X_t . For example, when the market price of risk is specified as

$$\Sigma \lambda_t = \lambda_0 + \lambda_1 X_t, \quad (5.3)$$

we can obtain the following relations between parameters:

$$\phi^q = \phi - \lambda_1 \quad c^q = c - \lambda_0. \quad (5.4)$$

For this reason, although both risk-neutral and risk-averse agents do not have perfect information about the current realization of state variables, we consider only risk-neutral agents' information problems for updating their information sets in each period, given constraints on their information-processing. In addition, the specification of the market price of risk in these two equations is not identical to the one used in the previous section in order to make our inattention factor comparable to Duffee's hidden factor.

Assumption 5.3 (Decomposition of the state vector) *The state vector (denoted by X_t) consists of both target and rational-inattention state vectors: $X_t = [F_t' \hat{F}_t']'$ where F_t denotes the vector of true factors and \hat{F}_t is the vector of inattention factors. The law of motion for true factors is given by*

$$F_t = c_F + G_1 F_{t-1} + \tilde{\epsilon}_t \quad (5.5)$$

where $\tilde{\epsilon}_t$ follows a normal distribution $\mathcal{N}(0, I)$. The true factors included in F_t are not affected by the vector of inattention factors \hat{F}_t .

We point out that this decomposition of the whole state vector into the two distinct groups of factors is prevalent in models with rational inattention, while it also facilitates the analysis of the impact of the inattention factors on prices of bonds. The law of motion for true state variables is represented by a linear dynamic system that is not affected by inattention factors. In fact, state variables of agents are augmented by some random variables that are not spanned by true state variables. Hence, the presence of information processing constraint

generates an augmented linear dynamic system that includes both true and agents' state variables.

In order to get some insight about the role of inattention factors in yield curve, we begin with a two-factor model and then move onto a generalized result. Following Sims (2011), when agents want to obtain a process that is close to the target process (denoted by x_t), their optimization problem can be written as

$$\max_{h_t, \sigma_t^2} E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(-(x_t - h_t)^2 + \lambda \log \left(\frac{(\phi_{11}^q)^2 \sigma_{t-1}^2 + \sigma_{\epsilon, x}^2}{\sigma_t^2} \right) \right) \right] \quad (5.6)$$

where β represents the time discount factor of risk-neutral agents, $\sigma_{\epsilon, x}^2$ is the variance of innovations to the stochastic process $\{x_t\}_{t=0}^{\infty}$, and the state variable x_t follows an AR(1) process of the form:

$$x_t = \phi_{11}^q x_{t-1} + \epsilon_{x,t}^q. \quad (5.7)$$

The amount of information updated at period t is expressed in terms of the ratio of the variance of h_t prior to the information update to the variance of h_t after the information update ($= \sigma_t^2$) and λ measures the shadow value of information cost.⁸

The optimal solution for h_t is $h_t = E[x_t | I_t]$. Given this optimal solution for h_t , the optimization problem turns out to be

$$\max_{\sigma_t^2} E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(-\sigma_t^2 + \lambda \log \left(\frac{(\phi_{11}^q)^2 \sigma_{t-1}^2 + \sigma_{\epsilon, x}^2}{\sigma_t^2} \right) \right) \right]. \quad (5.8)$$

The solution of this problem is

$$\lambda^{-1} + \sigma_t^{-2} = \frac{\beta (\phi_{11}^q)^2}{(\phi_{11}^q)^2 \sigma_t^2 + \sigma_{\epsilon, x}^2}. \quad (5.9)$$

In order to derive a joint autoregressive representation of (x_t, h_t) by using the solution of this problem, it is convenient to assume that agents observe a noisy signal $z_t = x_t + v_t$ for the true state variable x_t and v_t represents the noise. In this case, the conditional expectation of x_t is $E[x_t | z_t] = \gamma z_t$ where $\gamma = \sigma_x^2 / (\sigma_x^2 + \sigma_v^2)$. When the inattention factor h_t is defined as $h_t = E[x_t | z_t]$, we have $h_t = \gamma(x_t + v_t)$. The expectation error for the inattention factor can be used to derive the following equation: $h_t = E[h_t | I_{t-1}] + \gamma(x_t - E[x_t | I_{t-1}]) + \gamma v_t$. As a result, the solution of this problem produces a joint autoregressive representation of (x_t, h_t) :

$$\begin{pmatrix} x_t \\ h_t \end{pmatrix} = \begin{pmatrix} \phi_{11}^q & 0 \\ \gamma \phi_{11}^q & (1 - \gamma) \phi_{11}^q \end{pmatrix} \begin{pmatrix} x_{t-1} \\ h_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & \gamma \\ \gamma & \gamma \end{pmatrix} \begin{pmatrix} \epsilon_{x,t} \\ v_t \end{pmatrix}. \quad (5.10)$$

⁸The mutual information measures the reduction of uncertainty of X due to the knowledge of Y : $I(X, Y) = h(X) - h(X|Y)$ where $h(X)$ is computed at period $t-1$ information set and $h(X|Y)$ is computed at period t information set. $h(X|Y) = \frac{1}{2} \log((2\pi e)^n |\Sigma_t|)$ and $h(X) = \frac{1}{2} \log((2\pi e)^n |M_t|)$. Hence, the mutual information is $I(X, Y) = \frac{1}{2} (\log(|\Sigma_t|) - \log(|M_t|))$.

