Rational Inattention and Aggregate Fluctuations

Yulei Luo† University of Hong Kong

Eric R. Young‡ University of Virginia

Abstract

This paper introduces the rational inattention hypothesis (RI) – that agents process information subject to finite channel constraints – into a stochastic growth model with permanent technology shocks. We find that RI raises consumption volatility relative to output by introducing an endogenous demand shock. Furthermore, it is shown that incorporating RI can provide an additional internal propagation mechanism (measured by the impulse response function and the autocorrelation function of output growth) and generate higher variance of forecastable movements in output. However, we find that RI cannot resolve these puzzles in the RBC literature – weak internal propagation and low variance of forecastable movements in output, even with what appears to be a very low capacity channel.

JEL Classification Numbers: E13, E21, E32, G12.

Keywords: Rational Inattention, Consumption Volatility, Propagation Mechanism

† School of Economics and Finance, The University of Hong Kong, Hong Kong, email address: yluo@princeton.edu.
‡ Corresponding author. Department of Economics, University of Virginia, Charlottesville, VA 22904, email address: ey2d@virginia.edu.
1. Introduction

It is well-known that the standard RBC models have weak internal propagation mechanism: aggregate output and labor supply almost completely trace out the exogenous stochastic process of aggregate technology. That is, the model generates realistic output and employment dynamics only to the extent that it assumes them in the exogenous driving processes. For example, Cogley and Nason (1993) find that in a typical RBC model, output dynamics are determined primarily by impulse dynamics and that the endogenous propagation mechanism is very weak. In the data, US GDP has an obvious trend-reverting component, which can be characterized by hump-shaped responses to a transitory shock – as noted by Blanchard and Quah (1989) – while the standard RBC models can only generate monotonic responses of output and labor supply to transitory shocks. A related observation that standard models cannot replicate is that output growth displays positive autocorrelation at short horizons.\(^1\) The models generate zero persistence in output growth if the technology shock is assumed to follow a random walk.\(^2\) Rotemberg and Woodford (1996) highlight a third related anomaly: standard RBC models can produce only about 1 percent of the actual variance of forecastable movements in output; since all the dynamics are driven by exogenous shocks, with random walk technology there is no forecastable component.


In this paper, we explore whether introducing the Rational Inattention Hypothesis into a simple stochastic growth model can help overcome some of the deficiencies discussed above. This departure from the standard models is motivated by Sims (2003). Rational expectation models assume implicitly that agents can process information costlessly and respond immediately to market signals.

\(^1\) As documented in Cogley and Nason (1995), at lags of 1 and 2 quarters the sample autocorrelations are positive and statistically significant. For higher lags, the autocorrelations are mostly negative and statistically insignificant.

\(^2\) For standard RBC models – such as Christiano and Eichenbaum (1992) – output growth is close to white noise for plausible parameters.
or shocks to the economic system. This assumption is too strong to be consistent with the inborn ability of human beings: it requires unlimited information-processing capacity. As discussed in Sims (1998, 2003), if people have limited information-processing capacity their decisions can depend on observations only through their own communication channels. In other words, they cannot digest aggregate or individual information and market signals immediately. As a result, their responses to shocks may be delayed by the need to slowly absorb just how the state of the world has changed.

Recently, there have been several papers examining imperfect information-processing in alternative frameworks. For example, Woodford (2001) and Adam (2005) analyze optimal monetary policy and inflation and output dynamics with imperfect common knowledge, finding that the imperfect information models can generate highly persistent effects on real activity. Luo (2007) examines consumption dynamics under information process constraints in the permanent income hypothesis model, while Maćkowiak and Wiederholt (2005) explore the implications for optimal sticky prices. Sims (2005, 2006) investigates consumption and saving decision under more realistic preferences, but he is only able to explore a two-period model. Van Nieuwerburgh and Veldkamp (2005) examine the effects of information constraints on the home bias puzzle.

In this paper we study a simple stochastic growth model with permanent shocks to technology and information processing constraints to explore whether this friction can resolve some existing puzzles in the RBC literature. We suppose that the social planner devotes the limited information capacity to observing the univariate state of the economy, but since that capacity is limited the state is not perfectly observed. As a result, the consumption-savings and labor-leisure decisions are made relative to a noisy signal of the true state, and information about changes in the true state is not entirely incorporated into forecasts. In other words, the planner is forced to take some time to digest new information about the state. The result of this processing problem is that the planner faces a signal extraction problem in which the distribution of the noise is endogenous, a generalization of the problem studied in Kasa (1995).

The particular puzzles that we explore are the ones mentioned above. We first ask whether RI can improve the strength of the internal propagation mechanism in the basic model; we find the

---

3Extensions of the nonlinear model can be found in Lewis (2006), Tutino (2007), and Batchuluun, Luo, and Young (2007).

4We assume the shock to technology is permanent, so there is only one state variable (the ratio of capital to this permanent shock) in the model economy. This assumption simplifies computation greatly, but may limit the generality of our results.
improvement very small, even for very low channel capacities, because the estimated and actual states evolve in a very similar manner. As a result, the autocorrelation function for output growth has positive values at one and two lead/lags, but the coefficients are too small relative to the data. Second, we explore whether RI implies a hump-shaped response of output to a permanent technology shock; we find that RI does imply a delayed response of output to a technology shock, but the hump is too small. Third, we explore whether RI can increase the variance of the forecastable movements in output; here we find almost no improvement relative to the benchmark model studied in Rotemberg and Woodford (1996).

Rational inattention does greatly increase the relative volatility of consumption in the model, however. Since the planner cannot accurately observe the true capital stock, consumption must depend on a noisy signal instead. In effect, RI introduces a second shock into the model which tends to impact consumption volatility more than other aggregates (in our model, it has no effect on labor volatility and consequently only minor effects on output). When we use consumption volatility as a moment condition to estimate the channel capacity, we find that our model requires that individuals only process 0.6059 bits of macroeconomic information every quarter; that is, they only remove 57% of the uncertainty after observing the new signals. While this number might seem very low, it is difficult to measure the amount of attention individuals allocate to monitoring the aggregate economy; it is much more likely that the majority of their limited attention is dedicated to observing the more volatile idiosyncratic variables.

2. A Stochastic Growth Model with Rational Inattention

This section lays out the problem of a social planner suffering from Rational Inattention. In the standard model without distortions, Pareto-optima can be decentralized as competitive equilibria. Although real economies are decentralized, the allocation is one that would be chosen by a central authority. Information-processing constraints seem likely to break this equivalence, although we are not certain. As argued in Sims (2003), although the information processing randomness is idiosyncratic, it is still reasonable to assume that a considerable part of the responses from information constraints is common across agents. Hence, we can imagine here that some properties of aggregate fluctuations generated from our social planning economy may still hold in a corresponding

---

5 We will discuss the link between the rate of information transmission and the reduction in uncertainty in detail in section 2.

6 We discuss the problems associated with the competitive equilibrium relative to the planning solution in appendix A.
decentralized economy.

Following King, Plosser, and Rebelo (1988) and Rotemberg and Woodford (1996), we assume that aggregate technology shocks are permanent so that the model generates a stochastic growth path. One of the advantages of this assumption in our framework is that it is easy to derive the properties of the endogenous noise since there is an unique state variable (the ratio of capital stock to aggregate technology).\(^7\) Before setting up and solving the stochastic growth model with RI hypothesis, it is helpful to present the standard model without RI first. The problem can be stated as

\[
\max_{\{C_t, L_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left[ \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} V(1-L_t) \right] \tag{2.1}
\]

subject to

\[
C_t + I_t \leq Y_t \tag{2.2}
\]
\[
Y_t \leq K_t^\alpha (Z_t L_t)^{1-\alpha} \tag{2.3}
\]
\[
K_{t+1} \leq I_t + (1-\delta)K_t \tag{2.4}
\]
\[
\log Z_{t+1} = \mu + \log Z_t + \omega \varepsilon_{t+1} \tag{2.5}
\]

where \(C_t\) is consumption, \(L_t\) is total labor supply, \(I_t\) is gross investment, \(K_t\) is the capital stock, \(Y_t\) is total output, and \(Z_t\) is a random walk technology process with drift \(\mu\) and white noise errors with unit variance. This time-separable power utility function with \(\gamma = 1\) becomes the logarithmic form \(\log (C_t) + V(1-L_t)\), where the function \(V\) is suitably redefined. The parameters satisfy \(\alpha, \beta \in (0, 1), \delta \in [0, 1], \text{ and } \mu, \omega, \gamma > 0\). Combining equation (2.2), (2.3), and (2.4) yields the law of motion for capital,

\[
K_{t+1} = K_t^\alpha (Z_t L_t)^{1-\alpha} - C_t + (1-\delta)K_t, \tag{2.6}
\]

Note that the unit root assumption of \(Z_t\) here is roughly consistent with US time series data, as the model implies a unit root in output and consumption as well.\(^8\)

---

\(^7\)As discussed in Sims (2003) and Luo (2007), it is much more difficult to derive the properties of endogenous noise in the multivariate case since it involves how to allocate the agent’s channel capacity optimally when observing multiple state variables. Feasible methods to solve this problem are still being explored. Most researchers simply ignore the problem either by assuming that Gaussian noise is still optimal or by assuming that only one state variable is subject to the Shannon channel constraint.

