The B.E. Journal of Macroeconomics

Contributions

Volume 9, *Issue* 1 2009 *Article* 14

Rational Inattention and Aggregate Fluctuations

Yulei Luo* Eric R. Young†

*University of Hong Kong, ylou@econ.hku.hk

†University of Virginia, ey2d@virginia.edu

Recommended Citation

Yulei Luo and Eric R. Young (2009) "Rational Inattention and Aggregate Fluctuations," *The B.E. Journal of Macroeconomics*: Vol. 9: Iss. 1 (Contributions), Article 14.

Available at: http://www.bepress.com/bejm/vol9/iss1/art14

Rational Inattention and Aggregate Fluctuations*

Yulei Luo and Eric R. Young

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.

KEYWORDS: rational inattention, consumption volatility, propagation mechanism

*We would like to thank the helpful comments of Chris Sims and attendees at the 2006 Winter Econometric Society, especially our discussant Mathias Paustian. We also thank the comments of the editor Chris House and two anonymous referees and Xuan Tam for copyediting assistance above and beyond the call of duty. Luo thanks the Competitive Earmarked Research Grant (CERG) in Hong Kong for financial support. Young thanks the Bankard Fund for Political Economy at Virginia for financial support and LAEF at Santa Barbara for hospitality. All errors are our responsibility.

1 Introduction

It is well-known that the standard RBC models display a weak internal propagation mechanism: impulse responses for aggregate output and labor supply almost perfectly 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 is characterized by hump-shaped responses to transitory shocks – 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] In fact, the models generate zero persistence in output growth if the technology shock is assumed to follow a random walk. 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 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. Specifically, 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

[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.

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.[2] 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 improvement to be 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 shape 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). Finally, we examine the effects of idiosyncratic shocks on the propagation mechanism of an RI version of a linear–quadratic (LQ) RBC model and show that the presence of the idiosyncratic shock plays a role in strengthening the inertial responses to the aggregate productivity shock if individuals cannot distinguish the aggregate and idiosyncratic components in the aggregate productivity process.

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

[2] We 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.

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.[3]

We then alter the basic model to incorporate both aggregate and idiosyncratic shocks. With rationally-inattentive agents, idiosyncratic shocks are likely to dominate limited attention due to their relatively large variance. Here, we assume an explicit linear-quadratic environment for two reasons: it is easier to aggregate and prices become exogenous, leading the competitive and social planning problems to coincide. We assume that aggregate shocks are random walks while idiosyncratic shocks are white noise (as in Pischke 1995); furthermore, aggregate and idiosyncratic shocks cannot be separately observed. As the size of the idiosyncratic shocks gets big relative to the aggregate shocks, impulse responses of output to technology become more persistent and larger. Thus, that economy looks quite similar to an RI economy – when we combine the two effects together, we can get significant persistence in output growth. Unfortunately, the implication that prices are exogenous means that the general equilibrium effects that weaken persistence – falling returns – are absent.

Many propagation mechanisms have been introduced in business cycle models to improve their performance. Cogley and Nason (1995) and Burnside, Eichenbaum, and Rebelo (1996) consider the effects of various forms of adjustment costs in allocating labor, Perli and Sakellaris (1998) examines the propagation mechanism in a model with human capital, Schmitt-Grohé (2000) considers the impact of indeterminacy, Lettau and Uhlig (2000) and Boldrin, Christiano, and Fisher (2001) investigate the impact of habit formation, and Kasa (1995) and Saito (2005) explore the propagation power of signal extraction problems.

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 (2008) examines consumption dynamics under information process constraints in the permanent income hypothesis model, while Maćkowiak and Wiederholt (2008) explore the implications for optimal sticky prices. Sims (2005, 2006)

[3]Maćkowiak and Wiederholt (2008) formally examine how attention allocation affects firms' pricing decisions and show that firms allocate less attention in processing aggregate information.

investigates consumption and saving decision under more realistic preferences, but he is only able to explore a two-period model.[4] Van Nieuwerburgh and Veldkamp (2008) examine the effects of information processing constraints on diversification. This paper contributes to both literatures.

The remainder of the paper is organized as follows. Section 2 presents a stochastic growth model with rational inattention. Section 3 compare the model's predictions with the empirical evidence. Section 4 examines the effects of idiosyncratic shocks on the responses to the aggregate shock in a linear-quadratic version of the stochastic growth model. Section 5 concludes.

2 A Stochastic Growth Model with Rational Inattention

This section lays out the optimizing problem of a social planner facing information processing constraints. 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.[5] 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 decentralized economy. We examine a linear-quadratic setting with explicit idiosyncratic shocks in a later section of this paper.

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). 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

[4]Further extensions of the nonlinear model can be found in Lewis (2006), Tutino (2007), and Batchuluun, Luo, and Young (2007).

[5]We discuss the problems associated with the competitive equilibrium relative to the planning solution in Appendix A.1.

$$\max_{\{C_t, L_t\}} E_0 \sum_{t=0}^{\infty} \left[\beta^t \frac{C_t^{1-\gamma}}{1-\gamma} V\left(1 - L_t\right) \right] \tag{1}$$

subject to

$$C_t + I_t \leq Y_t \tag{2}$$
$$Y_t \leq K_t^{\alpha} (Z_t L_t)^{1-\alpha} \tag{3}$$
$$K_{t+1} \leq I_t + (1-\delta) K_t \tag{4}$$
$$\log Z_{t+1} = \mu_z + \log Z_t + \omega \varepsilon_{t+1} \tag{5}$$

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 μ_z 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 $\alpha, \beta \in (0,1)$, $\delta \in [0,1]$, and $\mu_z, \omega, \gamma > 0$. Combining equations (2), (3), and (4) yields the law of motion for capital,

$$K_{t+1} = K_t^{\alpha} (Z_t L_t)^{1-\alpha} - C_t + (1-\delta) K_t, \tag{6}$$

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.

Following Hansen (1985), we also assume indivisible labor:

$$V\left(1 - L_t\right) = \eta \left(1 - L_t\right).$$

Although the output process is nonstationary, the following ratios are stationary:

$$\widetilde{K}_t = \frac{K_t}{Z_t}, \widetilde{Y}_t = \frac{Y_t}{Z_t}, \widetilde{C}_t = \frac{C_t}{Z_t}, \widetilde{I}_t = \frac{I_t}{Z_t};$$

under the restrictions for preferences given above L_t is already stationary. Now we can normalize the evolution equation of capital (6) to obtain

$$\widetilde{K}_{t+1} = \exp\left(-\mu_z - \omega \varepsilon_{t+1}\right) \left(\widetilde{K}_t^{\alpha} {L_t}^{1-\alpha} - \widetilde{C}_t + (1-\delta) \widetilde{K}_t \right) \tag{7}$$

Similarly, the objective function can also be normalized in terms of $\widetilde{C}_t$ and the effective discount factor becomes $\beta \exp(-\mu_z - \omega \varepsilon_{t+1})(1-\gamma)$. The optimal decision rules take the form

$$c_t = \psi k_t \tag{8}$$
$$l_t = \phi k_t \tag{9}$$

for some coefficients (ψ, ϕ).[6] We use lowercase letters to denote deviations from the steady state of any stationary variable.

Our model with RI follows Sims (2003). 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. The sequential problem for the planner is

$$\widehat{v}\left(\widehat{k}_t\right) = \max_{\{c_t, l_t, \mathcal{D}_t\}_0^\infty} E_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t, l_t\right)\right] \tag{10}$$

subject to

$$k_{t+1} = \frac{\widetilde{K}^\alpha L^{1-\alpha}}{\widetilde{K}\exp\left(\mu_z\right)}\left(\alpha k_t + (1-\alpha) l_t\right) - \frac{\widetilde{C}}{\widetilde{K}\exp\left(\mu_z\right)} c_t + \frac{(1-\delta)\widetilde{K}}{\widetilde{K}\exp\left(\mu_z\right)} k_t - \omega\varepsilon_{t+1} \tag{11}$$

$$k_{t+1}|\mathcal{I}_{t+1} \sim \mathcal{D}_{t+1} \tag{12}$$

$$k_t|\mathcal{I}_t \sim \mathcal{D}_t, \tag{13}$$

given $k_0|\,\mathcal{I}_0 \sim N\left(\widehat{k}_0, \Sigma_0\right)$, and the requirement that the rate of information flow at $t+1$ implicit in the specification of the distributions, $\mathcal{D}_t$ and $\mathcal{D}_{t+1}$, be less than channel capacity. Here $\mathcal{D}_t$ is the posterior distribution of the underlying state variable (k_t), $\widehat{k}_t = E\left(k_t|\,\mathcal{I}_t\right)$ is the perceived state, and $\mathcal{I}_t$ is the information available at time t. (12) means that in the presence of RI, the social planner in the economy cannot observe the underlying state *perfectly*. As shown in Sims (2003, 2005), given the linear constraint and the (approximate) quadratic utility, $\mathcal{D}_t$ is normal, that is, Gaussian posterior uncertainty is optimal. (See the next paragraph for a detailed explanation.) The expectation is formed under the assumption that $\{c_t, l_t\}_0^\infty$ are chosen under the information processing constraints. 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.

Because observing the state under rational inattention involves information transfer at a limited channel capacity, 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 $\widetilde{k}_t$ in this model), but

[6] See Appendix A.2 for detailed derivations.

the so-called *information state*: the distribution of the underlying state k_t conditional on $\mathcal{I}_t$. In other words, it expands the state space to the space of distributions on k_t, creating a "curse of dimensionality" problem. Fortunately, given the restriction that the coefficient of relative risk aversion is close to 1, the original power utility can be approximated by a quadratic utility (see Luo 2008); in other words, the above problem can be approximated by a Linear-Quadratic-Gaussian (LQG) framework in which the conditional distributions are Gaussian, so that the first two moments, the conditional mean $\widehat{k}_t$ and the conditional covariance matrix $\Sigma_t = var\left(k_t|\mathcal{I}_t\right)$, are therefore sufficient to characterize the information state. Hence, $\widehat{v}\left(\widehat{k}_t\right)$ is the value function under information processing constraints and $v\left(k_t\right)$ 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: $\Delta v = v\left(k_t\right) - \widehat{v}\left(\widehat{k}_t\right)$.

We use the concept of *entropy* from information theory to characterize the uncertainty of a random variable and then use the reduction in entropy as a measure of information gain. Formally, entropy is defined as the expectation of the negative of the log of the density function, $-E\left[\log\left(f\left(X\right)\right)\right]$ (see Shannon 1948 and Cover and Thomas 1991 for details). For example, the entropy of a discrete distribution with equal weight on two points is simply $E\left[\log\left(f\left(X\right)\right)\right] = -0.5\log_2\left(0.5\right) - 0.5\log_2\left(0.5\right) = 1$, and the unit of information transmitted is called one "bit". In this case, an agent can remove all uncertainty about X if the capacity devoted to monitoring X is $\kappa = 1$ bit. With finite capacity κ, the social planner will choose a signal that reduces the uncertainty of the state. Formally, this idea can be described by the information constraint

$$\mathcal{H}\left(k_{t+1}|\mathcal{I}_t\right) - \mathcal{H}\left(k_{t+1}|\mathcal{I}_{t+1}\right) \leq \kappa, \tag{14}$$

where κ is the consumer's information channel capacity, $\mathcal{H}\left(k_{t+1}|\ \mathcal{I}_t\right)$ denotes the entropy of the state prior to observing the new signal at $t+1$, and $\mathcal{H}\left(k_{t+1}|\ \mathcal{I}_{t+1}\right)$ the entropy after observing the new signal. κ 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.[7] The greater the value of κ, the higher the degree of attention.