Given a joint autoregressive representation of (x_t, h_t) under risk-neutral measure, a joint autoregressive representation of (x_t, h_t) under the physical measure can be written as

$$\begin{pmatrix} x_t \\ h_t \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} x_{t-1} \\ h_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ \gamma & \gamma \end{pmatrix} \begin{pmatrix} \epsilon_{x,t} \\ \nu_t \end{pmatrix}. \quad (5.11)$$

The first-order condition for σ_t^2 and the information-processing constraint allow us to obtain the relation between the Lagrange multiplier and the upper bound on the capacity of processing information. Specifically, the first-order condition leads to a non-linear equation for σ^2 :

$$(\sigma^2 + \lambda)(\sigma^2(\phi_{11}^q)^2 + \sigma_{\epsilon,x}^2) = \lambda\beta(\phi_{11}^q)^2\sigma^2. \quad (5.12)$$

The substitution of the first-order condition into the information-processing constraint leads to the following equation:

$$\lambda = \frac{\sigma^2}{\beta(\phi_{11}^q)^2 \exp(-\kappa) - 1}. \quad (5.13)$$

Hence, we can solve these two equations to express the Lagrange multiplier and the variance of the inattention factor in terms of the upper bound on the capacity of processing information. As a result, in the case of a rational inattention problem for current states, the value of γ is endogenously determined by solving equations (5.12) and (5.13) given a relation of $\sigma^2 = \gamma \sigma_x^2$. However, values of σ_x^2 and σ_ν^2 are exogenously set in the case of a signal extraction problem for current states. In this case, the value of γ ($= \sigma_x^2 / (\sigma_x^2 + \sigma_\nu^2)$) is exogenously determined.

We now impose no-arbitrage conditions in pricing bonds. In this case, the logarithms of bond prices are affine in the state vector:

$$p_{n,t} = \bar{a}_n + \bar{b}'_n \begin{pmatrix} x_t \\ h_t \end{pmatrix}. \quad (5.14)$$

The coefficients \bar{a}_n and \bar{b}_n follow the difference equations:

$$\bar{b}'_{n+1} = \bar{b}'_n \phi^q + \bar{b}'_1, \quad \bar{a}_{n+1} = \bar{a}_n + \bar{b}'_n c^q + \frac{1}{2} \bar{b}'_n \Sigma \Sigma' \bar{b}_n + \bar{a}_1 \quad (5.15)$$

for $n = 1, 2, \dots, T$, and where ϕ^q and c^q are defined as $\phi^q = \phi - \lambda_1$ and $c^q = c - \lambda_0$.

Assumption 5.4 (Irrelevance of inattention factor for the short-term nominal interest rate) *In the two-factor example described above, the short-term interest rate at period t (denoted by r_t) is not affected by h_t . Hence, the short-term nominal interest rate can be written as*

$$r_t = -\bar{a}_1 - \bar{b}_{1,1} x_t \quad (5.16)$$

where $\bar{b}_{1,1}$ denotes the first element of \bar{b}_1 .

It might be worthwhile to discuss the implication of this assumption. We can set $\bar{a}_1 = 0$ and $\bar{b}_{1,1} = -1$ in the case of $x_t = r_t$ as in the example of Duffee (2011). Assumption 5.4 then implies that the short-term nominal interest rate is perfectly observable. However, in the case of $x_t \neq r_t$, more assumptions should be added to this interpretation in order to rationalize assumption 5.4. In this case, we might need to assume that the central bank's feedback rule requires the short-term nominal interest rate to respond to only perfectly observable variables when the central bank sets its target on the short-term nominal interest rate.

Proposition 5.1 (Irrelevance of inattention factor for the current yield curve)

Suppose that assumptions 5.1 - 5.4 hold. Then, a change in h_t does not affect the time- t yields.

Given assumptions 5.1 - 5.4, the formula of factor loadings implies that the vector of coefficients for the price at period t of bonds that mature at period $t + m$ can be written as

$$\bar{b}_m = \begin{pmatrix} \bar{b}_{1,1}(1 - \phi_{11}^q)^{-1}(1 - (\phi_{11}^q)^m) \\ 0 \end{pmatrix}. \quad (5.17)$$

Hence, all yields at period t are not affected by h_t .

Proposition 5.2 (Impact of inattention factor for expected excess returns for long-term bonds)

Suppose that assumptions 5.1 - 5.4 hold. Then, a change in h_t affects expected excess returns for long-term bonds. Specifically, suppose that agents purchase m -period bonds at period t and sell them at period $t + 1$. The excess return for holding m -period bonds is then given by

$$E_t[xr_{t,t+1}^{(m)}] = \bar{a}_{m-1} - \bar{a}_m + \bar{a}_1 - c_r x_t + c_h h_t \quad (5.18)$$

where c_r and c_h can be defined as

$$c_r = \bar{b}_{1,1} \{ (1 - \phi_{11}^q)^{-1} (1 - (\phi_{11}^q)^m - (1 - (\phi_{11}^q)^{m-1}) \phi_{11}) + 1 \} \quad (5.19)$$

$$c_h = \bar{b}_{1,1} (1 - \phi_{11}^q)^{-1} (1 - (\phi_{11}^q)^{m-1}) \phi_{12} \quad (5.20)$$

The result of this proposition can be obtained by substituting the recursive formula of factor loadings and the evolution equation of the state vector into the definition of the expected excess returns. Given this proposition, a non-zero value of ϕ_{12} helps allow for the possibility that the inattention factor affects the excess return of holding long-term bonds.