\(^8\)To put it a slightly different way, it is not inconsistent with US time series data on real GDP. See King et al. (1990).
Following Hansen (1985), we also assume indivisible labor: \( V (1 - L_t) = \eta (1 - L_t) \). Although the output process is nonstationary, the following ratios are stationary:

\[
\tilde{K}_t = \frac{K_t}{Z_t}, \quad \tilde{Y}_t = \frac{Y_t}{Z_t}, \quad \tilde{C}_t = \frac{C_t}{Z_t}, \quad \tilde{I}_t = \frac{I_t}{Z_t};
\]

under the restrictions for preferences given above \( L_t \) is already stationary. Now we can normalize the evolution equation of capital ((2.6)) to obtain

\[
\tilde{K}_{t+1} = \exp (-\mu - \omega \varepsilon_{t+1}) \left( \tilde{K}_t^\alpha L_t^{1-\alpha} - \tilde{C}_t + (1 - \delta) \tilde{K}_t \right)
\]

(2.7)

Similarly, the objective function can also be normalized in terms of \( \tilde{C}_t \) and the effective discount factor becomes \( \beta \exp(-\mu - \omega \varepsilon_{t+1}) (1 - \gamma) \). As usual, the solution to this standard welfare maximization problem is a set of contingency rules that specify how much to consume and work in each period as a function of the state in that period. We follow Campbell (1994) and use the log-linearization approach to solve the model. First, the efficiency conditions for the planning problem are

\[
\eta = \tilde{C}_t^{-\gamma} (1 - \alpha) \frac{\tilde{Y}_t}{L_t} \quad (2.8)
\]

\[
\tilde{C}_t^{-\gamma} = \beta E_t \left[ \exp (-\mu - \omega \varepsilon_{t+1}) \left( \tilde{C}_{t+1}^{-\gamma} \right) \left( 1 + \alpha \frac{\tilde{Y}_{t+1}}{\tilde{K}_{t+1}} - \delta \right) \right] \quad (2.9)
\]

\[
\exp (\mu + \omega \varepsilon_{t+1}) \tilde{K}_{t+1} = \tilde{K}_t^\alpha L_t^{1-\alpha} + (1 - \delta) \tilde{K}_t - \tilde{C}_t \quad (2.10)
\]

\[
\lim_{t \to \infty} \beta^{\gamma} \tilde{K}_{t+1} \tilde{C}_t^{-\gamma} = 0. \quad (2.11)
\]

Second, the balanced-growth path is defined by the three-equation system

\[
\eta = (1 - \alpha) \frac{\tilde{K}_t^\alpha L_t^{1-\alpha}}{\tilde{C}_t^\gamma} \quad (2.12)
\]

\[
1 = \exp (-\mu) \left( \tilde{K}_{t+1}^\alpha L_t^{1-\alpha} - \tilde{C}_t \right) + 1 - \delta \quad (2.13)
\]

\[
1 = \beta \exp (-\mu) \left( 1 - \delta + \alpha \tilde{K}_t^\alpha L_t^{1-\alpha} \right). \quad (2.14)
\]

A plausible calibration for this economy is given in Table 1 – it roughly matches the usual ratios in the US data as well as the time series behavior of the Solow residual.\(^9\)

\[^9\text{Specifically, we choose the parameters } (\beta, \delta, \eta) \text{ to match a capital-output ratio of 11.5, a consumption-output ratio of 0.5, and an efficiency coefficient of 0.1.}\]
We log-linearize the above system around the unique interior steady state and then derive optimal linear decision rules; this procedure lets us derive some approximate analytical results. The resulting linear system is

\[ c_t = \alpha k_t - \alpha l_t \]  
\[ k_{t+1} = -\omega E_{t+1} + \frac{\tilde{K}^{\alpha} L^{1-\alpha}}{K \exp(\mu)} (\alpha k_t + (1 - \alpha) l_t) - \frac{\tilde{C}}{K \exp(\mu)} c_t + \frac{(1 - \delta) \tilde{K}}{K \exp(\mu)} k_t \] 
\[ -\gamma c_t = E_t \left[ -\gamma c_{t+1} + \frac{\alpha \tilde{K}^{\alpha-1} L^{1-\alpha}}{1 - \delta + \alpha \tilde{K}^{\alpha-1} L^{1-\alpha}} ((\alpha - 1) k_{t+1} + (1 - \alpha) l_{t+1}) \right] \];

we use lowercase letters to denote deviations from the steady state of any stationary variable. The optimal decision rules take the form

\[ c_t = \psi k_t \]  
\[ l_t = \phi k_t \]

for some coefficients \((\psi, \phi)\). We solve for these coefficients using the approach from Campbell (1994), where we insert these guesses into the decision rules and solve the resulting undetermined coefficients system. Since the above system defines a quadratic equation we choose the root that implies stationarity in the law of motion for \(k\). Hence, the original nonstationary stochastic growth model can be regarded as a simple stationary optimal control problem, in which \(k_t\) is the unique state variable and \(c_t\) and \(l_t\) are control variables.

Our model with RI follows Sims (2003).\(^{10}\) We assume that the social planner maximizes the representative agent’s utility function subject to both the usual flow budget constraint and the information processing constraints that will be specified later.\(^{11}\) The sequential problem for the planner is

\[ \hat{v}(\tilde{k}_t) = \max_{\{c_t, l_t, D_t\}} E_0 \left[ \sum_{t=0}^{\infty} \beta^t u(c_t, l_t) \right] \]  

\(^{10}\)To introduce RI into our model we must assume that the coefficient of relative risk aversion \(\gamma\) is close enough to 1. Luo (2007) contains a discussion.

\(^{11}\)In Appendix C, we solve Hall’s version of the stochastic growth model with RI explicitly, in which we use quadratic utility and additive technology shocks. The closed-form solution of this special case shows that the important effects of RI are present in the log-linearized equations for the main model.
subject to

\[ k_{t+1} = \frac{\bar{K}^{\alpha} L^{1-\alpha}}{\bar{K} \exp(\mu)} (\alpha k_t + (1 - \alpha) l_t) - \frac{\bar{C}}{\bar{K} \exp(\mu)} c_t + \frac{(1 - \delta)}{\bar{K} \exp(\mu)} k_t \omega \varepsilon_{t+1} \]  \tag{2.21}

\[ k_{t+1} | I_{t+1} \sim D_{t+1} \]  \tag{2.22}

\[ k_t | I_t \sim D_t, \]  \tag{2.23}

given \( k_0 | I_0 \sim N(\hat{k}_0, \Sigma_0) \) and the requirement that the rate of information flow at \( t + 1 \) implicit in the specification of the distributions, \( D_t \) and \( D_{t+1} \), be less than channel capacity.\(^{12}\) \( k_t \) is the actual state variable, \( \hat{k}_t \) is the estimated state variable (to be explained in the next paragraph), and \( I_t \) is the information available at time \( t \). The expectation is formed under the assumption that \( \{c_t, l_t\}_0^\infty \) are chosen under the information processing constraints.\(^{13}\)

In the presence of RI, the social planner in the economy cannot observe the state(s) perfectly. Because the observing the state involves information transfer at a limited channel capacity \( \kappa \), the observations are contaminated by error; the social planner gets to choose the distribution of the endogenous error optimally, however. In this case the actual state variable is not the traditional one (e.g., the normalized capital stock level \( \bar{k}_t \) in this model), but the so-called information state: the distribution of the state variable \( k_t \) conditional on the information set available at time \( t \), \( I_t \). In other words, it expands the state space to the space of distributions on \( k_t \), creating a “curse of dimensionality” problem.\(^{14}\) Fortunately, the above problem can be approximated by a Linear-Quadratic-Gaussian framework in which the conditional distributions are Gaussian; the first two moments, the conditional mean \( \hat{k}_t \) and the conditional covariance matrix \( \Sigma_t \), are therefore sufficient to characterize the effective state. Hence, \( \hat{v}(\hat{k}_t) \) is the value function under information processing constraints, and \( v(k_t) \) is the value function derived from the standard model, where the social planner is assumed to have unlimited channel capacity and thus can observe the state perfectly. Finally, we define the loss function at \( t \) due to imperfect information as the difference between these two value functions: \( \Delta v = v(k_t) - \hat{v}(\hat{k}_t) \).

We use the concept of entropy from information theory to characterize the rate of information

\(^{12}\)As shown in Sims (2003, 2005), the two distributions are approximately normal in our setup.

\(^{13}\)It is important to note that the sequence of capital is not chosen by the planner; rather, it is determined as a residual once consumption and labor effort have been chosen. The planner cannot choose next period’s capital directly because then it would be perfectly observed.