[7]Note if we use the natural logarithm e the unit is called a 'nat'. Hence, 1 nat is equal to $\log_2\left(e\right) = 1.4427$ bits. After solving the calibrated model, we will discuss what rates of information transmission would be considered reasonable.

The information constraint restricts information flow in the following manner. Consider the planner's prior estimate of the mean of the information state tomorrow, $k_{t+1}|\,\mathcal{I}_t$; note that this object is an expectation conditional on information up to time t, and thus represents the expected value of tomorrow's wealth based on today's signals. The resulting posterior estimate of the state, $k_{t+1}|\,\mathcal{I}_{t+1}$, incorporates any new information gathered at time $t+1$ (such as a new signal) and is a control variable for the planner. The information processing constraint restricts the planner's ability to choose an updated state by limiting the reduction in entropy – roughly, the increase in the precision of the estimate – to be smaller than κ. We show below that the value of κ determines how much uncertainty can be removed after observing new signals (or equivalently, the fraction of relevant information processed every period by an agent).

Following Sims (2003, 2005) and Luo (2008), we know that $k_t|\,\mathcal{I}_t \sim N\left(\widehat{k}_t, \Sigma_t\right)$. Therefore, (14) can be rewritten as

$$\log|\Psi_t| - \log|\Sigma_{t+1}| \leq 2\kappa \tag{15}$$

where $\Sigma_{t+1} = var_{t+1}\left(k_{t+1}\right)$ and $\Psi_t = var_t\left(k_{t+1}\right)$ are the posterior and the prior variance-covariance matrices of the state vector and $|\cdot|$ is the determinant operator. 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:

$$\Psi_t \succeq \Sigma_{t+1}, \tag{16}$$

(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 κ.

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

$$c_t = \psi\widehat{k}_t \tag{17}$$

$$l_t = \phi\widehat{k}_t \tag{18}$$

where $\widehat{k}_t = E\left(k_t|\mathcal{I}_t\right)$ is the information state; the conditional distribution of

k_t given information at t is $N\left(\widehat{k}_t, \sigma_{k,t}^2\right)$.[8] The linearized dynamic resource constraint can be rewritten as

$$k_{t+1} = \frac{1}{\beta}k_t - \frac{\widetilde{C}}{\widetilde{K}\exp(\mu_z)}c_t + \frac{\widetilde{K}^{\alpha}L^{1-\alpha}}{\widetilde{K}\exp(\mu_z)}(1-\alpha)l_t - \omega\varepsilon_{t+1}; \tag{19}$$

taking the conditional variance implies $var_t\left(k_{t+1}\right) = \omega^2 + \frac{1}{\beta^2}\sigma_{k,t}^2$.

In Appendix A.3 we derive the out-of-steady state filtering equations explicitly. The conditional variance of the noise, $var_t\left[\xi_{k,t+1}\right]$, is time-varying:

$$var_t\left[\xi_{k,t+1}\right] = \frac{\beta^{-2}\sigma_{k,t}^2 + \omega^2}{\exp\left(2\kappa\right) - 1}. \tag{20}$$

However, in the LQG setup the conditional variance $\sigma_{k,t+1}^2$ turns out to be deterministic and policy-independent; consequently, only the behavior of the conditional mean $\widehat{k}_t$ matters for aggregate dynamics. We assume that the initial state of the model economy is $k_0 | \mathcal{I}_0 \sim N\left(\widehat{k}_0, \sigma_{k,0}^2\right)$. Given $\sigma_{k,0}^2$ and $\overline{\sigma}_k^2 = \frac{\omega^2}{\exp(2\kappa)-\beta^{-2}}$, (69) can be written as

$$\sigma_{k,t+1}^2 = \overline{\sigma}_k^2 + \lambda^t\left(\sigma_{k,0}^2 - \overline{\sigma}_k^2\right),$$

which shows how quickly the conditional variance approaches its steady state value. For example, if we set $\beta = 0.99$ and $\kappa = 0.2$ nats (which is close to our quarterly calibration), $\lambda = 0.684$ and $\sigma_{k,t+1}^2$ takes less than 2 quarters to get halfway to its steady state value $\overline{\sigma}_k^2$. Therefore, in this paper we focus on the steady state conditional variance case.

Using the information processing constraint it is straightforward to show that the steady state of the filtering process yields $\overline{\sigma}_k^2 = \frac{\omega^2}{\exp(2\kappa)-\beta^{-2}}$ where $\kappa \geq -\log\left(\beta\right)$ is assumed to guarantee convergence. The agent behaves as if observing a noisy measurement $k_{t+1}^* = k_{t+1} + \xi_{k,t+1}$ with error variance

$$var\left(\xi_{k,t+1}\right) = \frac{\left(\omega^2 + \beta^{-2}\overline{\sigma}_k^2\right)\overline{\sigma}_k^2}{\omega^2 + \left(\beta^{-2} - 1\right)\overline{\sigma}_k^2}. \tag{21}$$

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

[8] Since we only study a univariate case, to avoid confusion we use $\sigma_{k,t}^2$ to denote the conditional variance of the state variable.

the world: $\frac{\partial var\left(\xi_{k,t+1}\right)}{\partial \kappa} < 0$ as long as $\kappa > -\frac{1}{2}\log(\beta)$. For a typical quarterly business cycle calibration (say, $\beta = 0.99$) this would require $\kappa > 0.005$ nats, which seems sufficiently low that it can be safely ignored; note also that this lower bound is satisfied automatically whenever the filter converges to a steady state.

Next, the Kalman filter equation that describes the evolution of the information state can be written more simply as

$$\widehat{k}_{t+1} = (1-\theta)G\widehat{k}_t + \theta\left(k_{t+1} + \xi_{k,t+1}\right), \tag{22}$$

where $\theta = \frac{\overline{\sigma}_k^2}{var\left(\xi_{k,t+1}\right)}$ is the optimal weight on the observation and

$$G = \frac{1}{\beta} - \frac{\widetilde{C}}{\widetilde{K}\exp(\mu_z)}\psi + \frac{\widetilde{K}^{\alpha}L^{1-\alpha}}{\widetilde{K}\exp(\mu_z)}(1-\alpha)\phi.$$

Note that $\theta = \frac{\exp(2\kappa)-1}{\exp(2\kappa)}$ and thus $\lim_{\kappa\to\infty}\theta = 1$; furthermore, $var\left(\xi_{k,t+1}\right) \to 0$ as $\kappa \to \infty$. In other words, the weight of the new signal is increasing with channel capacity and increases to 1 as κ 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 underlying state k_t and for the evolution of the information state $\widehat{k}_t$ (see Appendix A.4 for derivations):

$$k_{t+1} = \frac{1}{\beta}k_t + \left(G - \frac{1}{\beta}\right)\widehat{k}_t - \omega\varepsilon_{t+1} \tag{23}$$

$$\widehat{k}_{t+1} = (1-\theta)G\widehat{k}_t + \theta\left(k_{t+1} + \xi_{k,t+1}\right). \tag{24}$$

We can now derive the expression of $\Delta\widehat{k}_{t+1}$, the change in the information state, as

$$\Delta\widehat{k}_{t+1} = (G-1)\,\widehat{k}_t - \frac{\theta}{\beta}\frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1-(1-\theta)\beta^{-1}L} - \theta\varepsilon_{t+1} + \theta\xi_{k,t+1}, \tag{25}$$

where L is the lag operator and we use the formula $k_t - \widehat{k}_t = -\frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1-(1-\theta)\beta^{-1}L}$.

We now have the expression of changes in log transformed consumption

and labor,

$$\Delta c_{t+1} = \psi \Delta \widehat{k}_{t+1} = \psi \left((G-1)\, \widehat{k}_t - \frac{\theta}{\beta} \frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1-(1-\theta)\beta^{-1}L} - \theta\omega\varepsilon_{t+1} + \theta\xi_{k,t+1} \right) \tag{26}$$

$$\Delta L_{t+1} = \phi \Delta \widehat{k}_{t+1} = \phi \left((G-1)\, \widehat{k}_t - \frac{\theta}{\beta} \frac{(1-\theta)\omega\varepsilon_t + \theta\xi_{k,t}}{1-(1-\theta)\beta^{-1}L} - \theta\omega\varepsilon_{t+1} + \theta\xi_{k,t+1} \right), \tag{27}$$

and we can recover all aggregate variables as follows

$$\Delta \log X_{t+1} = \Delta x_{t+1} + \mu_z + \omega\varepsilon_{t+1};$$

where $X = C, K, I$, or Y and $x = c, k, i$, or y. For the calibration reported in Table 1 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.

Table 1
Calibrated Parameter Values

δ	α	μ_z	η	β	ω
0.018	0.36	0.004	2.84	0.99	0.00732

3 Main Findings

3.1 Business Cycle Statistics

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.

Table 2

Business Cycle Statistics ($\kappa = \infty$, $\kappa = 0.2$ nats)

Variable	Std Dev.	Cross-Correlations				
y_t	$(0.90, 0.81)$	$(0.46, 0.52)$	$(0.71, 0.75)$	$(1.00, 1.00)$	$(0.71, 0.75)$	$(0.46, 0.52)$
c_t	$(0.44, 0.75)$	$(0.55, 0.15)$	$(0.75, 0.34)$	$(0.98, 0.62)$	$(0.65, 0.53)$	$(0.38, 0.42)$
i_t	$(2.32, 2.56)$	$(0.40, 0.53)$	$(0.66, 0.66)$	$(0.99, 0.72)$	$(0.72, 0.49)$	$(0.50, 0.30)$
l_t	$(0.48, 0.49)$	$(0.35, 0.52)$	$(0.63, 0.68)$	$(0.98, 0.81)$	$(0.73, 0.58)$	$(0.52, 0.39)$
Business Cycle Statistics ($\kappa = 0.2$, $\xi_{k,t} = 0\ \forall t$)						
y_t	0.78	0.55	0.77	1.00	0.77	0.55
c_t	0.64	0.20	0.48	0.91	0.73	0.55
i_t	1.56	0.85	0.94	0.86	0.62	0.41
l_t	0.33	0.75	0.91	0.93	0.71	0.50

Note: we use the expressions for main variables to simulate 1000 artificial samples of length 225, HP-filter data, and then compute the statistics from the standard RE model $\kappa = \infty$ and the RI model $\kappa = 0.2$ nats.