Given assumptions 5.1 - 5.4, the inattention factor is comparable to the hidden factor defined in Duffee (2011). The reason behind this argument is that a snapshot of the time- t yield curve conveys no information about the inattention factor. In fact, Duffee (2011) argues that a hidden factor has opposite effects on expected future interest rates and bond risk premia. For example, news that increases risk premia and simultaneously leads investors to believe the central bank will soon lower short-term interest rates. The rise in risk premia raises long-term bond yields immediately, while the expected drop in short rates induces an immediate decrease in these yields. As a result, our analysis opens the possibility that a hidden factor can arise endogenously when agents are subject to rational inattention.

We now move onto a general case in which there are more than two factors in order to spell out a general version of sufficient conditions for the presence of bounded rationality factors. In general, the behavior of the state vector X_t under the physical measure is described in a compact form as a first-order Gaussian VAR:

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t \quad (5.21)$$

where ϵ_t follows a normal distribution $\mathcal{N}(0, I)$. The law of motion for the state vector X_t under the equivalent measure is then given by

$$X_t = c^q + \phi^q X_{t-1} + \Sigma^q \epsilon_t. \quad (5.22)$$

In particular, when the market price of risk is specified as

$$\Sigma \lambda_t = \lambda_0 + \lambda_1 X_t, \quad (5.23)$$

there is a one-to-one relation between parameters in the evolution equation of X_t :

$$\phi^q = \phi - \lambda_1 \quad c^q = c - \lambda_0. \quad (5.24)$$

The presence of information constraint under the equivalent measure implies that matrix ϕ^q is block-triangular

$$\phi^q = \begin{pmatrix} \phi_{11}^q & 0 \\ \phi_{21}^q & \phi_{22}^q \end{pmatrix} \quad (5.25)$$

It is also possible to partition the matrix in the market price of risk as follows:

$$\lambda_1 = \begin{pmatrix} \lambda_{11,1} & \lambda_{12,1} \\ \lambda_{21,1} & \lambda_{22,1} \end{pmatrix} \quad (5.26)$$

In this case, the matrix ϕ under the physical measure is given by

$$\phi = \begin{pmatrix} \phi_{11}^q + \lambda_{11,1} & \lambda_{12,1} \\ \phi_{21}^q + \lambda_{21,1} & \phi_{22}^q + \lambda_{22,1} \end{pmatrix} \quad (5.27)$$

Hence, the presence of hidden factors requires that some elements of sub-matrix $\lambda_{12,1}$ are not zero. The reason for this result is that some non-zero elements of sub-matrix $\lambda_{12,1}$ guarantee a non-zero sub-matrix ϕ_{12} .

We now show how we can satisfy these sufficient conditions in the presence of rational inattention. For a moment, we assume that the law of motion for factors under rational expectation ($= F_t$) is given by

$$F_t = c_F + G_1 F_{t-1} + \tilde{\epsilon}_t. \quad (5.28)$$

Based on the solution to the linear-quadratic optimization problem included in appendix A, when \hat{F}_t denotes the resulting rational-inattention factors, the law of motion for this set of inattention factors can be written as

$$\hat{F}_t = \hat{c}_F + (I - \tilde{\Sigma}\Lambda^{-1})G_1 F_{t-1} + \tilde{\Sigma}\Lambda^{-1}G_1 \hat{F}_{t-1} + \tilde{\Sigma}\Lambda^{-1}\tilde{\epsilon}_t + \tilde{\Sigma}\Lambda^{-1}\hat{\epsilon}_t. \quad (5.29)$$

As a result, the whole state vector that consists of both true and rational-inattention state vectors ($X_t = [F_t' \hat{F}_t']'$) turns out to be

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t, \quad \epsilon_t = \begin{pmatrix} \tilde{\epsilon}_t \\ \hat{\epsilon}_t \end{pmatrix} \quad (5.30)$$

where matrices ϕ , Σ and vector c are defined as

$$\phi = \begin{pmatrix} G_1 & 0 \\ (I - \tilde{\Sigma}\Lambda^{-1})G_1 & \tilde{\Sigma}\Lambda^{-1}G_1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} I & 0 \\ \tilde{\Sigma}\Lambda^{-1} & \tilde{\Sigma}\Lambda^{-1} \end{pmatrix} \quad c = \begin{pmatrix} c_F \\ \hat{c}_F \end{pmatrix} \quad (5.31)$$

In addition, $\tilde{\Sigma}$ is the covariance matrix of inattention factors and Λ is the covariance matrix of the shocks to information set under rational inattention denoted by $\hat{\epsilon}$. Our result can be then summarized in the following proposition.

Proposition 5.3 (Sufficient conditions for inattention factors to be hidden factors) *Suppose that the short-term nominal interest rate is not affected by inattention factors:*