\(^{14}\)Sims (2005) contains a discussion of some of the complications that such a model would present. One major obstacle is the distribution chosen by the agents is a complicated discrete-continuous mixture, making even simulation methods difficult to apply. Further computational evidence is found in Batchuluun, Luo, and Young (2007).
flow and then use the reduction in entropy as a measure of information gain.\footnote{Entropy is a measure of the uncertainty about a random variable and is defined as the expectation of the natural log of the density function, \( E[\log(f(X))] \). See Shannon (1948) and Cover and Thomas (1991) for details.} With finite capacity, the social planner will choose a signal that reduces the uncertainty of the state. Formally, this idea can be described by the information constraint

\[
\mathcal{H}(k_{t+1}|I_t) - \mathcal{H}(k_{t+1}|I_{t+1}) \leq \kappa, \tag{2.24}
\]

where \( \kappa \) is the consumer’s information channel capacity, \( \mathcal{H}(k_{t+1}|I_t) \) denotes the entropy of the state prior to observing the new signal at \( t+1 \), and \( \mathcal{H}(k_{t+1}|I_{t+1}) \) the entropy after observing the new signal. \( \kappa \) imposes an upper bound on the amount of information flow – that is, the change in the entropy – that can be transmitted in any given period.\footnote{If the base for logarithms is 2, the unit used to measure information flow is called a ‘bit’, and if we use the natural logarithm \( \log(e) \) the unit is called a ‘nat’. Hence, 1 nat is equal to \( \log_2(e) = 1.4427 \) bits.}

Following Sims (2003, 2005) and Luo (2007), we know that \( k_t|I_t \sim N(\hat{k}_t, \Sigma_t) \). Therefore, ((2.24)) can be rewritten as

\[
\log|\Psi_t| - \log|\Sigma_{t+1}| \leq 2\kappa \tag{2.25}
\]

where \( \Sigma_{t+1} = \text{var}_{t+1}(k_{t+1}) \) and \( \Psi_t = \text{var}_t(k_{t+1}) \) are the posterior and the prior variance-covariance matrices of the state vector. Note that here we use the fact that the entropy of a Gaussian random variable is equal to half of its logarithm variance plus some constant term. In the univariate state case this information constraint completes the characterization of the optimization problem; for the multivariate state case, we need another information constraint:

\[
\Psi_t \succeq \Sigma_{t+1}, \tag{2.26}
\]

(where \( \succeq \) means the difference between the two matrices is a positive semi-definite matrix). This constraint embodies the restriction that precision in the estimates of the state cannot be improved by forgetting some components (making the change in their entropy negative) and using that extra capacity to reduce other components by more than \( \kappa \).\footnote{In other work we refer to this constraint as a ‘no subsidization’ constraint.}

Given that certainty equivalence holds in for our system of linear equations, introducing RI
implies that

\[ c_t = \psi \hat{k}_t \]  \hspace{1cm} (2.27)

\[ l_t = \phi \hat{k}_t \]  \hspace{1cm} (2.28)

where \( \hat{k}_t = E(k_t|I_t) \) is the information state; the conditional distribution of \( k_t \) given information at \( t \) is \( N(\hat{k}_t, \sigma^2_{k,t}) \).\(^{18}\) The linearized dynamic resource constraint can be rewritten as

\[
k_{t+1} = \frac{1}{\beta} k_t - \frac{\tilde{C}}{K \exp(\mu)} c_t + \frac{\tilde{K}^\alpha L^{1-\alpha}}{K \exp(\mu)} (1 - \alpha) l_t - \omega \varepsilon_{t+1}; \quad (2.29)
\]

taking the conditional variance implies

\[
\text{var}_t (k_{t+1}) = \omega^2 + \frac{1}{\beta^2} \sigma^2_{k,t}. \quad (2.30)
\]

From the information-processing constraint we can obtain

\[
\kappa = \frac{1}{2} \left( \log \left( \omega^2 + \frac{1}{\beta^2} \sigma^2_{k,t} \right) - \log \left( \sigma^2_{k,t+1} \right) \right).
\]

Hence, the steady state of the filtering process yields \( \overline{\sigma}^2_k = \frac{\omega^2}{\exp(2\kappa) - \beta^2} \). The agent behaves as if observing a noisy measurement \( k^*_{t+1} = k_{t+1} + \xi_{k,t+1} \) with error variance

\[
\text{var} (\xi_{k,t+1}) = \frac{\left( \omega^2 + \beta^{-2} \overline{\sigma}^2_k \right) \overline{\sigma}^2_k}{\omega^2 + (\beta^{-2} - 1) \overline{\sigma}^2_k} \quad (2.31)
\]

The planner is solving a signal extraction problem in which the variance of the noise is chosen optimally, subject to the entropy constraint. Agents with more channel capacity generally observe a less noisy signal about the state of the world:

\[
\frac{\partial \text{var} (\xi_{k,t+1})}{\partial \kappa} = \frac{2\omega^2 \exp(2\kappa) \beta^2 (1 - \exp(4\kappa) \beta^2)}{(1 - \exp(2\kappa) \beta^2)^2 (1 - \exp(2\kappa))^2} < 0
\]

as long as \( \frac{1}{4} \log (\beta^{-2}) > \kappa \). For a typical quarterly business cycle calibration (say, \( \beta = 0.99 \)) this would require \( \kappa < 0.005 \) nat, which seems sufficiently low that it can be safely ignored. More

\(^{18}\)Since we only study a one state case, to avoid confusion we use \( \sigma^2_{k,t} \) to denote the conditional variance of the state variable. We use the natural logarithm of the variance to measure uncertainty and the unit of information flow is the nat. The flow of information provides us the measure of the reduction in uncertainty.
patient agents choose a lower variance as well:

\[
\frac{\partial \text{var}(\xi_{k,t+1})}{\partial \beta} = \frac{2\omega^2 \exp(2\kappa)\beta}{(1 - \exp(2\kappa)\beta^2)^2 (1 - \exp(2\kappa))} < 0.
\]

More patient agents face higher costs in the future of having a suboptimal capital stock, so they compensate by observing more carefully.

We next derive the Kalman filter equation that describes the evolution of the perceived state:

\[
\hat{k}_{t+1} = \frac{1}{\beta} \hat{k}_t - \frac{\tilde{C}}{K \exp(\mu)} c_t + \frac{\tilde{K}^\alpha L^{1-\alpha}}{K \exp(\mu)} (1 - \alpha) l_t \\
+ \theta \left( k_{t+1} + \xi_{k,t+1} - \left( \frac{1}{\beta} \hat{k}_t - \frac{\tilde{C}}{K \exp(\mu)} c_t + \frac{\tilde{K}^\alpha L^{1-\alpha}}{K \exp(\mu)} (1 - \alpha) l_t \right) \right). 
\]

This can be written more simply as

\[
\hat{k}_{t+1} = (1 - \theta) G\hat{k}_t + \theta \left( k_{t+1} + \xi_{k,t+1} \right), 
\]

where \( \theta = \frac{\sigma^2_k}{\text{var}(\xi_{k,t+1})} \) is the optimal weight on the observation and \( G = \frac{1}{\beta} - \frac{\tilde{C}}{K \exp(\mu)} \psi + \frac{\tilde{K}^\alpha L^{1-\alpha}}{K \exp(\mu)} (1 - \alpha) \phi \). Note that

\[
\theta = \frac{1 + \frac{\beta^{-2} - 1}{\exp(2\kappa) - \beta^{-2}}}{1 + \frac{\beta^{-2} - 1}{\exp(2\kappa) - \beta^{-2}}} = \frac{\exp(2\kappa) - 1}{\exp(2\kappa)}. 
\]

Thus,

\[
\lim_{\kappa \to \infty} \frac{\exp(2\kappa) - 1}{\exp(2\kappa)} = 1.
\]

In other words, the weight of the new signal is increasing with channel capacity, and increases to 1 as \( \kappa \) gets infinitely-large, so that the standard model is a special case of our model where \( \kappa = \infty \). The weight is independent of the patience of the agent and the variance of the shocks hitting the economy.