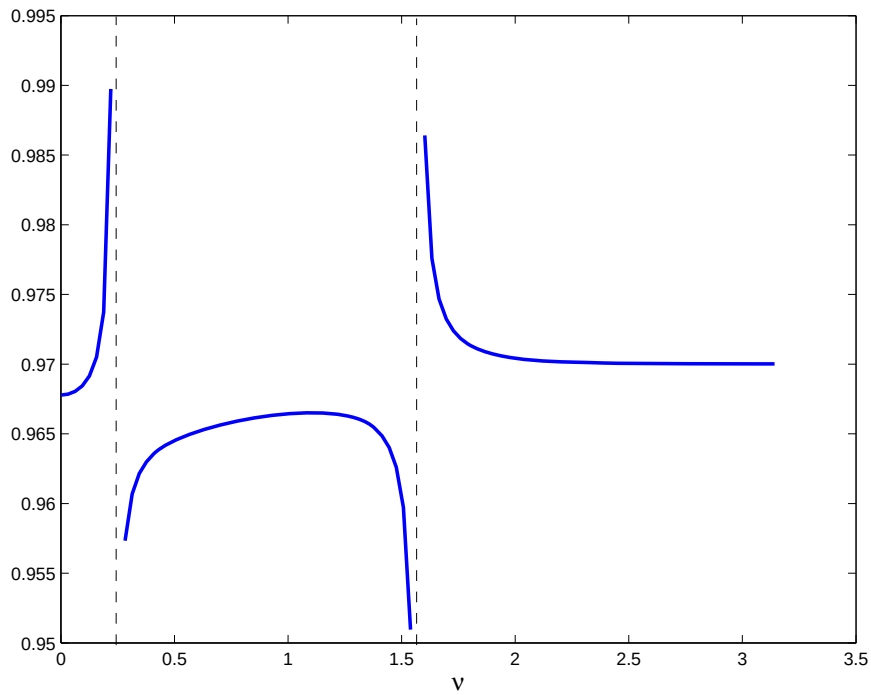
Output becomes less volatile, consumption and investment become more volatile (especially consumption, whose volatility nearly doubles), and hours remains the same. Consumption doubles in volatility because imperfectly observing the state of the world leads to a reduced ability to smooth consumption, similar to the results in Luo (2008). 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 (see 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 $\widehat{k}_t$ are highly correlated and have very similar volatilities, even when κ 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.[9] 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

[9]These are the moments of the unfiltered, normalized series.

frequency domain, which examines their similarity across all frequencies. In Figure 1 we plot the coherence between k_t and $\widehat{k}_t$ across the frequency band $\nu \in [0, \pi]$ (see Appendix A.5 for derivations of the coherence); the two series display strong positive coherence at almost every frequency, indicating that they are driven by movements at the same frequencies.[10] The coherence remains strong even for very low values of θ.

Figure 1: Coherence between k and $\widehat{k}$



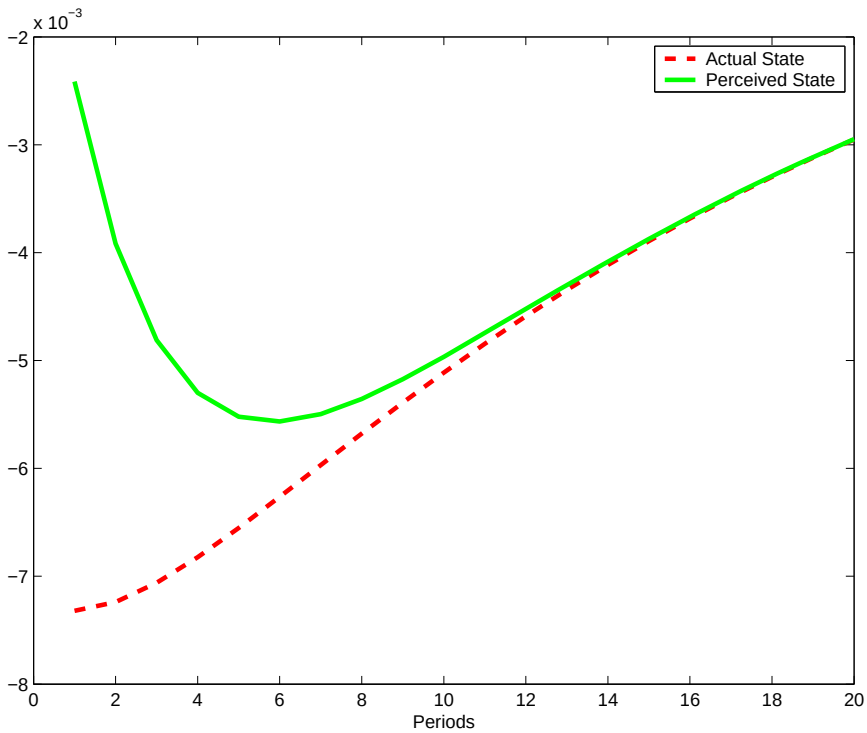
Note: The coherence is calculated across the frequency band $[0, \pi]$ when $\kappa = 0.2$ nats.

We can use the consumption volatility number, which seems to be the most sensitive to κ, to obtain an estimate of the channel capacity. We choose κ 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

[10]There are two poles in the joint process for $\left(k_t, \widehat{k}_t\right)$, one located between frequencies $[0.2199, 0.2827]$ and one 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.

signal. Our estimate is in the ballpark of some others found in the literature, although on the high side. Luo and Zhang (2008) obtain an estimate equivalent to $\theta = 0.14$ or $\kappa = 0.075$ using G-7 data on consumption and productivity. Adam (2005) finds $\theta = 0.4$ or $\kappa = 0.255$ based on the response of aggregate output to monetary policy shocks. To determine whether these estimates are plausible, we note that Landauer (1986) estimated individuals process about 2 bits per second in the laboratory. Given that this information flow is the upper limit, it is reasonable to think that agents devote only a small fraction of their capacity to aggregate conditions, both because idiosyncratic shocks are typically much larger and because the welfare costs of limited information appear to be tiny (see Luo 2008 and Luo and Young 2008). We discuss idiosyncratic shocks and their effects below.

Figure 2: Impulse Response to Technology Shock

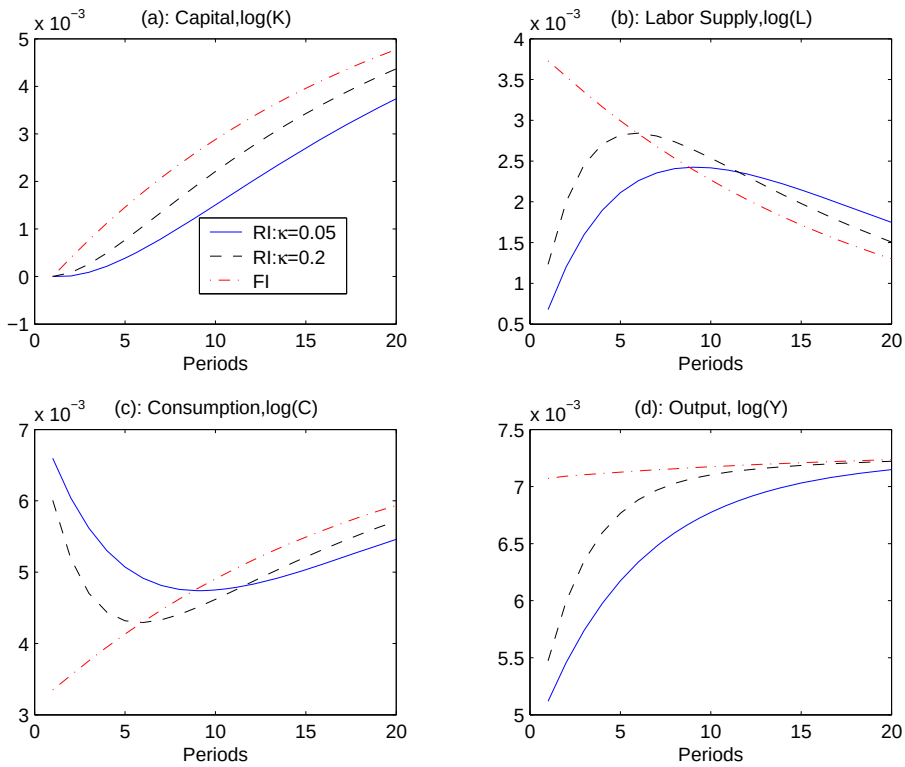


Note: The time unit of all figures (period) is one quarter and $\kappa = 0.2$.

3.2 Impulse Responses

In this section we focus on the implications of RI for the shape of the impulse response functions for the model. 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.

Figure 3: Impulse Response to Technology Shock

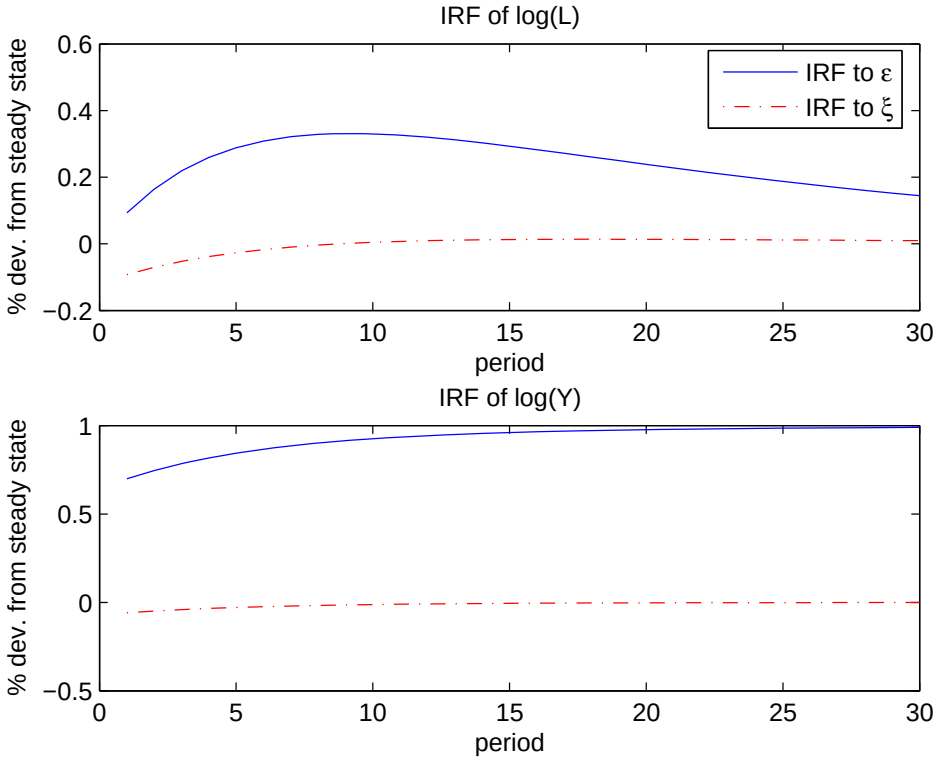


Note: Solid, dashed, dashed-dotted lines depict the impulse responses to the technology shock when $\kappa = 0.05$, $\kappa = 0.2$, and $\kappa = \infty$ nats, respectively.