$$r_t = -\bar{a}_1 - \bar{b}'_{1,F} F_t \quad (5.32)$$

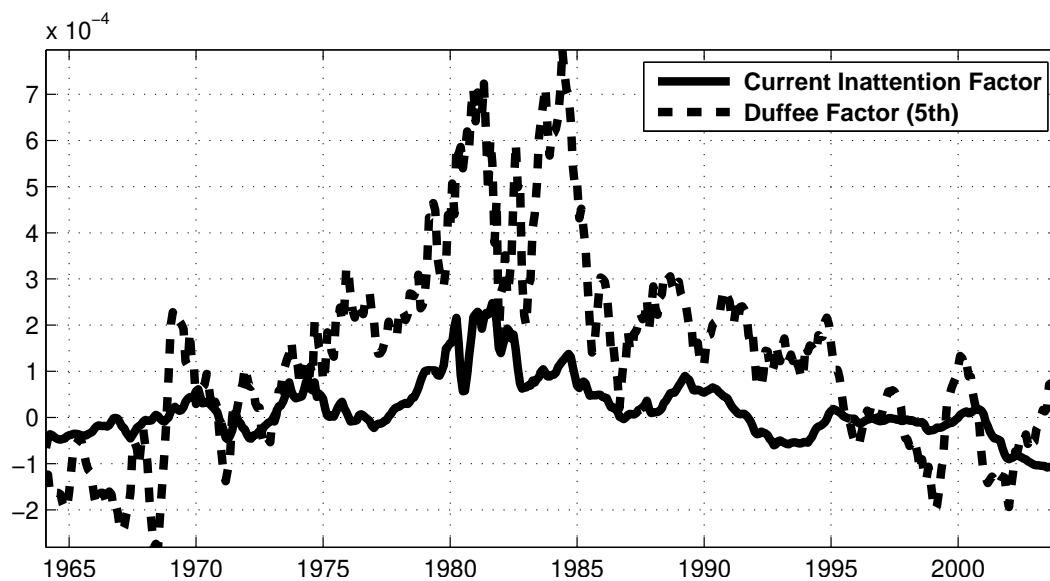
where $\bar{b}_{1,F}$ is a sub-vector of \bar{b}_1 that is relevant for F_t . In addition, the state vector consists of both true and rational-inattention state vectors ($X_t = [F_t' \hat{F}_t']'$):

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t, \quad \epsilon_t = \begin{pmatrix} \tilde{\epsilon}_t \\ \hat{\epsilon}_t \end{pmatrix} \quad (5.33)$$

where matrices ϕ and Σ are defined as

$$\phi = \begin{pmatrix} G_1 & 0 \\ (I - \tilde{\Sigma}\Lambda^{-1})G_1 & \tilde{\Sigma}\Lambda^{-1}G_1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} I & 0 \\ \tilde{\Sigma}\Lambda^{-1} & \tilde{\Sigma}\Lambda^{-1} \end{pmatrix} \quad (5.34)$$

Figure 7: Inattention Factor versus Hidden Factor



Note : The solid line shows the estimated rational inattention factor for current states. The dashed line depicts the fifth factor of Duffee (2011) that is estimated by using the Kalman filter.

In this case, inattention factors included in \hat{F}_t can be interpreted as hidden factors that are defined in Duffee (2011). The optimization problem for the generalized specification of inattention factor is included in the appendix.

In order to estimate the model with the rational inattention for current states, we assume that the first latent factor is subject to the rational inattention constraint. The coefficient of the inattention factor in the short-rate equation is set to be zero, which is a sufficient condition adopted in Duffee (2011). In addition, we assume that all yields are subject to measurement errors, so that we use the Kalman filter to estimate our model.⁹

As shown in Figure 7, both inattention and hidden factors tend to co-move over the sample period, while different identification conditions are used for the two factors. Each of these factors appears to have a long swing over the whole sample period. For example, the two factors rise in early 1980s and then move down before in 1985. In addition, these two factors stay below their sample averages in early 2000s.

In order to check the cyclical property of the rational inattention factor for current states,

⁹The estimates of structural parameters and their standard errors are included in the appendix.

we have done a regression of this factor on inflation and unemployment:

$$\hat{h}_t = \frac{-0.0001}{(0.0087 \times 10^{-3})} + \frac{0.0018}{(0.0965 \times 10^{-3})} \pi_t + \frac{0.0005}{(0.1519 \times 10^{-3})} u_t \quad (5.35)$$

where \hat{h}_t is our measure of the hidden factor, π_t denotes the inflation rate, and u_t is the unemployment rate. The estimates of coefficients are significant at the level of 0.01 and the adjusted R^2 of this regression is 0.4813. Hence, our regression results imply that the rational inattention factor for current states is affected by macro-economic conditions.

We now turn to estimation results of excess-returns regressions. Given the assumption that a single linear combination of states determines the compensation investors demand to face fixed-income risk from period t to period $t + 1$, we construct the risk premium factor and investigate how much this factor can explain actual excess returns of bonds, following Duffee (2011). In particular, the risk premium factor can be written as

$$RP_t = \lambda_1(L)' X_t \quad (5.36)$$

where RP_t denotes the risk premium factor and $\lambda_1(L)$ is the vector of coefficients of the level factor in the market price of risk. In order to see the impact of this risk premium factor on the excess return of bonds, we use a series of predictive regressions for excess returns of bonds:

$$xr_{t,t+1}^m = b_0 + b_1 RP_t + \epsilon_{t,t+1} \quad (5.37)$$

where orthogonality is imposed between the risk premium factor and the residual. The adjusted R^2 of these predictive regressions for bonds with maturities of 2, 3, and 4 months are 0.1401, 0.0867, and 0.0251 respectively. We note that, although our regressions are focused on short-term yields, our estimates of adjusted R^2 for bonds with maturities of 2- and 3-month are higher than those reported in Duffee (2011). For example, Duffee (2011)'s report on adjusted R^2 is 0.048 from predictive regressions of excess monthly returns to five-year zero-coupon bonds when the risk premium factor is constructed by using a multi-factor model.

6 Conclusion

We have incorporated rational inattention and robustness for model misspecification into macro-finance affine models in order to investigate the role of bounded rationality factor. Our empirical results indicate that substantial amounts of information capacity constraint and robustness preference for model misspecification are needed to explain the observed behavior

of yields when we use affine term-structure models. Moreover, out-of-sample forecasts of our models indicate that the incorporation of bounded rationality can enhance an affine term-structure model's capability to forecast yields. We also have evaluated the empirical performance of our approach in terms of the role of bounded rationality factors to help explain excess returns of bond holdings.