The evolution of the economy can be described by two equations, one each for the evolution of the true state \( k_t \) and for the evolution of the perceived state \( \hat{k}_t \) (see Appendix B for derivations):

\[
k_{t+1} = \frac{1}{\beta} k_t + \left( G - \frac{1}{\beta} \right) \hat{k}_t - \omega \varepsilon_{t+1} 
\]

\[
\hat{k}_{t+1} = (1 - \theta) G\hat{k}_t + \theta \left( k_{t+1} + \xi_{k,t+1} \right).
\]
We can now derive the expression of $\Delta \hat{k}_{t+1}$, the change in the information state, as

$$\Delta \hat{k}_{t+1} = (G - 1) \hat{k}_t - \frac{\theta}{\beta} \left( \frac{1 - \theta}{1 - (1 - \theta) \beta^{-1}} \right) \omega_\varepsilon_t + \theta \xi_{k,t} + \theta \varepsilon_{t+1} + \theta \xi_{k,t+1} \tag{2.38}$$

where $L$ is the lag operator and we use the formula $k_t - \hat{k}_t = \frac{(1 - \theta) \omega_\varepsilon_t + \theta \xi_{k,t}}{1 - (1 - \theta) \beta^{-1} L}$. Similarly, we can obtain the change in the true state as

$$\Delta k_{t+1} = G \hat{k}_t - k_t - \frac{1}{\beta} \left( \frac{1 - \theta}{1 - (1 - \theta) \beta^{-1}} \right) \omega_\varepsilon_t + \theta \xi_{k,t} + \theta \varepsilon_{t+1}. \tag{2.39}$$

We now have the expression of changes in log transformed consumption and labor,

$$\Delta c_{t+1} = \psi \Delta \hat{k}_{t+1} = \psi \left( (G - 1) \hat{k}_t - \frac{\theta}{\beta} \left( \frac{1 - \theta}{1 - (1 - \theta) \beta^{-1}} \right) \omega_\varepsilon_t + \theta \xi_{k,t} + \theta \varepsilon_{t+1} + \theta \xi_{k,t+1} \right) \tag{2.40}$$

$$\Delta L_{t+1} = \phi \Delta \hat{k}_{t+1} = \phi \left( (G - 1) \hat{k}_t - \frac{\theta}{\beta} \left( \frac{1 - \theta}{1 - (1 - \theta) \beta^{-1}} \right) \omega_\varepsilon_t + \theta \xi_{k,t} + \theta \varepsilon_{t+1} + \theta \xi_{k,t+1} \right), \tag{2.41}$$

and we can recover aggregate consumption as

$$\Delta \log C_{t+1} = \Delta c_{t+1} + \mu + \omega_\varepsilon_{t+1} \tag{2.42}$$

$$= \mu + \psi \left( (G - 1) \hat{k}_t - \frac{\theta}{\beta} \left( \frac{1 - \theta}{1 - (1 - \theta) \beta^{-1}} \right) \omega_\varepsilon_t + \theta \xi_{k,t} + \theta \varepsilon_{t+1} + \theta \xi_{k,t+1} \right) + \omega_\varepsilon_{t+1}.$$  

Similarly, we can recover all the other aggregate variables as follows

$$\Delta \log K_{t+1} = \Delta k_{t+1} + \mu + \omega_\varepsilon_{t+1}; \tag{2.43}$$

$$\Delta \log I_{t+1} = \Delta i_{t+1} + \mu + \omega_\varepsilon_{t+1}; \tag{2.44}$$

$$\Delta \log Y_{t+1} = \Delta y_{t+1} + \mu + \omega_\varepsilon_{t+1}. \tag{2.45}$$

For the calibration above we obtain the values $\psi = 0.5439$ and $\phi = -0.5109$, implying that $G = 0.9459$. The fact $G$ is close to 1 will have implications for the internal propagation in the model.


A standard tool for evaluating the empirical success of a business cycle model is to compare the predictions of a calibrated version of the model for a small set of unconditional second moments
to their empirical counterparts. Using the above expressions we simulate 1000 artificial samples of length 225, HP-filter the data, and compute these moments. In Table 2 we present statistics from the standard model without RI ($\kappa = \infty$) and the model with RI ($\kappa = 0.2$ nats). $\kappa = 0.2$ nats implies $\theta = 0.33$, so that about one-third of the new information is transmitted each period.

Output becomes less volatile, consumption and investment become more volatile (especially consumption, which nearly doubles), and hours remains the same. Consumption doubles in volatility because imperfectly observing the state of the world leads to a reduced ability to smooth consumption, similar to the results in Luo (2007). The comovements of the variables and output are basically unchanged, however, although the introduction of a second shock (the endogenous noise due to RI) orthogonal to the technology shock does tend to reduce the excessively high correlations in the model.\footnote{This result is not surprising given the discussion in Christiano and Eichenbaum (1992).}

The increase in consumption volatility may be misleading for this reason, so we conduct the following experiment. We assume that $\xi_{k,t} = 0 \forall t$; the random shock to the observations is always zero, but the planner uses the decision rules of an individual with RI. Relative to the case without RI, we observe an increase in consumption volatility and a large drop in investment volatility. The planner cannot smooth consumption as effectively even without the random changes in the observed state.

The reason that the changes in the implied comovements are not very large is that $k_t$ and $\hat{k}_t$ are highly correlated and have very similar volatilities, even when $\kappa$ is small. For example, when $\kappa = 0.2$ nats the contemporaneous correlation between these variables is 0.88 and their volatilities are 0.23 and 0.21 percent, respectively.\footnote{These are the moments of the unfiltered, normalized series.} Thus, even with very low channel capacities the resulting time series will be very similar in terms of their correlation patterns. Formally, we can compare the two time series using their coherence in the frequency domain, which examines their similarity across all frequencies. The system that determines the evolution of the states can be written as

$$X_t = (1 - \Psi L) \zeta_t$$
where

\[
\Psi = \begin{bmatrix}
\frac{1}{\beta} & G - \frac{1}{\beta} \\
\frac{\theta}{\beta} & \theta \left( G - \frac{1}{\beta} \right) + (1 - \theta) G
\end{bmatrix}
\]

\[
\zeta_t = \begin{bmatrix}
-\varepsilon_{t+1} \\
\theta \xi_{k,t+1} - \theta \varepsilon_{t+1}
\end{bmatrix}.
\]

The covariance matrix \( E_t [\zeta_t \zeta_t'] \) is then given by

\[
\Omega = \begin{bmatrix}
\omega^2 & \theta \omega^2 \\
\theta \omega^2 & \theta^2 \omega^2 + \theta^2 \text{var} (\xi_k)
\end{bmatrix}.
\]

Writing out the autocovariance of the vector MA(\( \infty \)) process yields

\[
\Gamma_s = \sum_{h=0}^{\infty} \Psi_{s+h} \Omega_{h} \Psi_{h}'
\]

for each \( s \), with typical element \( \gamma_{ij} (s) \). Then the spectral density at frequency \( \nu \) is

\[
S (\nu) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \Gamma_s \exp (-s\nu i) = \begin{bmatrix}
\gamma_{11} (\nu) & \gamma_{12} (\nu) \\
\gamma_{21} (\nu) & \gamma_{22} (\nu)
\end{bmatrix}
\]

with cross spectrum

\[
\gamma_{12} (\nu) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma_{12} (s) \exp (-s\nu i).
\]

The coherence of two series at frequency \( \nu \) is defined as

\[
K (\nu) = \frac{|\gamma_{12} (\nu)|}{\sqrt{\gamma_{11} (\nu) \gamma_{22} (\nu)}} \in [0, 1];
\]

coherece measures the linear comovement between \( X \) and \( Y \) at a given frequency. In Figure (1) we plot the coherence between \( k_t \) and \( \hat{k}_t \) across the frequency band \( \nu \in [0, \pi] \); the two series display strong positive coherence at almost every frequency, indicating that they are driven by movements at the same frequencies. The coherence remains strong even for very low values of \( \theta \).

\[21\]The eigenvalues of \( \Psi \) are \( \lambda_1 = G \in (0, 1) \) and \( \lambda_2 = \frac{1 - \theta}{\beta} \geq 0 \). Under the assumptions used previously to guarantee that the MA(\( \infty \)) process is stable, \( \Psi \) is also a stable matrix and therefore the infinite sum converges.

\[22\]There are two poles in the joint process for \( \{k_t, \hat{k}_t\} \), one located between frequencies \([0.2199, 0.2827]\) and one
We can use the consumption volatility number, which seems to be the most sensitive to $\kappa$, to obtain an estimate of the channel capacity. We choose $\kappa$ to match a consumption/output volatility ratio of 0.67, obtained from NIPA data over the period 1948-2005 using logged and HP-filtered data. Using this procedure we find an estimate of $\kappa = 0.42$ nats; then $\theta = 0.568$, implying that approximately 57% of the uncertainty is removed upon the receipt of a new signal. While this number may seem quite small, one must keep in mind that $\kappa$ measures the amount of information about the aggregate transmitted each quarter; presumably, agents in the real world would allocate much of their limited attention toward the large idiosyncratic movements in income observed in individual level data.

4. Impulse Responses

In this section we focus on the implications of RI for the shape of the impulse response functions for the model. The impulse response function is an important indicator of persistence. For example, if the variable of interest jumps up initially in response to a shock and then goes down immediately, then we know that this variable does not have any persistence at all. Here, we first examine the impulse responses of labor supply, output, capital stock, and consumption with response to a permanent technology shock in the presence of RI. We then examine the impact of a technology shock on output growth.