Before we examine the impulses of main macroeconomic variables, it is useful to examine the impulses of the true state k_t and the information state $\widehat{k}_t$, 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) $\widehat{k}_t$ 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 information 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.

Figure 4: Impulse Responses to Technology and Noise Shocks



Note: ε and ξ are the technology shock and the endogenous noise, respectively.

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 κ). The solid and dashed lines in Figure 3b show that labor supply reacts to the 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 (ε) and the endogenous noise (ξ_k). From the two-equation dynamic system, (23) and (24), it is clear that since the exogenous technology shock ε_t appears in the dynamic system of $\left(k_t, \widehat{k}_t\right)$ with a negative sign, the endogenous noise $\xi_{k,t}$ due to finite capacity serves as a negative demand shock. Figure 4 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 (2008), 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

$$\begin{aligned}\Delta \log Y_{t+1} =& \mu_z + \left(\alpha G + (1-\alpha)\,\phi\,(G-1)\right)\widehat{k}_t - \alpha k_t + \omega\,(1-\alpha)\,(1+\phi\theta)\,\varepsilon_{t+1} + \\ & (1-\alpha)\,\phi\theta\xi_{k,t+1} - \frac{(1-\theta)\,\omega}{\beta}\left(\frac{\alpha + (1-\alpha)\,\phi\theta}{1 - (1-\theta)\beta^{-1}L}\right)\varepsilon_t + \theta\left(\frac{\alpha + (1-\alpha)\,\phi}{1 - (1-\theta)\beta^{-1}L}\right)\xi_{k,t}.\end{aligned} \tag{28}$$

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, the same value that ensures convergence of the filter.[11]

We plot the calibrated impulse response function for the growth rate of the log of output to a one-standard-deviation increase in ε_t in Figure 5 for $\kappa = 0.2$ and $\kappa = \infty$. With $\kappa = 0.2$ nats the growth rate of output follows the process

$$\begin{aligned}\Delta \log Y_{t+1} = 0.004 + 0.3582\widehat{k}_t - 0.36k_t + 0.0039\varepsilon_{t+1} - 0.1078\xi_{k,t+1} - \\ 0.0012\sum_{j=0}^{\infty}\left(0.6767^j\varepsilon_{t-j}\right) + 0.0109\sum_{j=0}^{\infty}\left(0.6767^j\xi_{k,t-j}\right).\end{aligned} \tag{29}$$

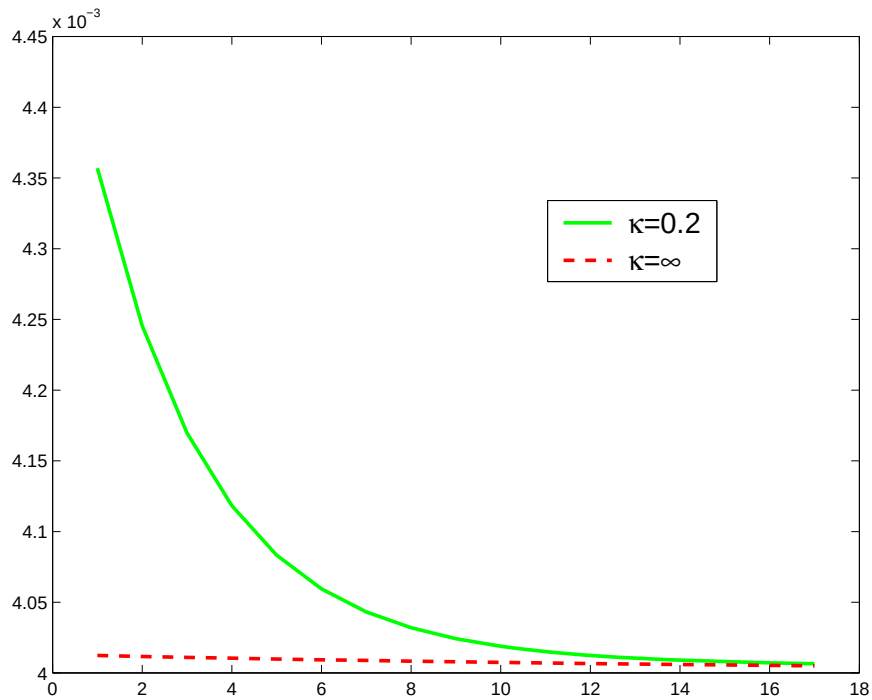
[11] As noted in Sims (2005), it is also the case that smaller values for κ make the linear approximation less valid. We do not explore values too close to the bound because we encountered many periods in which aggregate investment became negative, rendering the approximation invalid.

The effect of technology shocks on the growth rate of output dies off quite slowly, even in the absence of a strong capital accumulation mechanism, due to the significant effects of the lagged ε_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 $\widehat{k}_t$ nearly cancel each other, but rather by the learning process. In contrast, when $\kappa = \infty$ the output growth is given by

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

Clearly, this process will not display much persistence. Figure 5 contrasts 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 κ) generates a small amount of persistence into the output growth function. As noted above, the effect is quantitatively small.

Figure 5: Impulse Response of Output Growth



3.3 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.[12] 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:

$$\begin{aligned}\Delta \log Y_{t+1} =& \mu_z + (\alpha G + (1-\alpha)\phi(G-1))\widehat{k}_t - \alpha k_t + \omega(1-\alpha)(1+\phi\theta)\varepsilon_{t+1} + \\ & (1-\alpha)\phi\theta\xi_{k,t+1} - \frac{(1-\theta)\omega}{\beta}\left(\frac{\alpha+(1-\alpha)\phi\theta}{1-(1-\theta)\beta^{-1}L}\right)\varepsilon_t + \theta\left(\frac{\alpha+(1-\alpha)\phi}{1-(1-\theta)\beta^{-1}L}\right)\xi_{k,t}.\end{aligned} \tag{30}$$

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

$$\begin{aligned}& covar\left(\Delta \log Y_{t+j}, \Delta \log Y_t\right) \\ =& \left(\alpha(G-1) + (1-\alpha)\phi(G-1)\right)^2 covar\left(k_{t+j-1}, k_{t-1}\right) + \\ & \left(\alpha(G-1) + (1-\alpha)\phi(G-1)\right)\omega(1-\alpha)(1+\phi)\, covar\left(k_{t+j-1}, \varepsilon_t\right).\end{aligned} \tag{31}$$

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

$$\begin{aligned}covar\left(\Delta \log Y_{t+j}, \Delta \log Y_t\right) = & \left(3.2 \times 10^{-6}\right) covar\left(k_{t+j-1}, k_{t-1}\right) - \\ & \left(4.1 \times 10^{-6}\right) covar\left(k_{t+j-1}, \varepsilon_t\right);\end{aligned}$$

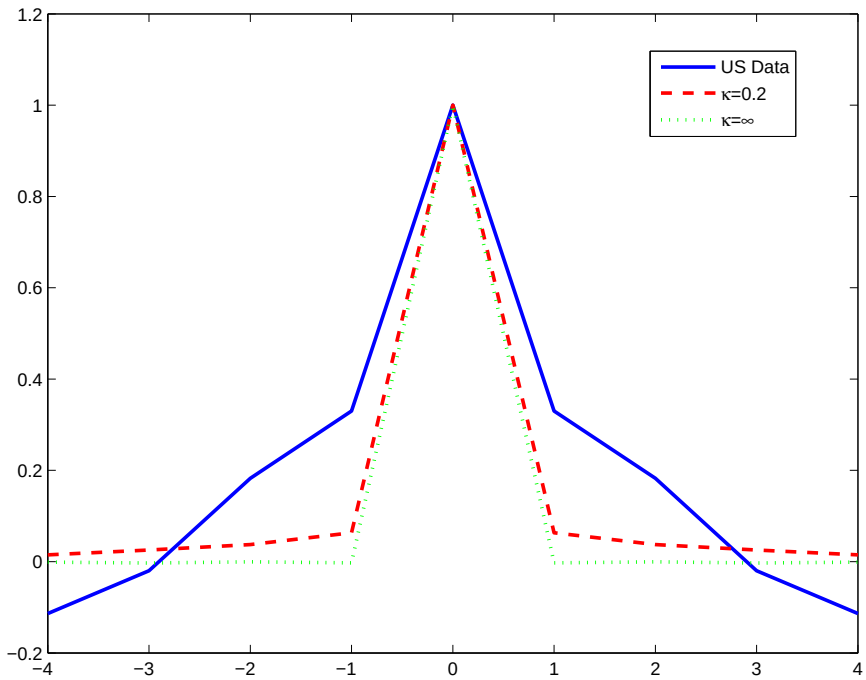
thus, output growth will be white noise in the absence of rational inattention unless capital accumulation is very strong, and previous results have shown the weakness of this mechanism.

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 ε 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

[12] These two values are around 0.4 and 0.2 in the US data, both of which are statistically distinguishable from 0; see Figure 6.

simulations used to compute the business cycle statistics, and compares them to the data. It is clear that RI does increase the autocorrelation of output growth, but the effect is small even for a relatively low value of κ – we cannot reproduce the observed values of 0.4 and 0.2 for the first two lags. 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.

Figure 6: Autocorrelation of Output Growth



Note: Leads correspond to negative x-axis entries and lags to positive entries. Lead/lags are in quarters.

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.[13] Changes in $\frac{C}{Y}$ are nonmonotone, and our chosen calibration target (which is very close to the value advocated by

[13] An alternative approach is to calibrate the model to match interest rates rather than

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 α have little effect.

3.4 Forecastable Movements

We follow the procedure in Rotemberg and Woodford (1996) by examining forecastable movements in output. For $j \geq 1$, let Δy_t^j denote the difference between y_{t+j} and y_t:

$$\Delta y_t^j = \log(Y_{t+j}) - \log(Y_t) = \sum_{i=1}^{j} \Delta \log(Y_{t+i}) \tag{32}$$

The forecastable change in output between t and $t+j$ is then given by $E_t\left(\Delta y_t^j\right)$. In the model with RI these expectations must be consistent with the information state $\widehat{k}_t$. Expected output at time t, conditional on $\widehat{k}_t$, is given by

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

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

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

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

$$E_t(y_{t+j}) = (\alpha + (1-\alpha)\phi)\,B^j \widehat{k}_t,$$

where $B = \frac{1-\delta}{\exp(\mu)} + \exp(i)\,\frac{\alpha+(1-\alpha)\phi-\psi}{\exp(\mu+k)}$. We can therefore obtain the standard deviation of the forecastable part of output growth

$$\sigma\left(E_t \Delta y_t^j\right) = (\alpha + (1-\alpha)\phi)\left(B^j - B\right)\sigma\left(\widehat{k}_t\right), \tag{33}$$

where $E_t \Delta y_t^j = (\alpha + (1-\alpha)\phi)(B^j - B)\,\widehat{k}_t$. Table 3 shows that RI tends to raise the volatility of the forecastable component of output growth, but not

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.