Appendix

A Parameter Estimates on Future States

A.1 Estimation Method

The dynamics of factors (denoted by X_t) are described by the following state equation:

$$X_t = c + \phi X_{t-1} + \Sigma \epsilon_t. \quad (\text{A.1})$$

The set of yields without measurement errors has the following representation:

$$Y_{1t} = A_1 + B_1 X_t + C_1 \psi_t \quad (\text{A.2})$$

and the short rate satisfies the following equation:

$$r_t = \delta_0 + \delta_1 X_t + \psi_t. \quad (\text{A.3})$$

To eliminate ψ_t from these equations, we subtract r_t from measurement equations of Y_{1t} :

$$\tilde{Y}_{1t} = \tilde{A}_1 + \tilde{B}_1 X_t \quad (\text{A.4})$$

where $\tilde{Y}_{1t} = Y_{1t} - C_1 r_t$, $\tilde{A}_1 = A_1 - C_1 \delta_0$, and $\tilde{B}_1 = B_1 - C_1 \delta_1$. By substituting this equation into the law of motion for X_t , we can derive a VAR(1) representation for yield factor Y_{1t} as follows:

$$\tilde{Y}_{1t+1} = \hat{A}_1 + \hat{\phi}_1 \tilde{Y}_{1t} + u_{t+1} \quad (\text{A.5})$$

where \hat{A}_1 , $\hat{\phi}_1$ and u_{t+1} are defined as

$$\hat{\phi}_1 = \tilde{B}_1 \phi \tilde{B}_1^{-1}, \quad \hat{A}_1 = (I - \hat{\phi}_1) \tilde{A}_1 + \tilde{B}_1 c, \quad \Omega \left(\equiv \text{var}(u_{t+1}) \right) = \tilde{B}_1 \Sigma \Sigma' \tilde{B}_1'. \quad (\text{A.6})$$

In addition, as it was discussed in section 4, the bond yields with measurement errors can be described as follows:

$$\tilde{Y}_{2t} = \hat{A}_2 + \hat{\phi}_2 \tilde{Y}_{1t} + \eta_t, \quad (\text{A.7})$$

where η_t is a vector of measurement errors, \hat{A}_2 and $\hat{\phi}_2$ are defined as

$$\hat{A}_2 = \tilde{A}_2 - \tilde{B}_2 \tilde{B}_1^{-1} \tilde{A}_1, \quad \hat{\phi}_2 = \tilde{B}_2 \tilde{B}_1^{-1}, \quad \tilde{A}_2 = A_2 - C_1 \delta_0, \quad \tilde{B}_2 = B_2 - C_2 \delta_1. \quad (\text{A.8})$$

By OLS estimation, we can obtain estimates of reduced-form parameters such as $\hat{\phi}_1$, \hat{A}_1 , $\hat{\phi}_2$, and \hat{A}_2 . Since the number of reduced-form parameters are larger than that of structural parameters in equations (A.6) and (A.8), we estimate ϕ , Σ , λ_0 , and λ_1 by using the Minimum Chi-Square Estimation (MCSE) as suggested in Hamilton and Wu (2011):

$$\min_{\phi, \Sigma, \lambda_0, \lambda_1} \left[\begin{aligned} & \|\hat{A}_{1,OLS} - (I - \hat{\phi}_1) \tilde{A}_1 - \tilde{B}_1 c\|^2 + \|\hat{\phi}_{1,OLS} - \tilde{B}_1 \phi \tilde{B}_1^{-1}\|^2 \\ & + \|\hat{A}_{2,OLS} - \tilde{A}_2 + \tilde{B}_2 \tilde{B}_1^{-1} \tilde{A}_1\|^2 + \|\hat{\phi}_{2,OLS} - \tilde{B}_2 \tilde{B}_1^{-1}\|^2 \end{aligned} \right] \quad (\text{A.9})$$

Table 1: Estimates of Parameters (Inattention for Future States)

ϕ	0.8617 (0.3647)	0	0
	0.2730 (0.1611)	0.9553 (0.1555)	0
	0.0773 (0.1481)	0.2749 (0.1289)	0.9337 (0.1673)
Σ	0.0013 ($10^{-4} \times 0.9891$)	0	0
	0	0.0004 ($10^{-4} \times 0.2083$)	0
	0	0	0.0006 ($10^{-4} \times 0.1766$)
λ'_0	0.4039 (0.2431 $\times 10^{-7}$)	0.0284 (0.1599 $\times 10^{-7}$)	0.4509 (0.4430 $\times 10^{-7}$)
λ_1	0.0536 (0.0027)	0	0
	0.4203 (0.0014)	0.0247 (0.0019)	0
	0.3985 (0.0003)	0.3553 (0.0002)	0.4532 (0.0002)