Before we examine the impulses of main macroeconomic variables, it is useful to examine the impulses of the true state ($k_t$) and the perceived state ($\hat{k}_t$), since they determine the dynamics of the main variables in the model. As Figure (2) shows, in the presence of RI ($\kappa = 0.2$ nats) $\hat{k}_t$ reacts to the shocks gradually and with delay (it displays an “inverse” hump shape), while $k_t$ jumps down initially and then transits back to its steady state level monotonically. The intuition is quite simple. Since the agent only has limited information-processing ability when analyzing the state, the perceived state takes more periods to start moving back to the steady state; the early periods are spent ’processing’ the fact that the technology state has changed.

We then plot the responses of capital, labor, output, and consumption with respect to the permanent technology shock for different degree of RI in Figure (3). Figure (3a) shows that capital takes more periods to converge to the steady state level for higher degrees of RI (lower values for $\kappa$). The solid and dashed lines in Figure (3b) show that labor supply reacts to the

---

$\text{between frequencies} \ [1.5394, 1.6022]$. These poles account for the apparent discontinuities in the plot; it is conventional to set the coherence at such points to 0.
innovations gradually and with delay under RI. In the absence of RI, the dashed-dotted line shows that labor supply jumps up initially with respect to the innovations and then goes back to the steady state immediately, clearly labor lacks a strong propagation mechanism in the standard full-information model. As a consequence, output also displays stronger persistence with RI since it is determined by both labor supply and capital stock. Figure (3d) shows that output also reacts to the innovations gradually under RI. Note that without RI, the figure also makes clear that output converges rapidly to its new steady state level. Its dynamics are not fundamentally different from the dynamics of productivity, in other words, the standard full-information case fails to display strong internal propagation mechanism. Figure (3c) shows that since the reaction of normalized consumption \((c_t)\) is smooth and delayed under RI, the recovered consumption \((\log(C_t))\) is more responsive and also displays an “inverse” hump shape.

Figure (4) illustrates the impulse responses of employment and output to both the exogenous technology shock \((\varepsilon_t)\) and the endogenous noise \((\xi_{k,t})\). From the two-equation dynamic system, (2.36) and (2.37), it is clear that since the exogenous technology shock \(\varepsilon_t\) appears in the dynamic system of \((k_t, \hat{k}_t)\) with a negative sign, the endogenous noise \(\xi_{k,t}\) due to finite capacity serves as a negative demand shock. Figure 3 clearly shows that \(\xi_{k,t}\) has the main features of an negative aggregate demand shock: it reduces employment and output temporarily but has no effect in the long-run. Our ”noise” shock is therefore quite similar to the ”news” shock in Lorenzoni (2006), providing an alternative theoretical foundation for demand shocks.

We now examine the response of output growth with respect to the technology shock. \(\Delta \log Y_{t+1}\), the change in the log of aggregate output, is given by:

\[
\Delta \log Y_{t+1} = \mu + \alpha \left( G\hat{k}_t - k_t - \frac{1}{\beta} \frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1 - (1-\theta)\beta^{-1}L} - \omega\varepsilon_{t+1} \right) + (1-\alpha) \phi \left( (G-1)\hat{k}_t - \frac{\theta}{\beta} \frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1 - (1-\theta)\beta^{-1}L} - \theta\omega\varepsilon_{t+1} + \theta\xi_{k,t+1} \right) + \omega\varepsilon_{t+1}. \tag{4.1}
\]

Combining terms we get

\[
\Delta \log Y_{t+1} = \mu + (\alpha G + (1-\alpha) \phi (G-1)) \hat{k}_t - \alpha k_t + \omega (1-\alpha) (1 + \phi\theta) \varepsilon_{t+1} + (1-\alpha) \phi\theta\xi_{k,t+1} - \frac{(1-\theta)\omega}{\beta} \left( \frac{\alpha + (1-\alpha) \phi\theta}{1 - (1-\theta)\beta^{-1}L} \right) \varepsilon_t + \theta \left( \frac{\alpha + (1-\alpha) \phi}{1 - (1-\theta)\beta^{-1}L} \right) \xi_{k,t}. \tag{4.2}
\]

For output growth to be stationary it must be the case that \(1-\theta < \beta\), which requires an adequately
large capacity channel (this point is also noted in Sims 2003). For the calibrated $\beta = 0.9908$, the cutoff value is $\kappa = -\frac{1}{2} \log (\beta) = 0.0047$ nats. As we approach the lower bound, we encounter increasingly-frequent violations of the nonnegativity constraint for gross investment, so we do not explore values too close to this limit.\footnote{As noted in Sims (2005), it is also the case that smaller values for $\kappa$ make the linear approximation less valid. The cutoff value turns out to be the same value as that needed to ensure that the variance of the noise is decreasing in channel capacity.}

We plot the calibrated impulse response function for the growth rate of the log of output to a one-standard-deviation increase in $\varepsilon_t$ in Figure (5) for $\kappa = 0.2$ and $\kappa = \infty$.\footnote{It is important that the reader recognizes the plot is of output growth, not the level of output.} With $\kappa = 0.2$ nats the growth rate of output follows the process

$$
\Delta \log Y_{t+1} = 0.004 + 0.3582 \hat{k}_t - 0.36k_t + 0.0039\varepsilon_{t+1} - 0.1078\xi_{k,t+1} - 0.0012 \sum_{j=0}^{\infty} (0.6767^j \varepsilon_{t-j}) + 0.0109 \sum_{j=0}^{\infty} (0.6767^j \xi_{k,t-j}).
$$

(4.3)

The effect of technology shocks on the growth rate of output die off quite slowly, even in the absence of a strong capital accumulation mechanism, due to the significant effects of the lagged $\varepsilon_t$ terms; it is also important to note that the persistence increase is not caused by the misperception of capital directly, as the coefficients on $k_t$ and $\hat{k}_t$ nearly cancel each other, but rather by the learning process.

In contrast, when $\kappa = \infty$ the output growth is given by

$$
\Delta \log Y_{t+1} = \mu + (\alpha (G - 1) + (1 - \alpha) \phi (G - 1)) k_t + \omega (1 - \alpha) (1 + \phi) \varepsilon_{t+1}
$$

(4.4)

$$
= 0.004 - 0.0018k_t + 0.0023\varepsilon_{t+1}.
$$

Clearly, this process will not display much persistence. Figure 5 shows the strong contrast between the two cases, where we plot the impulse response function of output growth to a one-standard-deviation shock to technology. In the case $\kappa = \infty$ there is essentially no effect, but introducing RI (reducing the value of $\kappa$) generates a small amount of persistence into the output growth function. As noted above, the effect is quantitatively small.

5. Autocorrelation Functions

Cogley and Nason (1995) show that output growth in US data has significant positive serial correlation for the first two periods after a permanent shock. Standard models cannot reproduce this
observation when hit by a technology that follows a random walk, such as the one we explored above – they predict output growth is essentially white noise. Modifications to permit the elastic response of capital utilization – following Greenwood, Hercowitz, and Huffman (1988) – do not alter this prediction, while introducing home production tends to make output growth negatively serially correlated. As discussed in Cogley and Nason (1995), the main discrepancy between the sample and model ACFs is the absence of positive dependence at lags 1 and 2.\textsuperscript{25} We use this section to explore the consequences of RI for the autocorrelation of output growth.

We restate the expression for the change in output derived earlier for convenience:

\[
\Delta \log Y_{t+1} = \mu + (\alpha G + (1-\alpha) \phi (G-1)) \hat{k}_t - \alpha k_t + \omega (1-\alpha) (1+\phi \theta) \varepsilon_{t+1} + (1-\alpha) \phi \theta \xi_{k,t+1} - \\
\frac{(1-\theta)}{\beta} \left( \frac{\alpha + (1-\alpha) \phi \theta}{1-(1-\theta) \beta^{-1} L} \right) \varepsilon_t + \theta \left( \frac{\alpha + (1-\alpha) \phi}{1-(1-\theta) \beta^{-1} L} \right) \xi_{k,t}. \tag{5.1}
\]

Note that in the case without RI ($\kappa = \infty$ which implies $\theta = 1$ and $\xi_{k,t} = 0 \forall t$), the above expression reduces to

\[
\Delta \log Y_{t+1} = \mu + (\alpha (G-1) + (1-\alpha) \phi (G-1)) k_t + \omega (1-\alpha) (1+\phi) \varepsilon_{t+1}
\]

and it implies that

\[
\text{covar} (\Delta \log Y_{t+j}, \Delta \log Y_t) \tag{5.2}
\]
\[
= E \left[ \left( (\alpha (G-1) + (1-\alpha) \phi (G-1)) k_{t+j-1} + \omega (1-\alpha) (1+\phi) \varepsilon_{t+j+1} \right) \times \right. \\
\left. \left( (\alpha (G-1) + (1-\alpha) \phi (G-1)) k_{t-1} + \omega (1-\alpha) (1+\phi) \varepsilon_t \right) \right] \tag{5.3}
\]
\[
= (\alpha (G-1) + (1-\alpha) \phi (G-1))^2 \text{covar} (k_{t+j-1}, k_{t-1}) + \\
(\alpha (G-1) + (1-\alpha) \phi (G-1)) \omega (1-\alpha) (1+\phi) \text{covar} (k_{t+j-1}, \varepsilon_t).
\]

The persistence in output growth depends directly on the strength of the capital accumulation channel. For the calibration we consider this expression becomes

\[
\text{covar} (\Delta \log Y_{t+j}, \Delta \log Y_t) = (3.2 \times 10^{-6}) \text{covar} (k_{t+j-1}, k_{t-1}) - (4.1 \times 10^{-6}) \text{covar} (k_{t+j-1}, \varepsilon_t);
\]

\textsuperscript{25}These two values are around 0.4 and 0.2 in the US data.
thus, output growth will be white noise in the absence of rational inattention unless capital accumulation is very strong, and previous results have shown the weakness of this mechanism.