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 (32) depends on the standard deviation of $\widehat{k}_t$ rather than k_t, as the coefficients do not depend on θ. We have previously noted that these two variables are very similar in their dynamic behavior even for small values of κ.

Table 3

Predicted Standard Deviations of Δy_t^j

Variable	$j = 1$	$j = 2$	$j = 4$	$j = 32$
$\kappa = 0.01$ nats	0.0003	0.0006	0.0011	0.0047
$\kappa = 0.2$ nats	0.0003	0.0005	0.0010	0.0045

Note: we report the volatility of the forecastable component of output growth when $\kappa = 0.01$ nats and $\kappa = 0.2$ nats, respectively.

4 Idiosyncratic and Aggregate Shocks

In the previous section, we solved a linear approximation to a stochastic growth model with RI and permanent technology shocks numerically and discussed the implications of RI for the dynamics of employment, consumption, capital stock, and output; critical for our discussion was the assumption that the social planner has the same constraints on information processing as the individual agents do, meaning that idiosyncratic shocks are unimportant. In this section we show that if individuals cannot distinguish an idiosyncratic shock from the aggregate shock in their productivity process (an assumption we refer to as *incomplete information*), the responses of main macroeconomic variables to the aggregate productivity shock under RI display more inertia than that found in the case in which individuals can separate the two components. For the purposes of this section we use an explicitly linear-quadratic setting because, with prices exogenous, the competitive and social planning allocations are identical.

4.1 LQ Business Cycle Model

Following the same procedure in Section 2, we first derive the optimal decisions under RE and then obtain the decisions under RI using the separation principle. Individuals are infinitely-lived; household i maximizes expected lifetime utility

$$\max_{\{c_t,l_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \left[c_t^i - \frac{1}{2}\left(c_t^i\right)^2 - \eta \left(l_t^i - \frac{1}{2}\vartheta \left(l_t^i\right)^2 \right) \right] \tag{34}$$

subject to

$$k_{t+1}^i = (1 + A_1)\, k_t^i - c_t^i + A_2 l_t^i + z_t^i, \tag{35}$$

where the time-separable utility is quadratic in terms of consumption and labor supply and output is linear in capital k_t^i and labor l_t^i.[14] There is an additive technology, z_t^i, that is the sum of *aggregate* permanent and *idiosyncratic* transitory components:

$$z_{t+1} = z_{t+1}^a + z_{t+1}^i, \tag{36}$$

where the superscripts a and i denote aggregate and idiosyncratic, respectively. Each of these components follows its own stochastic process; z_{t+1}^a follows a random walk

$$z_{t+1}^a = z_t^a + \varepsilon_{t+1}^a, \tag{37}$$

and z_{t+1}^i follows an *iid* process

$$z_{t+1}^i = \overline{z} + \epsilon_{t+1}^i, \tag{38}$$

where ε_{t+1}^a and ϵ_{t+1}^i are white noises with mean 0 and variance ω_ε^2 and ω_ϵ^2, respectively. To explore the effects of the idiosyncratic risk on aggregate dynamics under RI, following Pischke (1995) we suppose that households can only observe the sum of the aggregate shock and the idiosyncratic shock, not the realizations separately.[15] Specifically, given that the change in the total productivity is

$$\Delta z_{t+1}^i = \varepsilon_{t+1}^a + \epsilon_{t+1}^i - \epsilon_t^i, \tag{39}$$

individuals only observe an MA(∞) process for Δz_{t+1}^i:

$$\Delta z_{t+1}^i = \nu_{t+1}^i - \mu \nu_t^i, \tag{40}$$

where the innovation, ν_t^i, with mean 0 and variance ω_ν^2, is not a fundamental driving process – it contains information on current and lagged aggregate and idiosyncratic productivity shocks. Equating the variances and autocorrelation

[14] For simplicity we subsume depreciation into A_1. The coefficients satisfy $\eta > 0$, $\vartheta > 0$, $A_1 > 0$, and $A_2 > 0$.

[15] Similar assumptions in standard business cycle models – such as Kydland and Prescott (1982) – have little effect on aggregate dynamics.

coefficients of the original and derived processes (38) and (39), we have

$$\mu = -\frac{1-\sqrt{1-4\varrho^2}}{2\varrho} \text{ where } \varrho = -\frac{\omega_\epsilon^2}{\omega_\varepsilon^2 + 2\omega_\epsilon^2} \tag{41}$$

$$\omega_\nu^2 = \frac{\omega_\epsilon^2}{\mu}; \tag{42}$$

$\mu \in [0,1]$ will be large if the variance of the idiosyncratic shock ω_ϵ^2 is large relative to the variance of the aggregate shock ω_ε^2 and will converge to 0 as ω_ϵ^2 approaches to 0. Since studies of individual income data show that aggregate shocks account for very little of the variation in individual incomes, μ will be close to 1.[16]

To obtain a well-behaved stochastic steady state in this LQ model, we assume that $\beta(1+A_1) = 1$. Given this assumption, solving the LQ model gives the optimal consumption and labor supply under RE:

$$c_t^i = \frac{\vartheta\eta A_1}{\vartheta\eta - A_2^2}\left(k_t^i + \frac{1}{A_1}z_t^i - \frac{\mu}{A_1}\nu_t^i\right) + \frac{A_2(\eta - A_2)}{\vartheta\eta - A_2^2}, \tag{43}$$

$$l_t^i = \frac{A_1 A_2}{\vartheta\eta - A_2^2}\left(k_t^i + \frac{1}{A_1}z_t^i - \frac{\mu}{A_1}\nu_t^i\right) + \frac{\vartheta\eta(\eta - A_2)}{\vartheta\eta - A_2^2}. \tag{44}$$

For completeness, GDP is given by $y_t = A_1 k_t + A_2 l_t + z_t$.

4.2 Information Processing Constraints

To solve for the optimal decisions under RI, we reduce the original multivariate model to a univariate model by defining a new unique state variable:

$$a_t^i = k_t^i + \frac{1}{A_1}z_t^i - \frac{\mu}{A_1}\nu_t^i. \tag{45}$$

The original budget constraint can then be rewritten as

$$a_{t+1}^i = (1+A_1)\,a_t^i - c_t^i + A_2 l_t^i + \frac{1+A_1-\mu}{A_1}\nu_{t+1}^i. \tag{46}$$

The optimal distribution of the state a_t^i follows a normal distribution: $a_t^i \sim N(\widehat{a}_t^i, \sigma_t^2)$, where $\widehat{a}_t^i$ is the conditional expectation of the true state and $\sigma_t^2 =$

[16] Consistent with the market structure assumed in most of this literature, we assume that the idiosyncratic shocks to income are uninsurable through contingent claims markets. Given that idiosyncratic shocks are not observable even without RI, this assumption is not restrictive.

$var_t(a_t)$ is the conditional variance of the state. The information processing constraint,

$$\log\left(var_t\left(a_{t+1}^i\right)\right) - \log\left(var_{t+1}\left(a_{t+1}^i\right)\right) = 2\kappa,$$

implies that the steady state for the conditional variance σ_t^2 is $\overline{\sigma}^2 = \frac{\left(\frac{1+A_1-\mu}{A_1}\omega_\nu\right)^2}{\exp(2\kappa)-\beta^{-2}}$. The optimal choice of information structure leads the agent to act as if she observes a Gaussian noisy signal $a_{t+1}^{i,*} = a_{t+1}^i + \xi_{t+1}^i$, where a_{t+1}^i is the true signal and ξ_{t+1}^i is the endogenous noise due to RI. Hence, the evolution of the information state $\widehat{a}_t^i$ follows a recursive Kalman filter equation

$$\widehat{a}_{t+1}^i = (1-\theta)\,\widehat{a}_t^i + \theta\left(a_{t+1}^i + \xi_{t+1}^i\right), \tag{47}$$

where $\theta = 1 - 1/\exp(2\kappa)$ is the optimal weight on a new observation and $\omega_\xi^2 = var(\xi_t) = \frac{\overline{\sigma}^2}{\theta}$. Since the separation principle holds in this case, we can just replace the true state in the decision rules with the information state and obtain the optimal consumption and labor supply functions under RI:

$$c_t^i = \frac{\vartheta\eta A_1}{\vartheta\eta - A_2^2}\widehat{a}_t^i + \frac{A_2(\eta - A_2)}{\vartheta\eta - A_2^2}, \tag{48}$$

$$l_t^i = \frac{A_1 A_2}{\vartheta\eta - A_2^2}\widehat{a}_t^i + \frac{\vartheta\eta(\eta - A_2)}{\vartheta\eta - A_2^2}. \tag{49}$$

Combining (45) and (46) with (47) and (48) gives the expression for the change in individual households' information state:

$$\Delta\widehat{a}_{t+1}^i = \frac{\theta}{A_1}\frac{1+A_1-\mu}{1+A_1}\left(\frac{\varepsilon_{t+1}^a + \epsilon_{t+1}^i - \epsilon_t^i}{1-(1-\theta)(1+A_1)L}\right) + \theta\left(\xi_{t+1}^i - \frac{\theta(1+A_1)\xi_t^i}{1-(1-\theta)(1+A_1)L}\right).$$

After aggregating over all households, we obtain the aggregate change in the information state:

$$\begin{aligned}\Delta\widehat{a}_{t+1} =& \frac{\theta}{A_1}\frac{1+A_1-\mu}{1+A_1}\left(\frac{\varepsilon_{t+1}^a}{(1-(1-\theta)(1+A_1)L)(1-\mu L)}\right) + \\ & \theta\left(\overline{\xi}_{t+1} - \frac{\theta(1+A_1)\overline{\xi}_t}{1-(1-\theta)(1+A_1)L}\right),\end{aligned} \tag{50}$$

where $\overline{\xi}_t = E^i\left(\xi_t^i\right)$ is the common noise and $E^i(\cdot)$ is the population average.

As argued in Sims (2003), although the randomness in an individual's response to the aggregate shock should also be *idiosyncratic* because it arises from his *own* internal information processing constraint, a considerable part of the idiosyncratic error might be common across individuals. The intuition is that agents with finite capacity might make common mistakes because they process macroeconomic information from some common sources such as

newspapers, TV, or other media. However, the 'pure' RI theory says nothing about this common component, and to the best of our knowledge, no other existing theory can help us pin down the relative importance of the common error. Here we therefore assume that the common term of the idiosyncratic error, $\overline{\xi}_t$, is somewhere between 0 and the part of the idiosyncratic error ξ_t caused by the non-fundamental shock ν_t; the variance of ξ_t is

$$\omega_\xi^2 = \frac{1-\theta}{\theta\left(1-(1-\theta)\beta^{-2}\right)}\left(\frac{1+A_1-\mu}{A_1}\right)^2\omega_\nu^2, \tag{51}$$

which depends on both aggregate and idiosyncratic components. Formally, we assume that ξ_t consists of two independent noises, $\xi_t = \overline{\xi}_t + \xi_t^i$, where $\overline{\xi}_t = E^i(\xi_t)$ and ξ_t^i are the common and idiosyncratic components of the error generated by ν_t, respectively. A single parameter,

$$\lambda = \frac{var\left(\overline{\xi}_t\right)}{var\left(\xi_t\right)} \in [0,1],$$

can be used to measure the common source of coded information on the aggregate component (or the relative importance of $\overline{\xi}_t$ to ξ_t). Therefore, RI affects the impact of the exogenous shocks on consumption, labor supply, and output through the factor $\frac{1}{(1-(1-\theta)(1+A_1)L)(1-\mu L)}$.