B Parameter Estimation on Current States

Since we assume that bond yields of all maturities have measurement errors, we follow Duffee (2011)'s estimation method. Specifically, we employ the Kalman filter to extract unobserved factors. In this case, the transition equation of factor X_t can be written as follows:

$$X_{t+1} = c + \phi X_t + \Sigma \epsilon_{t+1}. \quad (\text{B.1})$$

In addition, the rational inattention factor h_t is not eliminated but added to the three factors of X_t :

$$\begin{pmatrix} X_t \\ h_t \end{pmatrix} = \phi^q \begin{pmatrix} X_{t-1} \\ h_{t-1} \end{pmatrix} + \Sigma \begin{pmatrix} \epsilon_{X,t} \\ \nu_t \end{pmatrix}. \quad (\text{B.2})$$

$$\text{where } \phi^q = \begin{pmatrix} \phi_{11}^q & 0 & 0 & 0 \\ \phi_{21}^q & \phi_{22}^q & 0 & 0 \\ \phi_{31}^q & \phi_{32}^q & \phi_{33}^q & 0 \\ \gamma\phi_{11}^q & 0 & 0 & (1-\gamma)\phi_{11}^q \end{pmatrix} \text{ and } \Sigma\Sigma' = \begin{pmatrix} 1 & 0 & 0 & \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma & 0 & 0 & \gamma^2(1+\sigma_v^2) \end{pmatrix}.$$

In our estimation, the measurement equation includes three-month, one-year, three-year, and five-year maturity bond yields: ($Y_t = [y_{3,t} \ y_{12,t} \ y_{36,t} \ y_{60,t}]'$), while the measurement equation can be written as follows:

$$Y_t = A + B \begin{pmatrix} X_{t-1} \\ h_{t-1} \end{pmatrix}. \quad (\text{B.3})$$

C Inattention Factors in the Model with Information-Processing Constraint on Latent Factors

Following Sims (2011), we specify the general case of linear-quadratic control with information cost as follows:

$$\min_{X_t, \hat{F}_t, \Sigma_t} E_0 \sum_{t=0}^{\infty} \beta^t (F_t' A F_t + F_t' B X_t + X_t' C X_t + \lambda H_t) \quad (\text{C.1})$$

subject to

$$F_{t+1} = G_1 F_t + \tilde{\epsilon}_{t+1} \quad (\text{C.2})$$

$$H_t = \frac{1}{2} (\log |M_t| - \log |\tilde{\Sigma}_t|) \quad (\text{C.3})$$

$$M_{t+1} = \Omega + G_1 \tilde{\Sigma}_t G_1' \quad (\text{C.4})$$

$$\tilde{\epsilon}_t | F_s, X_s, s < t \sim \mathcal{N}(0, \Omega) \quad (\text{C.5})$$

$$F_t | \mathcal{I}_t \sim \mathcal{N}(\hat{F}_t, \tilde{\Sigma}_t) \quad (\text{C.6})$$

$$\{X_t, X_{t-1}, \dots\} \subset \mathcal{I}_t \quad (\text{C.7})$$

In order to solve this problem, we break down the original optimization problem into two distinct optimization problems. The first one is a conventional linear-quadratic stochastic control problem. We do this by rewriting the quadratic part of the original objective function in terms of $\{\hat{F}_t\}_{t=0}^{\infty}$. We also rewrite the evolution equation of F_t in terms of \hat{F}_t . The next one is a deterministic optimization problem for information flow that helps determine the optimal $\tilde{\Sigma}_t$ subject to the information processing constraint.

As discussed above, we solve a prototypical linear-quadratic control problem whose quadratic loss is expressed in terms of the expected value of true state vector conditional on the current information set. We choose a stochastic process to minimize its distance from the expected value of true state vector conditional on the current information set. The linear-quadratic problem is specified as

$$\min_{X_t, \hat{F}_t} E_0 \sum_{t=0}^{\infty} \beta^t (\hat{F}_t' A \hat{F}_t + \hat{F}_t' B X_t + X_t' C X_t) \quad (\text{C.8})$$

Table 2: Estimates of Parameters (Inattention for Current States)

ϕ	0.0842 (0.2268 $\times 10^{-3}$)	0	0	1.0447 (0.1531 $\times 10^{-3}$)
	0.0272 (0.2551 $\times 10^{-3}$)	0.0071 (0.2062 $\times 10^{-3}$)	0	0
	-0.2656 (0.3150 $\times 10^{-3}$)	0.2786 (0.2460 $\times 10^{-3}$)	0.4217 (0.1972 $\times 10^{-3}$)	0
	0.0841 (0.2265 $\times 10^{-3}$)	0.5904 (0.2615 $\times 10^{-3}$)	0.4845 (0.2506 $\times 10^{-3}$)	0.0001 (0.0002 $\times 10^{-3}$)
Σ	0.7188 ($10^{-5} \times 0.0932$)	0	0	0.6952 ($10^{-5} \times 0.0963$)
	0	1.0000	0	
	0	0	1.0000	
	0.6952 ($10^{-5} \times 0.0963$)	0	0	0.7182 ($10^{-5} \times 0.1060$)
λ'_0	0.0127 (0.2057 $\times 10^{-5}$)	0.0185 (0.1992 $\times 10^{-5}$)	0.0404 (0.1746 $\times 10^{-5}$)	0.0041 (0.3002 $\times 10^{-5}$)
λ_1	0.0029 (0.2649 $\times 10^{-5}$)	0	0	0
	0.0075 (0.2958 $\times 10^{-5}$)	0.0050 (0.1496 $\times 10^{-5}$)	0	0
	0.0033 (0.2361 $\times 10^{-5}$)	-0.0025 (0.2389 $\times 10^{-5}$)	0.0001 (0.2113 $\times 10^{-5}$)	0
	0.0058 (0.2870 $\times 10^{-5}$)	0.0059 (0.2615 $\times 10^{-5}$)	0.0048 (0.2506 $\times 10^{-5}$)	0.0027 (0.2060 $\times 10^{-5}$)