In the model with RI, the model-generated autocorrelation function is more complicated due to the presence of the distributed lags of past shocks. Of course, these distributed lags are what induce additional persistence in output growth – any innovation to \( \varepsilon \) will contribute for many periods – in addition to the capital accumulation channel. Figure (6) presents the autocorrelation functions for the two cases, based on the ensemble average of the same 1000 simulations used to compute the business cycle statistics. It is clear that RI does increase the autocorrelation of output growth, but the effect is small even for a relatively low value of \( \kappa \) – we cannot reproduce the observed values of 0.4 and 0.2. Our chosen value of \( \kappa = 0.2 \) nats is the one that appears to produce the maximum amount of autocorrelation, as lower values reduce rather than increase the persistence of output growth.

Our results for the autocorrelation of output growth are robust to alternative calibration targets. For example, we chose a relatively-conservative target for \( \frac{K}{Y} \). Cooley and Prescott (1995), using a careful allocation of the data relative to the model, suggest a value of 13.2. When we use this value, we obtain only minor increases in the autocorrelation function. To get any significant improvement we would need a very high target value, much higher than any reasonable calibration could obtain. \(^{26}\) Changes in \( \frac{C}{Y} \) are nonmonotone, and our chosen calibration target (which is very close to the value advocated by Cooley and Prescott 1995) leads to a relatively high autocorrelation. Choosing alternative values for steady-state hours (such as 0.2, which would be the value if sleeping and personal care were included in leisure) or the share of capital income \( \alpha \) have little effect.

6. Forecastable Movements

Rotemberg and Woodford (1996) pointed out that forecastable movements in US output have a variance around 100 times larger than those predicted by a standard RBC model. For \( j \geq 1 \), let \( \Delta y^j_t \) denote the difference between \( y_{t+j} \) and \( y_t \):

\[
\Delta y^j_t = \log \left( Y_{t+j} \right) - \log \left( Y_t \right) = \sum_{i=1}^{j} \Delta \log \left( Y_{t+i} \right)
\]  

\(^{26}\)An alternative approach is to calibrate the model to match interest rates rather than capital/output ratios. If we calibrate the model to produce a quarterly return on capital of \( R = 0.016 \) (an annual return of 6.5 percent) then our implied capital/output ratio would be only 8.5, moving us in the wrong direction.
The forecastable change in output between $t$ and $t + j$ is then given by

$$E_t \left( \Delta y_t^j \right) = E_t \left( \sum_{i=1}^{j} \Delta \log (Y_{t+i}) \right).$$  \hspace{1cm} (6.2)

In the model with RI these expectations must be consistent with the information state $\hat{k}_t$. Expected output at time $t$, conditional on $\hat{k}_t$, is given by

$$E_t (y_t) = (\alpha + (1 - \alpha) \phi) \hat{k}_t;$$

since in general $\hat{k}_t \neq k_t$, expected output will not equal realized output. $\hat{k}_t$ is expected to evolve according to

$$E_t (\hat{k}_{t+1}) = \frac{1 - \delta}{\exp(\mu)} \hat{k}_t + \exp(i) \frac{\alpha + (1 - \alpha) \phi - \psi}{\exp(\mu + k)} \hat{k}_t,$$

or

$$E_t (\hat{k}_{t+j}) = \left[ \frac{1 - \delta}{\exp(\mu)} + \exp(i) \frac{\alpha + (1 - \alpha) \phi - \psi}{\exp(\mu + k)} \right]^j \hat{k}_t,$$

where $i$ is the steady state value of the log of gross investment. Thus, we have

$$E_t (y_{t+j}) = (\alpha + (1 - \alpha) \phi) \left[ \frac{1 - \delta}{\exp(\mu)} + \exp(i) \frac{\alpha + (1 - \alpha) \phi - \psi}{\exp(\mu + k)} \right]^j \hat{k}_t$$

$$= (\alpha + (1 - \alpha) \phi) B^j \hat{k}_t.$$

We can therefore obtain the expected change in output as

$$E_t \left( \Delta y_t^j \right) = (\alpha + (1 - \alpha) \phi) (B^j - B) \hat{k}_t;$$ \hspace{1cm} (6.3)

the standard deviation of the forecastable part of output growth is therefore

$$\sigma \left( E_t \Delta y_t^j \right) = (\alpha + (1 - \alpha) \phi) (B^j - B) \sigma (\hat{k}_t).$$ \hspace{1cm} (6.4)

Table 3 shows that RI tends to raise the volatility of the forecastable component of output growth, but not by very much. Even when $\kappa = 0.01$ nats we cannot produce a quantitatively significant increase over the $\kappa = \infty$ case.

The expression above shows clearly why there is little improvement – the only difference is that (6.4) depends on the standard deviation of $\hat{k}_t$ rather than $k_t$, as the coefficients do not depend on
We have previously noted that these two variables are very similar in their dynamic behavior even for small values of $\kappa$.

7. Conclusion

In this paper we have reconsidered some puzzles in the RBC literature by allowing for agents to suffer from Rational Inattention. It is reasonable to interpret our results in the following way: RI could be a component of a macroeconomic model that fits the data, but it cannot be the only modification. While it is difficult to calibrate the parameter which controls the capacity of the Shannon channel because the model fails to demonstrate sharp differences, one should certainly agree that processing only 0.2866 bits of macroeconomic information every quarter would have to be closer to a lower bound than an upper one (this corresponds to $\kappa = 0.2$ nats, the parametrization which produced the maximum autocorrelation in output growth). And our estimate using consumption volatility is not much larger ($\kappa = 0.42$ nats, corresponding to 0.6059 bits). Thus, RI seems to play only a minor role in resolving the extant puzzles.

There are some caveats to our results that we feel suggest a larger role for RI. First, we have derived our results under the assumption of certainty equivalence. This assumption is difficult to justify, as it rules out the precautionary behavior that seems pervasive in the microeconomic consumption literature. Sims (2005, 2006), Lewis (2006), Batchuluun, Luo, and Young (2007), and Tutino (2007) make some progress toward solving the fully-nonlinear problem but are restricted to very simple models for the following two reasons. The first obstacle has already been mentioned above – the infinite-dimensional state space. A second problem is the unknown form of the posterior distribution. We speculate that if we parameterize this object flexibly enough we can apply a projection algorithm to compute the laws of motion, although the discreteness apparent in such distributions is a troubling feature. The second obstacle is that something similar to an MCMC approach will be needed to construct the posterior, which would make the computation extremely slow. Luo and Young (2008) extend the basic RI linear-quadratic model to include risk-sensitivity, a feature which breaks certainty equivalence without losing the tractability of the linear-quadratic-Gaussian environment.

---

27 The recent introduction into economics of high-dimensional approximation methods may be quite helpful here, although the smoothness typically required could be a problem in an RI context where the distributions often hit the zero lower bound. Judd (2007) contains a summary of methods designed to break the curse of dimensionality.

28 That paper also demonstrates a wide range of observational equivalence results for the linear-quadratic setup (rational expectations, rational inattention, risk-sensitivity, and robust control are not distinguishable using
Second, we have assumed only one state variable. Lindé (2005) shows that the growth model performs better in the presence of growth-rate shocks with small autocorrelation, a modification that would increase the size of the state space to two variables. As we noted in the main body of the paper, solving RI models with multiple state variables is difficult at best. The 'no subsidization' constraint

$$\Psi_t \succeq \Sigma_{t+1}$$

is the main culprit, where this inequality is read 'the element-by-element difference of the matrices $\Psi_t$ and $\Sigma_{t+1}$ is a positive semi-definite matrix.' Positive semi-definiteness is a constraint on the eigenvalues of the matrix $\Psi_t - \Sigma_{t+1}$ and is not linear when the dimension is larger than 1, meaning that the optimal distribution of the state is not Gaussian; it is also likely that the pattern of binding and nonbinding restrictions embedded in this constraint will be complicated, since some variables may get no attention at all in some periods. Given previous results on the poor behavior of the fully-nonlinear model, proceeding along this direction seems difficult.

consumption-savings decisions), showing that the empirical verification of RI will be a difficult task.
References