In order to simplify expressions we consider the case where all noises are idiosyncratic (so that individuals live on isolated islands and do not interact with each other directly or indirectly via conversation, imitation, newspapers, or other media), meaning that $\lambda = 0$; in this special case (which is consistent with the pure RI theory developed by Sims 2003) the change in the perceived state can be written as

$$\Delta\widehat{a}_{t+1} = \frac{\theta}{A_1}\frac{1+A_1-\mu}{1+A_1}\left(\frac{\varepsilon_{t+1}^a}{\left(1-(1-\theta)(1+A_1)L\right)(1-\mu L)}\right). \tag{52}$$

This equation brings out two salient points in our aggregate RI model. First, both RI and incomplete information provide endogenous propagation mechanisms of the LQ-RBC model – they are characterized by the two factors, $\frac{1}{1-(1-\theta)(1+A_1)L}$ and $\frac{1}{1-\mu L}$, respectively. Second, under incomplete information, the presence of the idiosyncratic shock plays a role in strengthening the inertial responses to the aggregate productivity shock because μ is a function of the variance of the idiosyncratic shock. Furthermore, because we expect μ to be close to 1, the impact is expected to be initially small but highly persistent.

4.3 Results

We now show how μ and θ affect dynamics in the model. The relative volatility of the change in the information state relative to the innovation to the aggregate productivity can be written as

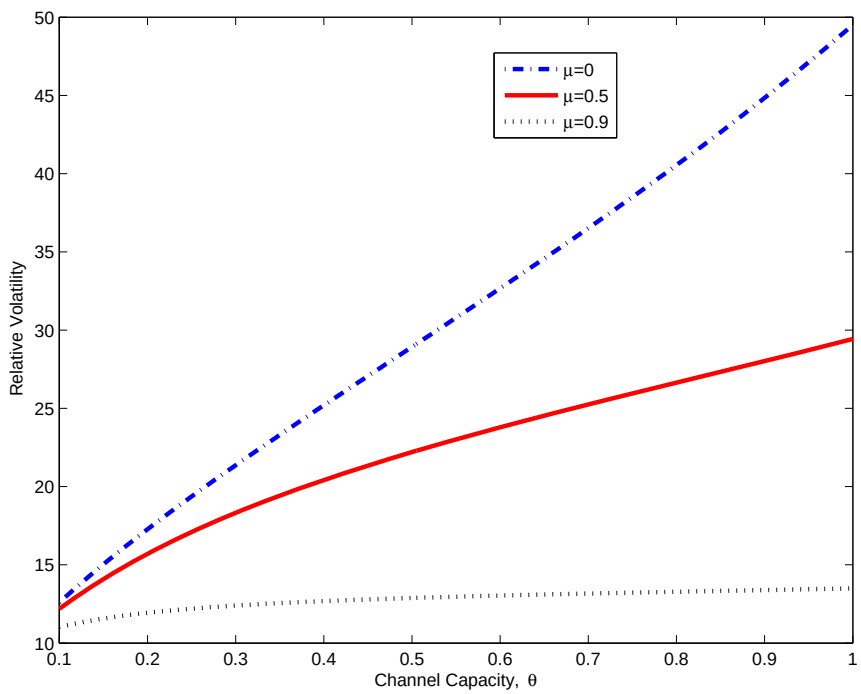
$$rv = \frac{sd\left[\Delta\widehat{a}_t\right]}{sd\left[\Delta z_t^a\right]}, \tag{53}$$

where the standard deviation of $\Delta\widehat{a}_t$ – denoted $sd\left[\Delta\widehat{a}_t\right]$ – can be written as

$$sd\left[\Delta\widehat{a}_t\right] = \sqrt{\begin{array}{c}\frac{\left(\frac{\theta}{A_1}\right)^2\left(\frac{1+A_1-\mu}{1+A_1}\right)^2(1+\mu(1-\theta)(1+A_1))}{(1+\mu(1-\theta)(1+A_1))\left[(1+\mu(1-\theta)(1+A_1))^2-(\mu+(1-\theta)(1+A_1))^2\right]}\omega_\varepsilon^2+ \\ \lambda^2\theta^2\left[1+\frac{(\theta(1+A_1))^2}{1-((1-\theta)(1+A_1))^2}\right]var\left[\xi_t\right]\end{array}}.$$

Specializing to the case where $\lambda = 0$ (no common information) we get

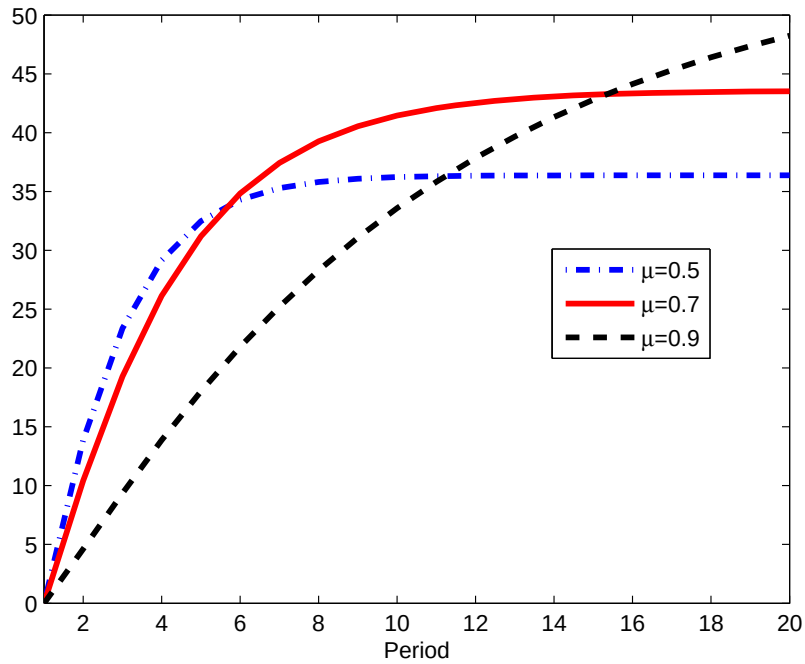
Figure 7: The Relative Volatility of Main Variables to Aggregate Productivity



$$rv = \frac{\theta}{A_1}\frac{1+A_1-\mu}{1+A_1}\times \sqrt{\frac{1+\mu(1-\theta)(1+A_1)}{(1-\mu(1-\theta)(1+A_1))\left((1+\mu(1-\theta)(1+A_1))^2-(\mu+(1-\theta)(1+A_1))^2\right)}}.$$

This expression shows how the volatility of aggregate variables is determined by (1) the degree of attention θ and (2) the relative importance of the idiosyncratic shock to the aggregate shock μ. Figure 7 illustrates how the combinations of (θ, μ) affect the relative volatility – rv is increasing with θ and decreasing with μ. Thus, rational inattention – $\theta < 1$ – does have implications when agents cannot distinguish aggregate and idiosyncratic shocks. Specifically, it increases the volatility of any aggregates that depend on $\widehat{a}_t$ (since by the separation principle the coefficients don't change) and it reinforces the volatility effects of incomplete information.

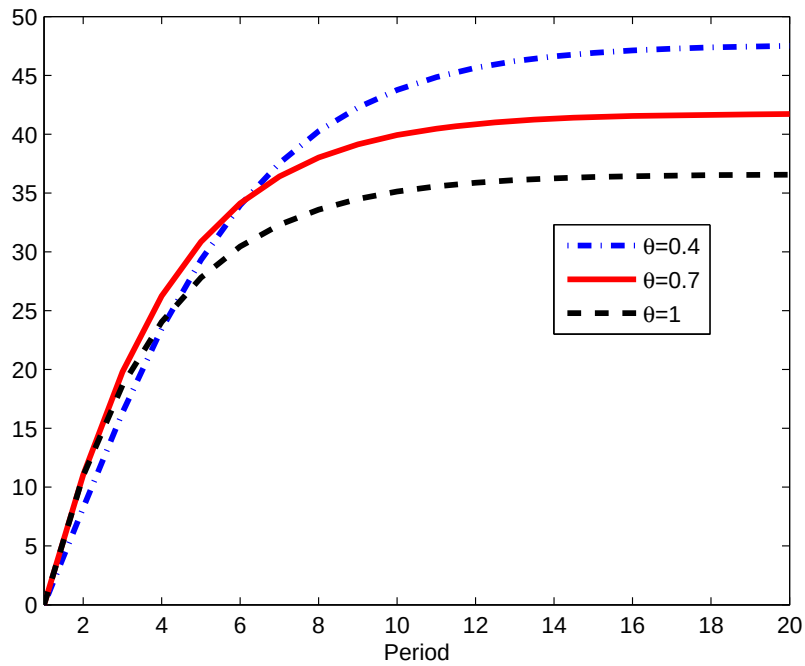
Figure 8: Autocorrelation of Output Growth



Turning to persistence, Figures 8 and 9 plot the dynamics of the state a_t (or $\widehat{a}_t$) in response to a technology shock under various different values for μ and θ. Both figures show that slower (and ultimately larger) adjustment occurs as μ increases (θ decreases). When agents cannot distinguish aggregate from idiosyncratic shocks, filtering assigns more weight to more volatile components. In our case, households assign higher weight to the idiosyncratic component, which is not persistent, because μ is close to one; thus, the total impact of the productivity shock is only discovered over time, leading to persistent responses. Figures 10 and 11 show that the persistence shows up in output growth the

autocorrelation function displays nonzero values at higher leads/lags than does the model we presented above without idiosyncratic shocks. In fact, we overpredict the persistence of output growth relative to the data, particularly for high values of μ and low values of θ (the ones that are reasonable given empirical evidence). That we overpredict the autocorrelation of output growth here is not surprising, given that we have eliminated a key mechanism – the decreasing marginal product of capital – that limits the power of shocks to generate persistent movements in output.