subject to

$$\hat{F}_{t+1} = G_1 \hat{F}_t + \hat{\epsilon}_{t+1} \quad (\text{C.9})$$

$$\hat{\epsilon}_t = \hat{F}_t - F_t + G_1(F_{t-1} - \hat{F}_{t-1}) + \tilde{\epsilon}_t \quad (\text{C.10})$$

where \hat{F}_t is the expected value of true state vector conditional on the current information set. As a result, when the law of motion for the factors under rational expectation ($= F_t$) is

$$F_t = G_1 F_{t-1} + \tilde{\epsilon}_t \quad (\text{C.11})$$

the dynamics of factors under rational inattention ($= \hat{F}_t$) is

$$\begin{pmatrix} F_t \\ \hat{F}_t \end{pmatrix} = \phi \begin{pmatrix} F_{t-1} \\ \hat{F}_{t-1} \end{pmatrix} + \Sigma \begin{pmatrix} \tilde{\epsilon}_t \\ \hat{\epsilon}_t \end{pmatrix} \quad (\text{C.12})$$

where matrices ϕ and Σ are defined as

$$\phi = \begin{pmatrix} G_1 & 0 \\ (I - \tilde{\Sigma}\Lambda^{-1})G_1 & \tilde{\Sigma}\Lambda^{-1}G_1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} I & 0 \\ \tilde{\Sigma}\Lambda^{-1} & \tilde{\Sigma}\Lambda^{-1} \end{pmatrix}. \quad (\text{C.13})$$

In addition, $\tilde{\Sigma}$ is the covariance matrix of inattention factors and Λ is the covariance matrix of the shocks to information set under rational inattention denoted by $\hat{\epsilon}$.

Second, the resulting deterministic optimization problem for information flow turns out to be

$$\max_{\tilde{\Sigma}_t} \sum_{t=0}^{\infty} \beta^t (-\text{trace}(\tilde{\Sigma}_t A) + \lambda H_t) \quad (\text{C.14})$$

subject to

$$H_t = \frac{1}{2}(\log |M_t| - \log |\tilde{\Sigma}_t|) \quad (\text{C.15})$$

$$M_{t+1} = \Omega + G_1 \tilde{\Sigma}_t G_1' \quad (\text{C.16})$$

The first-order condition for $\tilde{\Sigma}_t$ is

$$A = \beta \lambda G_1 M_{t+1}^{-1} G_1' - \lambda \tilde{\Sigma}_t^{-1} \quad (\text{C.17})$$

We thus find a value of Σ_t by solving the following equation:

$$\lambda^{-1} A = \beta G_1 (\Omega + G_1 \Sigma_t G_1')^{-1} G_1' - \Sigma_t^{-1}. \quad (\text{C.18})$$

We now discuss how to compute covariance matrices $\tilde{\Sigma}$ and Λ . Following Harvey and Rossi (2004), we can write an updating equation for $\tilde{\Sigma}_t$ as follows:

$$\tilde{\Sigma}_{t+1} = M_t - M_t (M_t + \Lambda_t)^{-1} M_t \quad (\text{C.19})$$

where M_t is defined as

$$M_t = G_1 \tilde{\Sigma}_t G_1' + \Omega \quad (\text{C.20})$$

We now discuss how to compute Λ and $\tilde{\Sigma}$. The steady-state version of this updating equation is

$$\tilde{\Sigma} = M - M(M + \Lambda)^{-1} M \quad (\text{C.21})$$

The well-known solution to this equation is

$$\Lambda^{-1} = \tilde{\Sigma}^{-1} - M^{-1} \quad (\text{C.22})$$

Hence, once we compute the matrix $\tilde{\Sigma}$, it is possible to compute Λ and M by using equations specified above. Given this fact, we now use the first-order condition to compute $\tilde{\Sigma}$. In particular, following Luo and Young (2010), the first-order condition can be rewritten as

$$\tilde{\Sigma}^{-1} = (Q\tilde{\Sigma}Q + Q_0)^{-1} - A/\lambda \quad (\text{C.23})$$

where Q and Q_0 are defined as

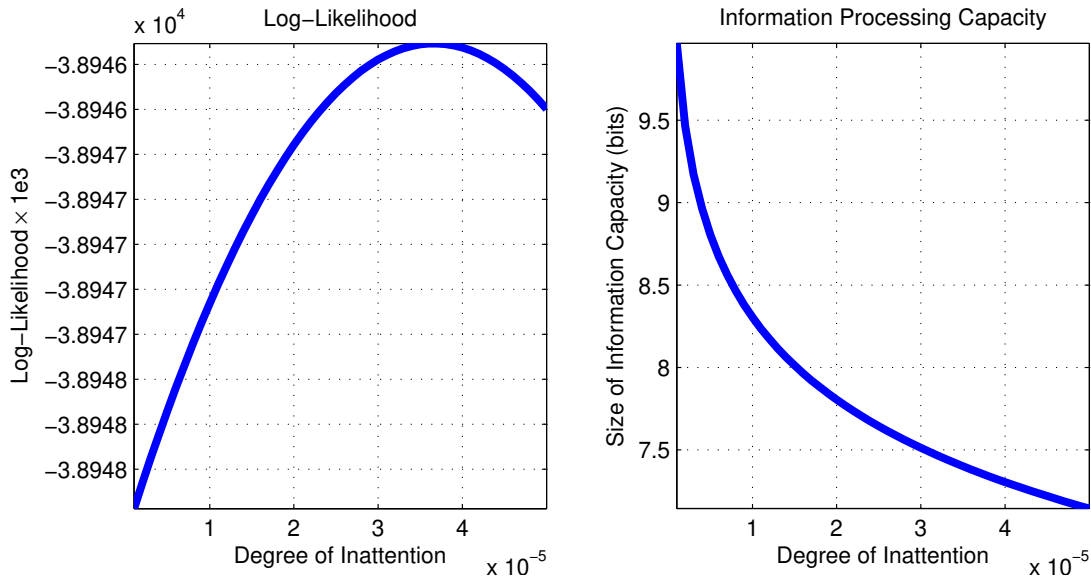
$$Q = \beta^{-1/2}(G'_1)^{-1}G_1 \quad Q_0 = \beta^{-1}(G'_1)^{-1}\Omega G_1^{-1} \quad (\text{C.24})$$