8. Appendix

8.1. Appendix A. Decentralizing the Planning Problem

In this appendix we make some comments on the problem of decentralizing an economy with rational inattention. The competitive equilibrium version of our model features households who solve the sequential problem

$$\max_{\{c_t, l_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t)$$

subject to

$$c_t + i_t \leq r_t k_t + w_t l_t$$

$$k_{t+1} \leq (1 - \delta) k_t + i_t$$

$$\log (\Psi_t^e) - \log (\Sigma_{t+1}^e) \leq 2\kappa$$

$$\Psi_t^e \leq \Sigma_{t+1}^e.$$ 

The superscript 'e' denotes the equilibrium problem. The firm’s problem supplies expressions for the prices:

$$r_t = \alpha \exp(z_t) K_t^{1-\alpha} L_t^{\alpha-1}$$

$$w_t = (1 - \alpha) \exp(z_t) K_t^\alpha L_t^{-\alpha}.$$ 

Assuming that $z_t$ follows a random walk with drift, the two state variables for the household problem are $\frac{k_t}{\exp(z_t)}$ and $\frac{K_t}{\exp(z_t)}$. The household faces the problem of allocating attention between observing individual wealth and observing the aggregate capital stock; the planner apparently needs only to observe the aggregate capital stock. However, both the planner and the individual households actually need to observe the entire distribution of individual capital stocks $\{k_i^t\}$, where $\sum_i k_i^t = K_t$ defines the aggregate; in this model it may not be innocuous to assume that the distribution is completely summarized by the aggregate. If we assume that these are the state variables for the economy, the problems of the households and the planner become symmetric, implying that a decentralization must exist by the Second Welfare Theorem (all RI does is put constraints on expectations, which can be subsumed into the utility function) provided the objective function remains a concave programming problem. However, deriving the outcomes of a model with RI and multiple state variables is difficult, as we note in the main body of the paper, so actually computing
the decentralization is nontrivial.

Sims (2005) makes a related point regarding the nature of competitive equilibria with rational inattention. He is concerned with trying to understand just how an economy would allocate goods in the presence of agents with limited capacity, noting that it would involve theorizing at the market microstructure level and incorporating details regarding inventories, retailers, and other market microstructure arrangements. Similar concerns in the alternative implementation of rational inattention arise in Mankiw and Reis (2007), where price-setting firms would attempt to exploit rationally-inattentive agents; restrictions on exactly which agents are inattentive are needed to study competitive equilibria in that environment.

8.2. Appendix B. Deriving the variance-covariance matrix of \((k, \hat{k})\)

Taking unconditional variances on both sides of (2.36) and (2.37), we find that

\[
\begin{bmatrix}
1 & 0 \\
-\theta & 1
\end{bmatrix}
\Sigma_k
\begin{bmatrix}
1 & -\theta \\
0 & 1
\end{bmatrix}
= 
\begin{bmatrix}
\frac{1}{\beta} & G - \frac{1}{\beta} \\
0 & (1-\theta)G
\end{bmatrix}
\Sigma_k
\begin{bmatrix}
\frac{1}{\beta} & 0 \\
G - \frac{1}{\beta} & (1-\theta)G
\end{bmatrix}
+ 
\begin{bmatrix}
\omega^2 & 0 \\
0 & \theta^2 \text{var}(\xi_k)
\end{bmatrix}
\]

where

\[
\Sigma_k = 
\begin{bmatrix}
\text{var}(k) & \text{covar}(k, \hat{k}) \\
\text{covar}(k, \hat{k}) & \text{var}(\hat{k})
\end{bmatrix}
\]

This expression is a standard discrete Lyapounov equation. For the case with \(\kappa = \infty\) we use the fact that \(k_t = \hat{k}_t \forall t\), \(\xi_{k,t} = 0 \forall t\), and \(\theta = 1\) to obtain

\[
\text{var}(k) = \frac{\omega^2}{1 - G^2}. \quad (8.1)
\]

The solution to this equation when \(\kappa < \infty\) is given by

\[
\begin{bmatrix}
\text{var}(k) \\
\text{covar}(k, \hat{k}) \\
\text{var}(\hat{k})
\end{bmatrix}
= 
\begin{bmatrix}
1 - \beta^{-2} & -2\beta^{-1}(G - \beta^{-1}) & -\beta^{-2}(G\beta - 1)^2 \\
-\theta - G\beta^{-1} & 1 - (1-\theta)G(G - \beta^{-1}) & 0 \\
\theta^2 - (1-\theta)^2G^2 & -2\theta & 1
\end{bmatrix}^{-1}
\begin{bmatrix}
\omega^2 \\
0 \\
\theta^2 \text{var}(\xi_k)
\end{bmatrix}
\]

We can derive the solutions in closed-form, although they are not particularly intuitive.
8.3. Appendix C. A Special Case with Quadratic Utility and Additive Shocks

In the main text, we solve a linear approximation to a stochastic growth model with RI and permanent technology shocks numerically and discuss the implications of RI for the dynamics of employment, consumption, capital stock, and output. In this appendix we provide an analytical solution to an explicitly linear-quadratic economy to show that the linear approximation in the main text captures the important elements of RI in a linear environment. A more complete presentation of a linear-quadratic RI model can be found in Luo and Young (2007).

To derive the closed-form solution, we first state the social planning problem as

\[
\max_{\{c_t, l_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[ c_t - \frac{1}{2} c_t^2 - \eta \left( l_t - \frac{1}{2} \vartheta l_t^2 \right) \right]
\]  

subject to

\[
k_{t+1} = k_t - c_t + A_1 k_t + A_2 l_t + z_t
\]  

\[
z_{t+1} = \rho z_t + \varepsilon_{t+1}
\]

where the time-separable utility is quadratic in terms of consumption and labor supply and output is linear in capital \((k_t)\) and labor \((l_t)\). There is an additive technology disturbance, \(z_t\), which follows an AR(1) process with iid shocks \(\varepsilon_t\) \((E(\varepsilon) = 0 \text{ and } \text{var}(\varepsilon) = \omega^2)\). For simplicity we assume that there is no depreciation, although any linear depreciation schedule could be folded into \(A_1\).

The efficiency conditions for this problem are

\[
1 - c_t = \lambda_t
\]  

\[
\eta (1 - \vartheta l_t) = A_2 \lambda_t
\]  

\[
\lambda_t = \beta (1 + A_1) E_t [\lambda_{t+1}].
\]

Combining (8.5) and (8.6) gives the intratemporal optimal condition

\[
l_t = \frac{1}{\vartheta} - \frac{A_2}{\vartheta \eta} + \frac{A_2}{\vartheta \eta} c_t.
\]

For simplicity, we assume that \(\beta (1 + A_1) = 1\) and thus both consumption and labor supply are
martingales:

\[ c_t = E_t [c_{t+1}] \quad \text{and} \quad l_t = E_t [l_{t+1}]. \tag{8.9} \]

Forward iteration of (8.8) with the No-Ponzi condition results in the intertemporal resource constraint

\[
(1 + A_1) k_t = \sum_{i=0}^{\infty} (1 + A_1)^{-i} E_t [c_{t+i}] - A_2 \sum_{i=0}^{\infty} (1 + A_1)^{-i} E_t [l_{t+i}] - \sum_{i=0}^{\infty} (1 + A_1)^{-i} E_t [z_{t+i}] . \tag{8.10}
\]

Substituting (8.4) into (8.10) yields

\[
c_t - A_2 l_t = A_1 k_t + \left( \frac{A_1}{1 + A_1 - \rho} \right) z_t .
\]

Combining this equation with the intratemporal optimal condition (8.8), we obtain

\[
c_t - A_2 \left( \frac{1}{\vartheta} - \frac{A_2}{\partial \eta} + \frac{A_2}{\partial \eta} c_t \right) = A_1 k_t + \left( \frac{A_1}{1 + A_1 - \rho} \right) z_t ,
\]

which implies that the optimal consumption function and labor supply are

\[
c_t = \frac{\partial \eta A_1}{\vartheta - A_2^2} \left( k_t + \left( \frac{1}{1 + A_1 - \rho} \right) z_t \right) + \frac{A_2 (\eta - A_2)}{\partial \eta - A_2^2} \tag{8.11}
\]

\[
l_t = \frac{A_1 A_2}{\partial \eta - A_2^2} \left( k_t + \left( \frac{1}{1 + A_1 - \rho} \right) z_t \right) + \frac{\partial \eta (\eta - A_2)}{\vartheta - A_2^2} \tag{8.12}
\]

Denote \( a_t = k_t + \left( \frac{1}{1 + A_1 - \rho} \right) z_t \) as the new unique state variable (a measure of permanent income). We can now write the original resource constraint as

\[
a_{t+1} = (1 + A_1) a_t - c_t + A_2 l_t + \left( \frac{1}{1 + A_1 - \rho} \right) \varepsilon_{t+1} . \tag{8.13}
\]

In the absence of RI, the evolution of the new state variable can be written as

\[
\Delta a_{t+1} = \left( \frac{1}{1 + A_1 - \rho} \right) \varepsilon_{t+1} , \tag{8.14}
\]

which implies that \( a_t \) is a random walk. As a result, consumption growth, labor supply growth, and output growth are all white noise. Hence, this simplified model generates similar results as our benchmark model in the main text, in which all variables are approximately random walks.

In the presence of RI, the agent cannot observe the state perfectly due to information processing.
constraint. Consequently, in this Linear Quadratic Gaussian framework, the optimal distribution of the state, \( a_t \), follows a normal distribution:

\[
a_t \sim N \left( \hat{a}_t, \sigma^2_{a,t} \right),
\]

where \( \hat{a}_t \) is the conditional expectation of the true state and \( \sigma^2_{a,t} = \text{var}_t (a_t) \) is the conditional variance of the state. The information-processing constraint,

\[
\log (\text{var}_t (a_{t+1})) - \log (\text{var}_{t+1} (a_{t+1})) = 2\kappa,
\]

implies that the steady state for the conditional variance \( \sigma^2_{a,t} \) is

\[
\sigma^2_a = \frac{\omega^2 / (1 + A_1 - \rho)}{\exp (2\kappa) - \beta^2},
\]

and the agent is assumed to observe a Gaussian noisy signal \( a^*_{t+1} \),

\[
a^*_{t+1} = a_{t+1} + \xi_{a,t+1},
\]

where \( a_{t+1} \) is the true signal and \( \xi_{a,t+1} \) is the endogenous noise due to RI. Hence, the evolution of the information state, \( \hat{a}_t \), follows a recursive Kalman filter equation

\[
\hat{a}_{t+1} = (1 - \theta) \hat{a}_t + \theta (a_{t+1} + \xi_{a,t+1})
\]

where \( \theta = 1 - 1 / \exp (2\kappa) \) is the optimal weight on a new observation. Since certainty equivalence holds in this case, we can just replace the true state in the decision rules with the information state and then obtain the optimal consumption and labor supply function under RI as

\[
c_t = \frac{\partial \eta A_1}{\partial \eta - A_2^2} \hat{a}_t + \frac{A_2 (\eta - A_2)}{\partial \eta - A_2^2},
\]

\[
l_t = \frac{A_1 A_2}{\partial \eta - A_2^2} \hat{a}_t + \frac{\partial \eta (\eta - A_2)}{\partial \eta - A_2^2},
\]

and output can be written as

\[
y_t = \frac{1 - \rho}{1 + A_1 - \rho} z_t + A_1 a_t + \frac{A_1 A_2}{\partial \eta - A_2^2} \hat{a}_t.
\]
Furthermore, we can write the evolution of the true state as

\[ a_{t+1} = (1 + A_1) a_t - A_1 \hat{a}_t + \left( \frac{1}{1 + A_1 - \rho} \right) \varepsilon_{t+1} \]  

(8.20)

where we omit the constant terms since they are irrelevant. Hence, the model economy can be characterized by two dynamic equations, (8.20) and (8.16). Combining these two equations we find that the difference between the true state and the information state follows MA(\infty) process in terms both fundamental shocks and endogenous noise.\(^{29}\)

\[ a_t - \hat{a}_t = \frac{1}{1 - (1 - \theta)(1 + A_1)} L \left( \left( \frac{1 - \theta}{1 + A_1 - \rho} \right) \varepsilon_t - \left( \frac{\theta}{1 + A_1 - \rho} \right) \xi_{a,t} \right) \]  

(8.21)

and thus we can rewrite the evolution of \( a_t \) and \( \hat{a}_t \) as

\[ \Delta a_{t+1} = \frac{A_1}{1 - (1 - \theta)(1 + A_1)} L \left( \left( \frac{1 - \theta}{1 + A_1 - \rho} \right) \varepsilon_t - \left( \frac{\theta}{1 + A_1 - \rho} \right) \xi_{a,t} \right) + \left( \frac{1}{1 + A_1 - \rho} \right) \varepsilon_{t+1} \]  

(8.22)

\[ \Delta \hat{a}_{t+1} = \frac{\theta}{1 + A_1 - \rho} \left( \frac{\varepsilon_{t+1}}{1 - (1 - \theta)(1 + A_1) L} \right) + \theta \left( \xi_{a,t+1} - \frac{\theta (1 + A_1)}{1 + A_1 - \rho} \frac{\xi_{a,t}}{1 - (1 - \theta)(1 + A_1) L} \right) \]  

(8.23)

(8.22) means that in the presence of RI, the true state \( a_t \) is no longer a random walk. Instead, the changes in \( a_t \) are predictable by past technology shocks as well as past noise shocks.\(^{30}\) (8.23) implies that RI can affect the impact of the exogenous shocks on consumption, labor supply, and output by a factor, \( \frac{\theta}{1 - (1 - \theta)(1 + A_1) L} \). In other words, this factor provides an additional propagation mechanism so that the technology shocks have persistent effects on the changes in consumption, labor supply, and output. Thus, with AR(1) technology shocks, the simple model has a potential to generate the positive serial correlation in both output growth and consumption growth found in the US data. As noted in the main body of the paper, plausible calibrations imply only weak effects, though.

\(^{29}\) Of course, here we also impose a condition that \((1 - \theta)(1 + A_1) < 1\).

\(^{30}\) Without RI (\( \kappa = \infty \) which implies \( \theta = 1 \)) (8.22) reduces to (8.14).
### Table 1
Calibrated Parameter Values

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\alpha$</th>
<th>$\mu$</th>
<th>$\eta$</th>
<th>$\beta$</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.018</td>
<td>0.36</td>
<td>0.004</td>
<td>2.84</td>
<td>0.99</td>
<td>0.00732</td>
</tr>
<tr>
<td>Variable</td>
<td>Std Dev.</td>
<td>Cross-Correlations</td>
<td>Std Dev.</td>
<td>Cross-Correlations</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>$y_t$</td>
<td>(0.90, 0.81)</td>
<td>(0.46, 0.52)</td>
<td>(0.71, 0.75)</td>
<td>(1.00, 1.00)</td>
<td>(0.71, 0.75)</td>
</tr>
<tr>
<td>$c_t$</td>
<td>(0.44, 0.75)</td>
<td>(0.55, 0.15)</td>
<td>(0.75, 0.34)</td>
<td>(0.98, 0.62)</td>
<td>(0.65, 0.53)</td>
</tr>
<tr>
<td>$i_t$</td>
<td>(2.32, 2.56)</td>
<td>(0.40, 0.53)</td>
<td>(0.66, 0.66)</td>
<td>(0.99, 0.72)</td>
<td>(0.72, 0.49)</td>
</tr>
<tr>
<td>$l_t$</td>
<td>(0.48, 0.49)</td>
<td>(0.35, 0.52)</td>
<td>(0.63, 0.68)</td>
<td>(0.98, 0.81)</td>
<td>(0.73, 0.58)</td>
</tr>
</tbody>
</table>

Business Cycle Statistics ($\kappa = 0.2, \xi_{k,t} = 0 \forall t$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Std Dev.</th>
<th>Cross-Correlations</th>
<th>Std Dev.</th>
<th>Cross-Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t$</td>
<td>0.78</td>
<td>0.55</td>
<td>0.77</td>
<td>1.00</td>
</tr>
<tr>
<td>$c_t$</td>
<td>0.64</td>
<td>0.20</td>
<td>0.48</td>
<td>0.91</td>
</tr>
<tr>
<td>$i_t$</td>
<td>1.56</td>
<td>0.85</td>
<td>0.94</td>
<td>0.86</td>
</tr>
<tr>
<td>$l_t$</td>
<td>0.33</td>
<td>0.75</td>
<td>0.91</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Table 3

Predicted Standard Deviations of $\Delta y^j_t$

<table>
<thead>
<tr>
<th>Variable</th>
<th>$j = 1$</th>
<th>$j = 2$</th>
<th>$j = 4$</th>
<th>$j = 32$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa = 0.01$</td>
<td>0.0003</td>
<td>0.0006</td>
<td>0.0011</td>
<td>0.0047</td>
</tr>
<tr>
<td>$\kappa = 0.2$</td>
<td>0.0003</td>
<td>0.0005</td>
<td>0.0010</td>
<td>0.0045</td>
</tr>
</tbody>
</table>
Figure 1: Coherence between $k$ and $\tilde{k}$
Figure 2: Impulse Response to Technology Shock

IRFs: Normalized Actual and Perceived Capital ($\kappa=0.2$)
Figure 3: Impulse Response to Technology Shock
Figure 4: Impulse Response to Noise Shock

IRF of log(L)

IRF of log(Y)

% dev. from steady state

period

% dev. from steady state

period

IRF to $\epsilon$

IRF to $\xi$

38
Figure 5: Impulse Response of Output Growth

Impulse Response Functions for Output Growth

\[ \kappa = 0.2 \]

\[ \kappa = \infty \]
Figure 6: Autocorrelation of Output Growth

Autocorrelation Functions for Output Growth

- Lead
- Autocorrelation Functions
- $\kappa = 0.2$
- $\kappa = \infty$