Figure 9: Autocorrelation of Output Growth

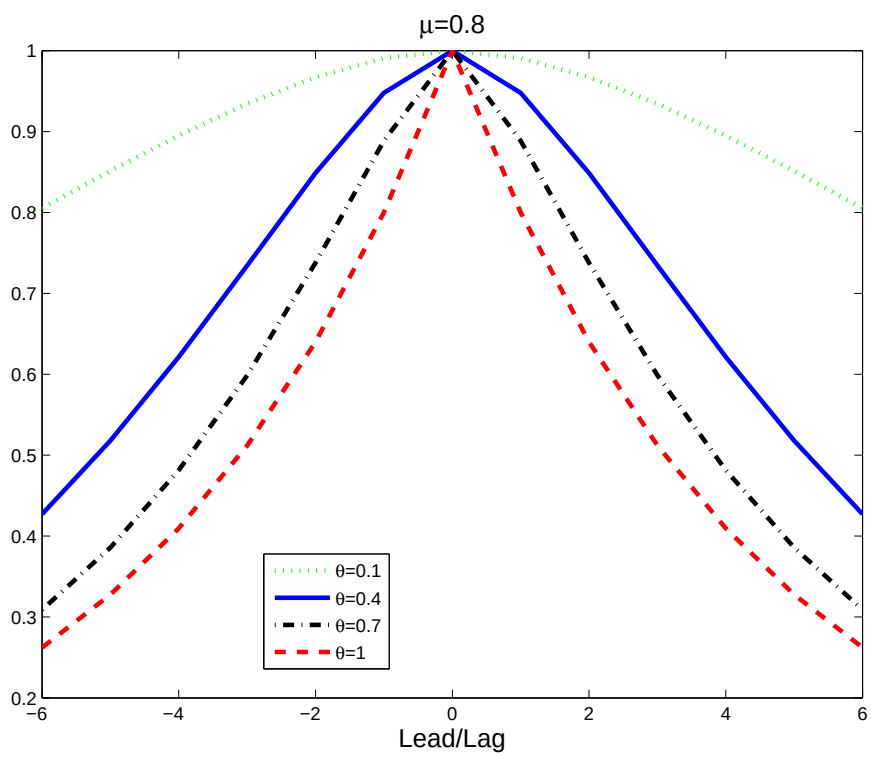


5 Conclusion

In this paper we have reconsidered some puzzles in the RBC literature by via the introduction of 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 is not very sensitive to this parameter, 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 value corresponds to $\kappa = 0.2$ nats, the parametrization which produced the maximum autocorrelation in output growth in the business cycle model). 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. However, when we conduct experiments in a model in which aggregate and idiosyncratic shocks cannot be distinguished, the RI model displays stronger propagation, too strong in fact, particularly when idiosyncratic shocks are the dominant source of fluctuations in individual productivity (as they are in the data). Given that the LQ model with idiosyncratic shocks we use above ignores general equilibrium effects through the interest rate that would tend to weaken output growth persistence, we think that RI can have a significant effect.

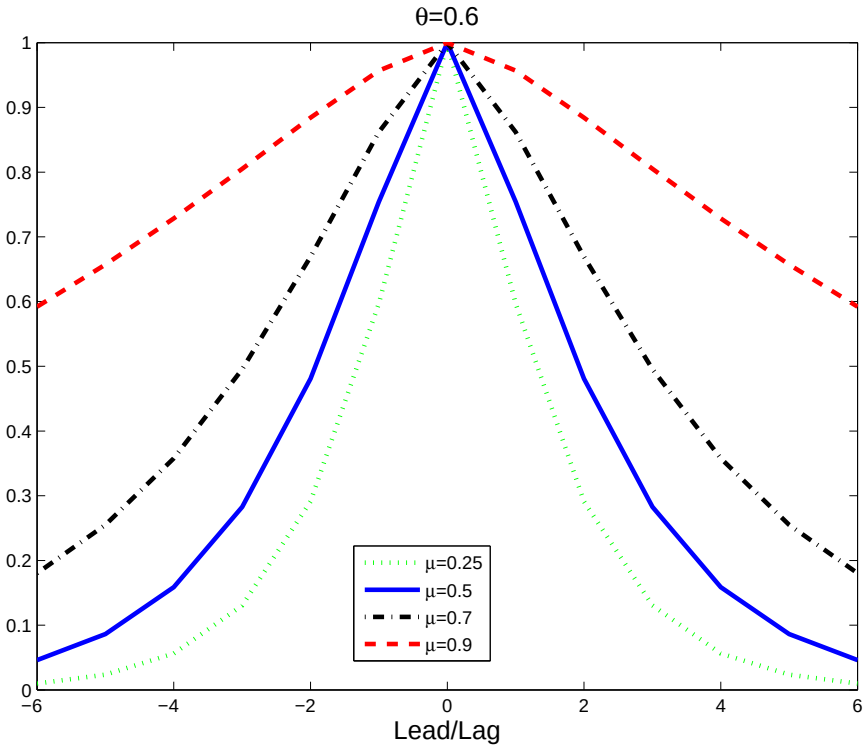
Figure 10: Autocorrelation of Output Growth



Note: Leads correspond to negative x-axis entries and lags to positive entries. Lead/lags are in quarters.

There are some caveats to our results that we feel suggest even 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; furthermore, without precautionary savings one important channel through which RI would affect the economy is shut off. 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 fairly 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. Furthermore, given the unknown

Figure 11: Autocorrelation of Output Growth



Note: Leads correspond to negative x-axis entries and lags to positive entries. Lead/lags are in quarters.

form something similar to an MCMC approach will be needed to construct the posterior, making 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.

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 and given previous results on the poor behavior of the fully-nonlinear model, solving RI models with multiple state variables seems difficult but worthwhile.

A Appendix

A.1 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

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

subject to

$$\begin{aligned}
&c_t + i_t \leq r_t k_t + w_t l_t \\
&k_{t+1} \leq \left(1-\delta\right) k_t + i_t \\
&\log\left(\Psi_t^e\right) - \log\left(\Sigma_{t+1}^e\right) \leq 2\kappa \\
&\Psi_t^e \succeq \Sigma_{t+1}^e.
\end{aligned}$$

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

$$\begin{aligned}
r_t &= \alpha \exp\left(z_t\right) K_t^{\alpha-1} L_t^{1-\alpha} \\
w_t &= \left(1-\alpha\right) \exp\left(z_t\right) K_t^{\alpha} L_t^{-\alpha}.
\end{aligned}$$

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 $\{k_t^i\}$, where $\sum_i k_t^i = K_t$ 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 bargaining. Similar concerns arise in the sticky information literature, such as Mankiw and Reis (2006), where price-setting firms would want to exploit agents whose information has gone stale. That literature has typically proceeded by restricting the information problem to one variable per decision-maker, and is thus inconsistent with limited Shannon channel capacity.

A.2 Deriving Optimal Decisions

In this appendix, we follow Campbell (1994) and use a log-linearization approach to solve the standard welfare maximization problem proposed in Section 2. First, the efficiency conditions for the planning problem are

$$\eta = \widetilde{C}_t^{-\gamma} (1-\alpha) \frac{\widetilde{Y}_t}{L_t} \tag{54}$$

$$\widetilde{C}_t^{-\gamma} = \beta E_t \left[\exp\left(-\mu_z - \omega\varepsilon_{t+1}\right) \left(\widetilde{C}_{t+1}^{-\gamma}\right) \left(1 + \alpha \frac{\widetilde{Y}_{t+1}}{\widetilde{K}_{t+1}} - \delta\right) \right] \tag{55}$$

$$\exp\left(\mu_z + \omega\varepsilon_{t+1}\right) \widetilde{K}_{t+1} = \widetilde{K}_t^{\alpha} L_t^{1-\alpha} + (1-\delta) \widetilde{K}_t - \widetilde{C}_t \tag{56}$$

$$\lim_{t\to\infty} \beta^t \widetilde{K}_{t+1} \widetilde{C}_t^{-\gamma} = 0. \tag{57}$$

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

$$\eta = (1-\alpha)\frac{\widetilde{K}^{\alpha}L^{-\alpha}}{\widetilde{C}^{\gamma}} \tag{58}$$

$$1 = \exp(-\mu_z)\left(\widetilde{K}^{\alpha-1}L^{1-\alpha} - \frac{\widetilde{C}}{\widetilde{K}} + 1 - \delta\right) \tag{59}$$

$$1 = \beta\exp(-\mu_z)\left(1-\delta+\alpha\widetilde{K}^{\alpha-1}L^{1-\alpha}\right). \tag{60}$$

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

$$c_t = \alpha k_t - \alpha l_t \tag{61}$$

$$k_{t+1} = -\omega\varepsilon_{t+1} + \frac{\widetilde{K}^{\alpha}L^{1-\alpha}}{\widetilde{K}\exp(\mu_z)}(\alpha k_t + (1-\alpha)l_t) - \frac{\widetilde{C}}{\widetilde{K}\exp(\mu_z)}c_t + \frac{(1-\delta)\widetilde{K}}{\widetilde{K}\exp(\mu_z)}k_t \tag{62}$$

$$-\gamma c_t = E_t\left[-\gamma c_{t+1} + \frac{\alpha\widetilde{K}^{\alpha-1}L^{1-\alpha}}{1-\delta+\alpha\widetilde{K}^{\alpha-1}L^{1-\alpha}}((\alpha-1)k_{t+1} + (1-\alpha)l_{t+1})\right]; \tag{63}$$

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

$$c_t = \psi k_t \tag{64}$$

$$l_t = \phi k_t \tag{65}$$

for some coefficients (ψ, ϕ). We solve for these coefficients using the method of undetermined coefficients 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 non-stationary 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.

A.3 Out-of-Steady-State Filtering

As shown Section 2, in the steady state $\overline{\sigma}_k^2 = var_t[k_t]$ and $var_t[k_{t+1}] = \beta^{-2}\overline{\sigma}_k^2 + \omega^2$. Given the updating equation for the conditional variance,

$$var_{t+1}[k_{t+1}] = var_t[k_{t+1}]\left(var_t[k_{t+1}] + var_t\left[\xi_{k,t+1}\right]\right)^{-1} var_t\left[\xi_{k,t+1}\right], \tag{66}$$

we obtain that in the steady state

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

If the economy is *not* in the steady state of the filter, the information processing constraint

$$\kappa = \frac{1}{2}\left(\log\left(\omega^2 + \beta^{-2}\sigma_{k,t}^2\right) - \log\left(\sigma_{k,t+1}^2\right)\right),$$

implies that

$$\sigma_{k,t+1}^2 = \left(\omega^2 + \beta^{-2}\sigma_{k,t}^2\right)\exp\left(-2\kappa\right) \tag{67}$$

and (65) can be rewritten as

$$\sigma_{k,t+1}^2 = \left(\beta^{-2}\sigma_{k,t}^2 + \omega^2\right)\left(\beta^{-2}\sigma_{k,t}^2 + \omega^2 + var_t\left[\xi_{k,t+1}\right]\right)^{-1} var_t\left[\xi_{k,t+1}\right].$$

Therefore, the conditional variance of the noise in this case, $var_t\left[\xi_{k,t+1}\right]$, would be time-varying:

$$var_t\left[\xi_{k,t+1}\right] = \frac{\beta^{-2}\sigma_{k,t}^2 + \omega^2}{\exp\left(2\kappa\right) - 1}. \tag{68}$$

However, in the LQG setup the conditional variance $\sigma_{k,t+1}^2$ turns out to be deterministic and policy-independent; consequently, only the behavior of the conditional mean $\widehat{k}_t$ matters for aggregate dynamics.