We then solve this equation for $\tilde{\Sigma}$ and then check if the difference between M and $\tilde{\Sigma}$ is positive definite.

D Models with Macroeconomic Variables

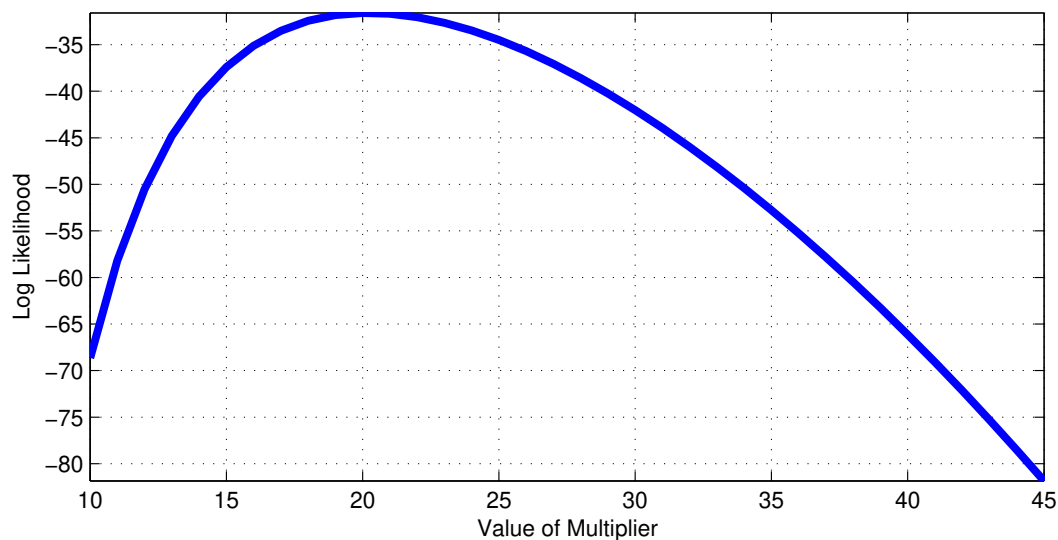
This section contains estimation results of models with both yield data and macroeconomic variables such as the PCE inflation rate and unemployment rate.

Figure 8: Estimated Size of Rational Inattention



Note: The left panel shows the relation between the degree of rational inattention ($1 - \tau$) and log-likelihood value. The right panel depicts the relation between the degree of rational inattention ($1 - \tau$) and the information processing capacity κ (bits).

Figure 9: Estimate of Multiplier in Risk-Sensitive Approach



Note: This figure shows the relation between the value of multiplier θ and log-likelihood value. The likelihood value is maximized at $\theta = 20.2$, while the case of pure rational expectations corresponds to $\theta = \infty$.

References

- Ang, Andrew and Monika Piazzesi. 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50, pp. 745-787.
- Bauer, Michael, Glenn Rudebusch, and Jing Cynthia Wu. 2011. Unbiased Estimation of Dynamic Term Structure Models. Unpublished Manuscript. University of Chicago and Federal Reserve Bank of San Francisco.
- Chen, R. and L. Scott, 1993. Maximum Likelihood Estimation for A Multi-Factor Equilibrium Model of the Term structure of Interest Rates. *Journal of Fixed Income* 3, pp. 14-31.
- Cochrane, John and Monika Piazzesi. 2005. Bond Risk Premia. *American Economic Review* 95(1). pp. 128-160
- Duffee, Greg. 2011. Information in (and not in) the Term Structure. *Review of Financial Studies* 24, pp. 2895-2934.

- Hamilton, James and Jing Cynthia Wu. 2011. Testable Implications of Affine Term Structure Models. Forthcoming in Journal of Econometrics.
- Hansen, Lars and Thomas Sargent, 2011. Wanting Robustness in Macroeconomics. Handbook of Monetary Economics Vol. 3 (Ch 20). pp. 1097-1157.
- Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu, 2011. A New Perspective on Gaussian Dynamic Term Structure Models. Review of Financial Studies 24 (3). pp. 926-970.
- Litterman, R. and Jose Scheinkman. 1991. Common Factors Affecting Bond Returns. Journal of Fixed Income 1, pp. 51-61.
- Luo, Yulei and Eric Young. 2010. Signal Extraction and Rational Inattention. Unpublished Manuscript. University of Hong Kong and University of Virginia.
- Piazzesi, Monika. 2010. Affine Term Structure Models. Handbook of Financial Econometrics (Ch 12). pp. 691-765.
- Sims, Christopher. 2011. Rational Inattention and Monetary Economics. Handbook of Monetary Economics Vol. 3 (Ch 4). pp. 155-161.