In this non-steady state case, we assume that the initial state of the model economy is $k_0 | \mathcal{I}_0 \sim N\left(\widehat{k}_0, \sigma_{k,0}^2\right)$. The social planner's optimization problem can be written as

$$\widehat{v}\left(\widehat{k}_t\right) = \max_{\{c_t, l_t\}} E_t\left[u\left(c_t, l_t\right) + \widehat{v}\left(\widehat{k}_{t+1}\right)\right] \tag{69}$$

subject to (1) the resource constraint (19), (2) the Kalman filter equation governing the conditional mean $\widehat{k}_t$, (22), and (3) the transition equation governing the conditional variance $\sigma_{k,t}^2$

$$\sigma_{k,t+1}^2 = \lambda\sigma_{k,t}^2 + \exp\left(-2\kappa\right)\omega^2, \tag{70}$$

where $\lambda = \beta^{-2}\exp\left(-2\kappa\right) < 1$.

Note that given that given $\sigma_{k,0}^2$ and $\overline{\sigma}_k^2 = \frac{\omega^2}{\exp(2\kappa)-\beta^{-2}}$, (69) can be written as

$$\sigma_{k,t+1}^2 = \overline{\sigma}_k^2 + \lambda^t\left(\sigma_{k,0}^2 - \overline{\sigma}_k^2\right),$$

which shows how quickly the conditional variance approaches its steady state value. For example, if we set $\beta = 0.99$ and $\kappa = 0.2$ nats quarterly, $\lambda = 0.684$

and $\sigma_{k,t+1}^2$ takes less than 2 quarters to get halfway to its steady state value $\overline{\sigma}_k^2$. Therefore, in this paper we focus on the steady state conditional variance case.

A.4 Deriving the Variance-covariance Matrix of $\left(\widehat{k}, k\right)$

Taking unconditional variances on both sides of (23) and (24), we find that

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

where

$$\Sigma_k = \begin{bmatrix} var(k) & covar\left(k, \widehat{k}\right) \\ covar\left(k, \widehat{k}\right) & var\left(\widehat{k}\right) \end{bmatrix}.$$

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

$$var(k) = \frac{\omega^2}{1 - G^2}. \tag{71}$$

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

$$\begin{bmatrix} var(k) \\ covar\left(k, \widehat{k}\right) \\ var\left(\widehat{k}\right) \end{bmatrix} = \begin{bmatrix} 1 - \beta^{-2} & -2\beta^{-1}\left(G - \beta^{-1}\right) & -\beta^{-2}\left(G\beta - 1\right)^2 \\ -\theta - G\beta^{-1} & 1 - (1-\theta) G \left(G - \beta^{-1}\right) & 0 \\ \theta^2 - (1-\theta)^2 G^2 & -2\theta & 1 \end{bmatrix}^{-1} \begin{bmatrix} \omega^2 \\ 0 \\ \theta^2 var(\xi_k) \end{bmatrix}.$$

we can derive the solutions in closed-form, although they are not particularly intuitive.

A.5 Deriving the Coherence of $\left(k, \widehat{k}\right)$

The system that determines the evolution of the states, $\left(k, \widehat{k}\right)$, can be written as

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

where

$$\begin{aligned} \Psi &= \begin{bmatrix} \frac{1}{\beta} & G - \frac{1}{\beta} \\ \frac{\theta}{\beta} & \theta\left(G - \frac{1}{\beta}\right) + (1-\theta) G \end{bmatrix} \\ \zeta_t &= \begin{bmatrix} -\varepsilon_{t+1} \\ \theta \xi_{k,t+1} - \theta \varepsilon_{t+1} \end{bmatrix}. \end{aligned}$$

The covariance matrix $E_t\left[\zeta_t\zeta_t'\right]$ is then given by

$$\Omega = \begin{bmatrix} \omega^2 & \theta\omega^2 \\ \theta\omega^2 & \theta^2\omega^2 + \theta^2 var\left(\xi_k\right) \end{bmatrix}.$$

Writing out the autocovariance of the vector MA(∞) process yields

$$\Gamma_s = \sum_{h=0}^{\infty} \Psi_{s+h}\Omega\Psi_h'$$

for each s, with typical element $\gamma_{ij}\left(s\right)$.[17] Then the spectral density at frequency ν is

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

with cross spectrum

$$s_{12}\left(\nu\right) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\gamma_{12}\left(s\right)\exp\left(-s\nu i\right).$$

The coherence of two series at frequency ν is defined as

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

coherence measures the linear comovement between X and Y at a given frequency.

References

[1] Adam, Klaus (2005), "Optimal Monetary Policy in the Presence of Imperfect Common Knowledge," *Journal of Monetary Economics* **54(2)**, pp. 267-301.

[2] Batchuluun, Altantsetseg, Yulei Luo, and Eric R. Young (2007), "Portfolio Choice with Information-Processing Limits," manuscript, University of Virginia and Hong Kong University.

[17] The eigenvalues of Ψ 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(∞) process is stable, Ψ is also a stable matrix and therefore the infinite sum converges.

[3] Blanchard, Oliver J. and Danny Quah (1989), "The Dynamic Effects of Aggregate Supply and Demand Disturbances," *American Economic Review* **79(4)**, pp. 655-73.

[4] Boldrin, Michele, Lawrence J. Christiano, and Jonas D.M. Fisher (2001) "Habit Persistence, Asset Returns, and the Business Cycle," *American Economic Review* **91(1)**, pp. 149-66.

[5] Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo (1993), "Labor Hoarding and the Business Cycle," *Journal of Political Economy* **101(2)**, pp. 245-73.

[6] Campbell, John Y. (1994), "Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model," *Journal of Monetary Economics* **33(2)**, pp. 463-506.

[7] Christiano, Lawrence J. and Martin Eichenbaum (1992), "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations," *American Economic Review* **82(3)**, pp. 430-50.

[8] Cogley, Timothy and James M. Nason (1993), "Impulse Dynamics and Propagation Mechanisms in a Real Business Cycle Model," *Economics Letters* **43(1)**, pp. 77-81.

[9] Cogley, Timothy and James M. Nason (1995) "Output Dynamics in Real-Business-Cycle Models," *American Economic Review* **85(3)**, pp. 492-511.

[10] Cover, Thomas, and Joy Thomas (1991), *Elements of Information Theory*, New York: John Wiley and Sons.

[11] Galí, Jordi (1999), "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review* **89(2)**, pp. 249-71.

[12] Galí, Jordi and Pau Rabanal (2005), "Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data?" *NBER Macroeconomics Annual 2004*, 2005 MIT Press.

[13] Hansen, Gary D. (1985), "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics* **16(2)**, pp. 309-27.

[14] Kasa, Kenneth (1995), "Signal Extraction and the Propagation of Business Cycles," Federal Reserve Bank of San Francisco Working Papers in Applied Economic Theory No. **95-14**.

[15] King, Robert G., Charles I. Plosser, and Sergio Rebelo (1988) "Production, Growth, and Business Cycles: II. New Directions," *Journal of Monetary Economics* **21(2)**, pp. 309-41.

[16] King, Robert G. and Sergio Rebelo (1999), "Resuscitating Real Business Cycles," in Taylor, John and Michael Woodford eds., *Handbook of Macroeconomics* Volume 1B, pp. 928-1002.

[17] Kydland, Finn E. and Edward C. Prescott (1982), "Time to Build and Aggregate Fluctuations," *Econometrica* **50(6)**, pp. 1345-70.

[18] Landauer, Thomas K. (1986), "How Much Do People Remember? Some Estimates of the Amount of Learned Information in Long-Term Memory," *Cognitive Science* **10**, pp. 477-93.

[19] Lettau, Martin and Harold Uhlig (2000) "Can Habit Formation be Reconciled with Business Cycle Facts?" *Review of Economic Dynamics* **3(1)**, pp. 79-99.

[20] Lewis, Kurt F. (2006), "The Life-Cycle Effects of Information-Processing Constraints," mimeo, University of Iowa.

[21] Lindé, Jesper (2004), "The Effects of Permanent Technology Shocks on Labor Productivity and Hours in the RBC Model," Sveriges Riksbank Working Paper **161**.

[22] Lorenzoni, Guido (2006), "A Theory of Demand Shocks," Manuscript, MIT.

[23] Luo, Yulei (2007), "Rational Inattention, Portfolio Choice, and the Equity Premium," manuscript, University of Hong Kong.

[24] Luo, Yulei (2008), "Consumption Dynamics Under Information Processing Constraints," *Review of Economic Dynamics* **11(2)**, pp. 366-85.

[25] Luo, Yulei and Eric R. Young (2008), "Risk-sensitive Consumption and Savings under Rational Inattention," manuscript, University of Hong Kong and University of Virginia.

[26] Luo, Yulei and Qiang Zhang (2008), "The Current Account and Macroeconomic Adjustment under Rational Inattention," manuscript, University of Hong Kong and University of Leicester.

[27] Maćkowiak, Bartosz and Mirko Wiederholt (2008), "Optimal Sticky Prices under Rational Inattention," forthcoming, *American Economic Review*.

[28] Mankiw, N. Gregory and Ricardo Reis (2007), "Sticky Information in General Equilibrium," *Journal of the European Economic Association* **5(2-3)**, pp. 603-13.

[29] Moscarini, Giuseppe (2004), "Limited Information Capacity as a Source of Inertia," *Journal of Economic Control and Dynamics* **28(10)**, pp. 2003-35.

[30] Perli, Roberto and Plutarchos Sakellaris (1998), "Human Capital Formation and Business Cycle Persistence," *Journal of Monetary Economics* **42(1)**, pp. 67-92.

[31] Rotemberg, Julio J. and Michael Woodford (1996), "Real-Business-Cycle Models and the Forecastable Movements in Output, Hours, and Consumption," *American Economic Review* **86(1)**, pp. 71-89.

[32] Saito, Masashi (2005), "Shifting Expectations about Technology Growth as a Propagation Mechanism," manuscript, Boston University.

[33] Schmitt-Grohé, Stephanie (2000), "Endogenous Business Cycles and the Dynamics of Output, Hours, and Consumption," *American Economic Review* **90(5)**, pp. 1136-59.

[34] Shannon, Claude (1948), "A Mathematical Theory of Communication," *The Bell System Technical Journal* **27**, pp. 623-56.

[35] Sims, Christopher A. (1998), "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* **49**, pp. 317-56.

[36] Sims, Christopher A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics* **50(3)**, pp. 665-90.

[37] Sims, Christopher A. (2005), "Rational Inattention: A Research Agenda," Technical Report, Princeton University.

[38] Sims, Christopher A. (2006), "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review* **96(2)**, pp. 158-163.

[39] Tutino, Antonella (2007), "The Rigidity of Choice: Life Cycle Savings with Information-Processing Limits," manuscript, Princeton University.

[40] Van Nieuwerburgh, Stijn and Laura Veldkamp (2008), "Information Acquisition and Under-Diversification," manuscript, New York University.

[41] Woodford, Michael (2002) "Imperfect Common Knowledge and the Effects of Monetary Policy," Princeton University, in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds.