Optimal Foresight*

Ryan Chahrour
Boston College

Kyle Jurado
Duke University

February 4, 2021

Abstract

Agents have foresight when they receive information about a random process above and beyond the information contained in its current and past history. In this paper, we propose an information-theoretic measure of the quantity of foresight in an information structure, and show how to separate informational assumptions about foresight from physical assumptions about the dynamics of the processes itself. We then develop a theory of endogenous foresight in which the type of foresight is chosen optimally by economic agents. In a prototypical dynamic model of consumption and saving, we derive a closed-form solution to the optimal foresight problem.

JEL classification: D83, D84, E21

Keywords: Expectations, news, information choice

*Chahrour: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 (e-mail: `ryan.chahrour@bc.edu`); Jurado: Department of Economics, Duke University, 419 Chapel Drive, Durham, NC 27708 (e-mail: `kyle.jurado@duke.edu`).

1 Introduction

A popular hypothesis in macroeconomics and finance is that economic agents have foresight: they receive information about the future of a random process that is not revealed by its own past history. For example, the large and growing macroeconomic literature on "news," starting with Cochrane (1994) and Beaudry and Portier (2006), explicitly analyzes this hypothesis. In addition, but perhaps less obviously, many papers introduce foresight implicitly, by representing a structural driving process as the sum of several independent "components," each of which is separately observed by agents. A classic example is the Friedman and Kuznets (1945) type representation of income in terms of persistent and transitory components.

The first goal of this paper is to provide a measure of the quantity of foresight in an information structure. We suggest measuring foresight by the information-theoretic concept of conditional mutual information. According to this measure, the quantity of foresight that agents have about a random process is equal to the information agents have about the future history of that process above and beyond the information contained in its own current and past history alone. This measure provides a way to consistently compare seemingly different information structures based on the quantity of foresight they contain.

The second goal of this paper is to show how to determine the type of foresight implied by an information structure. This is useful for disentangling physical assumptions about the law of motion of structural processes from informational assumptions about foresight. In cases when foresight is introduced implicitly, through independent-component representations, any change in the law of motion of one component changes both the type of foresight in the information structure, as well as the implied dynamics of the structural processes. This makes it difficult to perform comparative statics exercises, where one set of assumptions is altered holding fixed the other. It also makes it difficult to independently validate these two sets of assumptions. Our approach is to make foresight *explicit*, by representing foresight as coming exclusively from noise-ridden signals about the future of the structural process.

The third goal of this paper is to develop a theory of endogenous foresight in which forward-looking agents optimally choose what information to process about the future. Of course, the first best would be to choose to have perfect foresight, but the agent faces various cognitive limitations which make it infeasible for him to achieve this. We model these limitations as a constraint on the total quantity of foresight the agent can receive, but allow him to otherwise choose his information structure optimally.

By way of results, we provide three propositions, each addressing one of the objectives described above. The first presents a closed-form expression for the quantity of foresight in a popular persistent-transitory representation of a random process in terms of the underlying parameters. The second presents the type of foresight implied by this representation in the form of a noise-ridden signal of the future values of the process. The third presents a closed-form solution to the optimal foresight problem in a prototypical dynamic optimizing model of consumption and saving. All of these results can be generalized for use in other settings, which is ongoing work.

Related literature

Our approach to constrained information choice is related to the approach used in the rational inattention literature initiated by Sims (1998), but it is distinct in several respects.[1] First, this literature and the subsequent literature on endogenous information choice imposes a "no foresight constraint," which prevents agents from having any foresight about the structural disturbances in the model. This constraint was first introduced by Sims (2003), who suggested that it would be unrealistic to allow agents to condition their information on future disturbances.[2] By contrast, we allow agents to have foresight regarding structural disturbances. Second, this literature assumes that it is costly to process information about past and present structural disturbances. By contrast, in this paper we assume that agents can costlessly process information about current and past disturbances, and only face costs in processing information about future disturbances.

One way to partially circumvent the no foresight constraint in the rational inattention literature has been to introduce foresight implicitly, using independent-component representations. This approach allows agents to have some foresight regarding the sum of the components, even if they have no foresight regarding each component separately. Examples of this type of implicit foresight in the rational inattention literature include Luo (2008) and the related models in Section 6 of Sims

[1]See Veldkamp (2011) for a broad introduction to theories of endogenous information choice and Maćkowiak et al. (2018b) for a recent survey of the rational inattention literature.

[2]Cf. Sims 2003, p. 672 where he describes the "more realistic situation."

(2003) and Section 5 of Miao et al. (2020), which allow consumers to inform themselves about different independent components of their income process. In all these cases, however, the type of foresight is still constrained by the exogenously specified independent-component representation of the fundamental process. In this paper, we allow that representation to be determined endogenously by agents' optimal information choice.

There are three papers in the rational inattention literature that are more explicit about introducing foresight. The first is Gaballo (2016), who presents an overlapping generations equilibrium model in which agents receive a noise-ridden private signal about next period's average price level. The main difference is that information is not endogenously chosen by agents in that model; the signal structure is determined exogenously. However, in Appendix (B), we use the theory of foresight that we develop here to formally prove that the exogenous signal structure also happens to be *optimal*. This result provides an information-theoretic justification for the particular signal structure chosen in that paper.

The second paper is Maćkowiak et al. (2018a). In Section 7 of the paper, they formulate a business-cycle model with rational inattention and news. In this model, technological disturbances are assumed to affect the level of technology with a delay of k periods. This means that the history of technological disturbances contains more information than the history of technology. The main differences with this paper are that (i) agents are not able to costlessly observe current and past technology, and (ii) agents are restricted to only have foresight about technology k periods into the future. By contrast, in this paper agents have perfect hindsight, and are not restricted in the type of foresight they can have, only in the quantity.

The third paper is Jurado (2020), which provides a frequency-domain solution to the canonical dynamic rational inattention problem proposed by Sims (2003) and analyzed in the time domain by Maćkowiak et al. (2018b). That paper also highlights the importance of the no foresight constraint in the rational inattention literature and analyzes how removing this constraint affects the problem. A key result is that without the no foresight constraint, agents construct an endogenous band-pass filter, and only pay attention to those frequencies that contribute most to the variation in their target. The main difference with this paper is that Jurado (2020) follows the rest of the rational inattention literature in assuming that information about the past is costly to process.

2 Defining foresight

This section defines what we mean by foresight and also provides a measure of the quantity of foresight that an information structure contains.

Foresight refers to information about the future history of a process beyond what is contained in its own current and past history. A natural way to quantify this is by conditional mutual information, which measures the reduction in uncertainty, as measured by entropy, that foresight provides. More specifically, let $\mathcal{I}_t$ denote an arbitrary time-t information set, which is assumed to be generated by the current and past history of a collection of stationary random processes (potentially infinitely many), and consider a random process $\{y_t\}$ which is stationary and stationarily related to this collection. Then the quantity of foresight in the information structure $\{\mathcal{I}_t\}$ regarding the process $\{y_t\}$ can be defined as follows.

Definition 1. The quantity of foresight regarding $\{y_t\}$ contained in the information structure $\{\mathcal{I}_t\}$ is

$$\lim_{T\to\infty} I((y_{t+1},\dots,y_{t+T}),\mathcal{I}_t|y^t).$$

where $I((y_{t+1},\dots,y_{t+T}),\mathcal{I}_t|y^t)$ denotes the conditional mutual information between $(y_{t+1},\dots,y_{t+T})$ and $\mathcal{I}_t$, conditional on $y^t = (y_t, y_{t-1},\dots)$.

To understand this definition, first note that conditional mutual information can be expressed in terms of (differential) entropy as follows[3]

$$I((y_{t+1},\dots,y_{t+T}),\mathcal{I}_t|y^t) = H((y_{t+1},\dots,y_{t+T})|y^t) - H((y_{t+1},\dots,y_{t+T})|\mathcal{I}_t, y^t).$$

The first term denotes the conditional entropy of $(y_1,\dots,y_T)$ with no foresight; it is the degree of uncertainty about the process $\{y_t\}$ over the next T periods in the future (as of time t) that remains after conditioning on its own current and past history. The second term is the conditional entropy with foresight. The difference represents the degree to which $\mathcal{I}_t$ reduces uncertainty about the future relative to only knowing y^t. By taking limits as $T\to\infty$, the definition represents the reduction in uncertainty about the *entire future history* of the process $\{y_t\}$ that is provided by observing $\mathcal{I}_t$. The assumption of stationarity ensures that this definition is the same regardless of which time period is treated as the present; e.g.

$$\lim_{T\to\infty} I((y_{t+1},\dots,y_{t+T}),\mathcal{I}_t|y^t) = \lim_{T\to\infty} I((y_1,\dots,y_T),\mathcal{I}_0|y^0).$$

[3]Cover and Thomas (2006) is a standard reference on conditional mutual information and entropy.

It is also worth noting that foresight is a *directed* measure of information flow. To understand this, suppose that $\mathcal{I}_t$ is generated by the current and past values of the stationary process $\{x_t\}$. The fact that foresight is directed means that the roles of $\{y_t\}$ and $\{x_t\}$ cannot be reversed; i.e.

$$\lim_{T\to\infty} I((y_{t+1},\ldots,y_{t+T}),x^t|y^t) \neq \lim_{T\to\infty} I((x_{t+1},\ldots,x_{t+T}),y^t|x^t).$$

This is unlike the average rate of information flow between $\{x_t\}$ and $\{y_t\}$, which is undirected (i.e. symmetric).

When all processes are Gaussian, as we will maintain throughout this paper, conditional mutual information can be expressed in terms of the covariance matrices of forecast errors with and without foresight

$$I((y_{t+1},\ldots,y_{t+T}),\mathcal{I}_t|y^t) = -\frac{1}{2}\ln\frac{\det\hat{\Sigma}_T}{\det\Sigma_T}, \tag{1}$$

where $\hat{\Sigma}_T$ is the covariance matrix with foresight, and Σ_T is the covariance matrix without foresight,

$$\begin{aligned}\hat{\Sigma}_T &\equiv \mathrm{var}((y_{t+1},\ldots,y_{t+T}) - E[(y_{t+1},\ldots,y_{t+T})|\mathcal{I}_t,y^t])\\ \Sigma_T &\equiv \mathrm{var}((y_{t+1},\ldots,y_{t+T}) - E[(y_{t+1},\ldots,y_{t+T})|y^t]).\end{aligned}$$

Using information-theoretic measures like conditional mutual information to quantify information transmission is familiar from the economic literature on rational inattention. In that literature, agents choose their information structures (i.e. what they pay attention to) subject to a constraint on the rate of information flow. As we discuss in the introduction, one important difference with respect to what we do is that in rational inattention models, agents are not allowed to choose information structures that contain any amount of foresight about the underlying structural disturbances. If we call these disturbances $\{\varepsilon_t\}$, then this requirement can be expressed as

$$\lim_{T\to\infty} I((\varepsilon_{t+1},\ldots,\varepsilon_{t+T}),\mathcal{I}_t|\varepsilon^t) = 0 \tag{2}$$

for all possible information structures $\{\mathcal{I}_t\}$.[4]

[4] See Jurado (2020) for discussion of the relationship between this way of articulating the no foresight constraint and the way it is more commonly articulated in the rational inattention literature.

3 Computing foresight

A common way to introduce foresight into economic models is by representing a random process as the sum independent components, each of which is separately observed by an economic agent. Perhaps the most popular of such representations is the persistent-transitory representation. This section derives closed-form expressions for the quantity of foresight in this representation. Because it may not always be feasible to obtain closed-form expressions for the amount of foresight in an information structure, we also present an algorithm that can be used across a wide variety of information structures in Appendix (C).

Let $\{y_t\}$ denote a stationary process; for the sake of concreteness we refer to it as income. The persistent-transitory representation decomposes income into the sum of two independent components,

$$y_t = z_t + \sigma_u u_t \qquad z_t = \rho z_{t-1} + \sigma_\eta \eta_t, \tag{3}$$

where $\sigma_u, \sigma_\eta > 0$, $0 < \rho < 1$, and $\{u_t\}$ and $\{\eta_t\}$ are independent orthonormal Gaussian white noise processes. The persistent component is z_t and the transitory component is $\sigma_u u_t$. At each point in time, the agent's information set is equal to the closed linear space spanned by the current and past history of disturbances, $\mathcal{I}_t = \text{span}(\eta^t, u^t)$.

To compute the quantity of foresight regarding the income process that is contained in the information structure $\{\mathcal{I}_t\}$, we can use equation (1). First, note that the j-step-ahead forecast error in income according to the persistent-transitory representation (3) is

$$\hat{e}_{t+j|t} = y_{t+j} - E[y_{t+j}|\mathcal{I}_t] = \sigma_u u_{t+j} + \sum_{k=1}^{j} \rho^{j-k} \sigma_\eta \eta_{t+k}.$$

Stacking these up for $j = 1, \dots, T$,

$$\begin{bmatrix} \hat{e}_{t+1|t} \\ \hat{e}_{t+2|t} \\ \hat{e}_{t+3|t} \\ \vdots \\ \hat{e}_{t+T|t} \end{bmatrix} = \underbrace{\sigma_u I_T}_{Q_u} \begin{bmatrix} u_{t+1} \\ u_{t+2} \\ u_{t+3} \\ \vdots \\ u_{t+T} \end{bmatrix} + \underbrace{\begin{bmatrix} \sigma_\eta & 0 & 0 & \cdots & 0 \\ \rho\sigma_\eta & \sigma_\eta & 0 & \cdots & 0 \\ \rho^2\sigma_\eta & \rho\sigma_\eta & \sigma_\eta & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma_\eta & \rho^{T-2}\sigma_\eta & \rho^{T-3}\sigma_\eta & \cdots & \sigma_\eta \end{bmatrix}}_{Q_\eta} \begin{bmatrix} \eta_{t+1} \\ \eta_{t+2} \\ \eta_{t+3} \\ \vdots \\ \eta_{t+T} \end{bmatrix}.$$

From this we can see that $\hat{\Sigma}_T = Q_u Q_u' + Q_\eta Q_\eta'$. By exploiting the structure of Q_u and Q_η, it is possible to show that[5]

$$\det \hat{\Sigma}_T = \left[(1 - r^2) \left(\frac{\rho}{\theta} \right)^T + r^2 (\rho\theta)^T \right] \sigma_u^{2T}, \tag{4}$$

where the coefficient $0 \leq r^2 \leq 1$ is given by $r^2 \equiv (\rho - \theta(1 + \sigma_\eta^2/\sigma_u^2))/(\rho(1 - \theta^2))$.

Next, we need to find an expression for $\det \Sigma_T$ in equation (1). This requires us to forecast income only on the basis of its own past history. The key step at this point is to find the Wold representation of the income process. In this case, the Wold representation is[6]

$$y_t = \rho y_{t-1} + \sigma_\varepsilon (\varepsilon_t - \theta \varepsilon_{t-1}), \tag{5}$$

where $\{\varepsilon_t\}$ is an orthonormal Gaussian white noise process with the special property that $\text{span}(\varepsilon^t) = \text{span}(y^t)$ for all t. In this representation, $\sigma_\varepsilon^2 = \sigma_u^2 \rho/\theta$, and θ is the root of the polynomial $\mathcal{P}(\theta) = \rho\theta^2 - (1 + \rho^2 + \sigma_\eta^2/\sigma_u^2)\theta + \rho$ that lies inside the unit circle. This is the familiar result from time series analysis that the sum of an AR(1) process and white noise is an ARMA(1,1) process (e.g. Hamilton, 1994, ch. 4).

The j-step-ahead forecast error in income according to the Wold representation (5) is

$$e_{t+j|t} = y_{t+j} - E[y_{t+j}|y^t] = \sigma_\varepsilon \varepsilon_{t+j} + (\rho - \theta) \sum_{k=1}^{j-1} \rho^{j-1-k} \sigma_\varepsilon \varepsilon_{t+k}.$$

Stacking these up for $j = 1, \ldots, T$,

$$\begin{bmatrix} e_{t+1|t} \\ e_{t+2|t} \\ e_{t+3|t} \\ \vdots \\ e_{t+T|t} \end{bmatrix} = \underbrace{\begin{bmatrix} \sigma_\varepsilon & 0 & 0 & \cdots & 0 \\ (\rho - \theta)\sigma_\varepsilon & \sigma_\varepsilon & 0 & \cdots & 0 \\ \rho(\rho - \theta)\sigma_\varepsilon & (\rho - \theta)\sigma_\varepsilon & \sigma_\varepsilon & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-2}(\rho - \theta)\sigma_\varepsilon & \rho^{T-3}(\rho - \theta)\sigma_\varepsilon & \rho^{T-4}(\rho - \theta)\sigma_\varepsilon & \cdots & \sigma_\varepsilon \end{bmatrix}}_{Q_\varepsilon} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+2} \\ \varepsilon_{t+3} \\ \vdots \\ \varepsilon_{t+T} \end{bmatrix}.$$

From this it is easy to see that

$$\det \Sigma_T = \sigma_\varepsilon^{2T}, \tag{6}$$

[5] See the proof of Proposition (1) in Appendix (A).

[6] The details are provided in the proof of Proposition (2) in Appendix (A).

since $\Sigma_T = Q_\varepsilon Q'_\varepsilon$, and Q_ε is lower triangular, so its determinant is the product of its diagonal elements.

Now that we have computed the determinants of the forecast error covariance matrices with and without foresight, we can use these expressions to compute the conditional mutual information about $(y_{t+1}, \ldots, y_{t+T})$. Plugging (4) and (6) into (1) and using the fact that $\sigma_\varepsilon^2 = \sigma_u^2 \rho/\theta$, we find that

$$I((y_{t+1}, \ldots, y_{t+T}), (y^t, z^t)|y^t) = -\frac{1}{2} \ln \left[(1 - r^2) + r^2\theta^{2T}\right],$$

Since $|\theta| < 1$, we can see that the second term vanishes as $T \to \infty$, which means that we have arrived at the following result.

Proposition 1. *The quantity of foresight in the information structure from the persistent-transitory representation* (3) *is*

$$\lim_{T\to\infty} I((y_{t+1}, \ldots, y_{t+T}), (\eta^t, u^t)|y^t) = -\frac{1}{2} \ln(1 - r^2),$$

where $0 \leq r^2 \leq 1$ *is given by*

$$r^2 \equiv \frac{\rho - \theta(1 + \sigma_\eta^2/\sigma_u^2)}{\rho(1 - \theta^2)}$$

and $0 < \theta < 1$ *is given by*

$$\theta \equiv \frac{1}{2\rho}\left(1 + \rho^2 + \sigma_\eta^2/\sigma_u^2 - \sqrt{(1 + \rho^2 + \sigma_\eta^2/\sigma_u^2)^2 - 4\rho^2}\right).$$

Notice that the quantity of foresight in this representation depends only on ρ and the ratio σ_η^2/σ_u^2. The following corollary summarizes the way that foresight depends on these parameters.

Corollary 1. *Let* F *denote the quantity of foresight in the persistent-transitory representation* (3)*. Then*

(i) F *is monotonically increasing in* ρ *with limiting values* $\lim_{\rho\to 0} F = 0$ *and* $\lim_{\rho\to 1} F = -\frac{1}{2}\ln(1 - \bar{r}^2)$, *where*

$$\bar{r}^2 \equiv \frac{1 - \bar{\theta}(1 + \sigma_\eta^2/\sigma_u^2)}{1 - \bar{\theta}^2}$$

and

$$\bar{\theta} \equiv \frac{1}{2}\left(2 + \sigma_\eta^2/\sigma_u^2 - \sqrt{\left(2 + \sigma_\eta^2/\sigma_u^2\right)^2 - 4}\right)$$

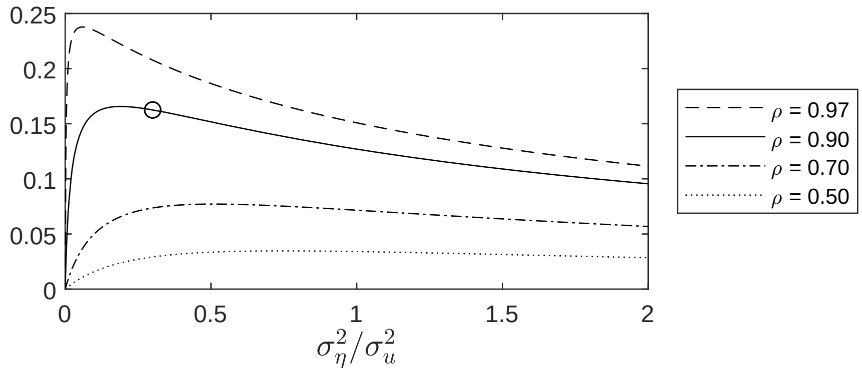

Figure 1: Quantity of foresight in the persistent-transitory representation. The circle shows the amount of information at the baseline parameter values $\rho = 0.9$, $\sigma_u^2 = 0.01$, and $\sigma_\eta^2 = 0.003$.

(ii) F has limiting values $\lim_{\sigma_\eta^2/\sigma_u^2 \to 0} F = 0$ *and* $\lim_{\sigma_\eta^2/\sigma_u^2 \to \infty} F = 0$.

We can illustrate these results in a numerical example. We select baseline parameter values consistent with the numerical exercise in Section 6 of Sims (2003),[7]

$$\rho = 0.9, \quad \sigma_u^2 = 0.01, \quad \text{and} \quad \sigma_\eta^2 = 0.003.$$

At these values, the expression in Proposition (1) implies that the quantity of foresight is 0.16 nats (0.23 bits). How "big" is this? While these units do have familiar interpretations from computer processing, we must be careful not to rely too heavily on those intuitions when it comes to agents in an economic model. This is because economic models are drastically simplified versions of reality; small amounts of information in an economic model can correspond to large amounts of information in the actual world.

Figure (1) shows the effects of changing the parameters on the total quantity of foresight. Each line shows the effects as the ratio σ_η^2/σ_u^2 varies over the range of values on the horizontal axis. The different lines show these effects for different values of ρ. From the figure, we can see that regardless of the value of σ_η^2/σ_u^2, the quantity of foresight is monotonically increasing in ρ. Moreover, the quantity of foresight approaches zero as σ_η^2/σ_u^2 approaches either 0 or infinity for any value of ρ.

[7]In that exercise, the income process has two persistent components. We eliminate one and keep all other parameters unchanged.

4 Explicit foresight

As in the example from the previous section, foresight is typically introduced into economic models by representing a structural process (e.g. income, technology, dividends) as the sum of independent components, which are all separately observable by agents in the model. A difficulty with this approach is that foresight is introduced only *implicitly*; each of the independent components affects both the law of motion of the structural process and the type of foresight agents have. This makes it difficult to perform comparative statics exercises, such as altering informational assumptions about the quantity of foresight, holding fixed physical assumptions about the structural process. It also makes it difficult to validate these assumptions independently. This section describes how to disentangle these two sets of assumptions, by representing foresight *explicitly* in terms of subjective signals about the structural process.

To alter informational assumptions regarding foresight without changing the physical assumptions regarding the structural process, it is necessary to *hold the Wold representation of the structural process fixed*. The most straightforward way to do this is to construct an equivalent representation of the agent's information structure in which the Wold innovations of the structural process appear in the set of structural disturbances. To illustrate, consider the persistent-transitory representation (3). First, write income in terms of its Wold innovations as in equation (5). This isolates the physical assumptions that the persistent-transitory representation makes regarding the income process. Second, construct a set of subjective signals that isolate the type of foresight agents have regarding the income process. In this case, it is possible to show that $\text{span}(\eta^t, u^t) = \text{span}(y^t, s^t)$, where[8]

$$s_t = (1-\theta)\sum_{j=0}^{\infty}\theta^j y_{t+j} + \sigma_v v_t, \tag{7}$$

$\sigma_v = (1-\theta)\sigma_\varepsilon\sigma_u/\sigma_\eta$, and $\{v_t\}$ is an orthonormal Gaussian white noise process which is independent of $\{y_t\}$. In other words, separately observing the history of the persistent and transitory components is equivalent to observing the history of income and a noisy signal of an exponential average of current and future income. The noise in this signal captures purely expectational disturbances; these are disturbances that do not affect income but they do affect the agent's expectations.

[8]See the proof of Proposition (2) in Appendix (A).

By separating representation (3) into a physical law of motion for income (5) and a subjective signal (7), it becomes possible to analyze these physical and informational assumptions separately. For example, to analyze the effect of increasing agents' information about income far out into the future, we could replace the law of motion for the signal $\{s_t\}$ in (7) with a different signal that places more weight on future income,

$$\tilde{s}_t = (1-\tilde{\theta})\sum_{j=0}^{\infty}\tilde{\theta}^j y_{t+j} + \sigma_v v_t,$$

with $\tilde{\theta} > \theta$ (and σ_v is defined as before). The change in information from $\mathcal{I}_t = \text{span}(y^t, s^t)$ to $\tilde{\mathcal{I}}_t = \text{span}(y^t, \tilde{s}^t)$ alters the quantity of foresight the agent has regarding income, but does not alter the dynamics of income itself, which is held fixed at (5).

We summarize this discussion with a proposition. It is not difficult to see that this result can be extended to apply to representations other than the persistent-transitory representation in (3). The general recipe is: (i) derive the Wold representation of the structural process using standard results from time series analysis, and (ii) create a set of subjective signals that generates the same information structure by projecting any other variables observed by agents onto the space spanned by all past, present, and future values of the structural process.

Proposition 2. *Consider the persistent-transitory representation* (3)*, with information structure* $\{\mathcal{I}_t\}$*,* $\mathcal{I}_t \equiv span(\eta^t, u^t)$*. Then* $\mathcal{I}_t = span(y^t, s^t)$ *when*

$$y_t = \rho y_{t-1} + \sigma_\varepsilon(\varepsilon_t - \theta\varepsilon_{t-1})$$
$$s_t = (1-\theta)\sum_{j=0}^{\infty}\theta^j y_{t+j} + \sigma_v v_t,$$

where $\{\varepsilon_t\}$ *and* $\{v_t\}$ *are independent orthonormal Gaussian white noise processes,* $\sigma_\varepsilon \equiv \sigma_u\sqrt{\rho/\theta}$*,* $\sigma_v \equiv (1-\theta)\sigma_\varepsilon\sigma_u/\sigma_\eta$*, and* θ *is defined as in Proposition* (1)*.*

This proposition reveals that in the persistent-transitory representation (3), it is the implicit parameter θ which controls the magnitude of the signal weights on future values of $\{y_t\}$. From the expression in the Proposition, we can see that θ depends only on ρ and the ratio σ_η^2/σ_u^2. We summarize its dependence on these parameters in a corollary.

Corollary 2. *The parameter* θ *from Proposition* (2) *has the following properties.*

(i) θ *is monotonically increasing in* ρ *with limiting values* $\lim_{\rho\to 0}\theta = 0$ *and* $\lim_{\rho\to 1}\theta = \bar{\theta}$*, where* $\bar{\theta}$ *is defined as in Corollary* (1)*.*

(ii) θ *is monotonically decreasing in* σ_η^2/σ_u^2 *with limiting values* $\lim_{\sigma_\eta^2/\sigma_u^2\to 0}\theta = \rho$ *and* $\lim_{\sigma_\eta^2/\sigma_u^2\to\infty}\theta = 0$*.*

The intuition behind part (i) of the corollary is that as ρ increases, the persistent component becomes more informative about income farther out into the future. This corresponds to an increase in the weights of the signal s_t on future income. As ρ approaches zero, income becomes white noise and the signal contains no information about the future. The intuition behind part (ii) of the corollary is that the value of the ratio σ_η^2/σ_u^2 determines how much of the variation in income is driven by the persistent component relative to the transitory component. When it is arbitrarily small, income becomes white noise; when it is arbitrarily large, income becomes an AR(1) process. In either case, the implied signal s_t becomes completely uninformative. In the first case, θ is large but the signal is uninformative because it becomes infinitely noisy, $\sigma_v^2/\sigma_\varepsilon^2 = \sigma_u^2/\sigma_\eta^2 \to \infty$. In the second case, the signal is uninformative because it places no weight on future income, $\theta \to 0$.

These results can be visualized in a numerical example. Using the same baseline parameter values used to construct Figure (1), the expression in Proposition (1) implies that $\theta = 0.56$. Figure (2) illustrates the effects of changing the parameters on the magnitude of θ. Each line shows what happens as the ratio σ_η^2/σ_u^2 varies over the range of values on the horizontal axis. The different lines show these effects for different values of ρ. From the figure, we can see both how θ is monotonically increasing in ρ and monotonically decreasing in σ_η^2/σ_u^2. We can also see how the point at which each line crosses the vertical axis is equal to the corresponding value of ρ.

5 Optimal foresight

Having discussed what foresight is and how to measure it, we now ask: what type of foresight would be optimal from an agent's perspective? We answer this question in the context of a simple model of consumption and saving, in which we can provide a closed-form characterization of optimal foresight. This example is useful because it is a standard reference point for models of information choice and shares many features with a wider class of dynamic optimizing models with forward-looking behavior.

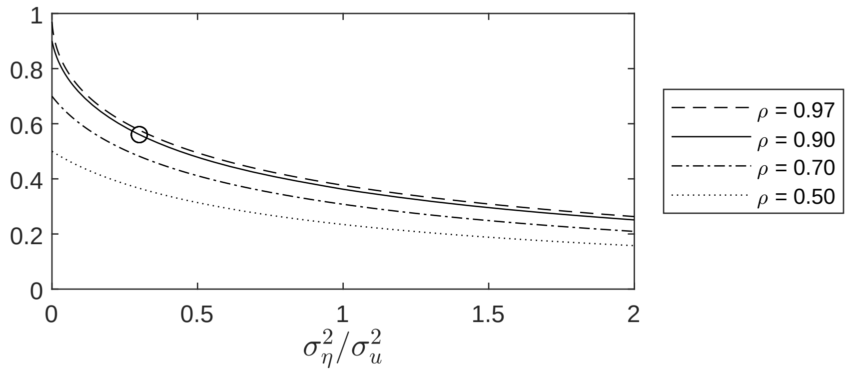

Figure 2: Value of the discounting parameter θ implied by the persistent-transitory representation. The circle shows the value of θ at the baseline parameter values $\rho = 0.9$, $\sigma_u^2 = 0.01$, and $\sigma_\eta^2 = 0.003$.

This section has two parts. The first derives a closed-form expression for the agent's forecasting rule with optimal foresight, and illustrates how foresight affects the responses of endogenous variables to the structural disturbances. The second compares the model's predictions under optimal foresight with its predictions under exogenous foresight, in the form of the persistent-transitory representation (3).

5.1 Consumption with endogenous foresight

In the model, a consumer seeks to maximize expected lifetime utility

$$E \sum_{t=0}^{\infty} \beta^t u(C_t),$$

where C_t is consumption, $0 < \beta < 1$ is a subjective time discount factor, u is an increasing, concave period utility function. Each period, consumption C_t and savings B_t are subject to the dynamic budget constraint

$$C_t + B_t \leq (1 + r_{t-1})B_{t-1} + Y_t,$$

where Y_t is (random) labor income and r_t is the interest rate at which the consumer can borrow at time t. The consumer is also prevented from engaging in Ponzi schemes by the constraint $\lim_{t\to\infty} \prod_{s=0}^{t-1}(1 + r_s)^{-1} B_t \geq 0$.

The interest rate faced by the consumer is allowed to depend on his current level of savings, $r_t = r + \Phi(B_t)$, where $r > 0$ is a constant, and Φ is a strictly decreasing function. This function represents a savings-elastic risk premium faced by the consumer; higher levels of savings (lower levels of debt) are associated with lower interest rates. From a theoretical perspective, we can think of this as a simple way of introducing financial frictions in the market for consumer debt. We introduce this assumption primarily for technical convenience, because it ensures that the optimal consumption process will be stationary, and it will allow us to nest non-stationarity as a limit case.

We limit ourselves to characterizing the optimal dynamics of this model only in response to random disturbances that generate sufficiently small fluctuations around the deterministic steady state. To do so, we first construct a linear-quadratic (LQ) analogue to the exact nonlinear problem, which has two special properties. First, its optimality conditions are the same ones that result from performing a first-order Taylor approximation to the exact nonlinear optimality conditions, as is commonly done in the literature. Second, the objective function is *purely quadratic* in the endogenous variables (it contains no linear terms), which means that the first-order accurate optimality conditions are sufficient for computing a second-order accurate approximation to the lifetime utility of the consumer.[9] This second property is important for formally articulating the consumer's information problem.

Our approximation is performed in terms of the log deviation of consumption and income from their deterministic steady-state values, $c_t \equiv \ln(C_t/C)$ and $y_t \equiv \ln(Y_t/Y)$, and the deviation of savings from its steady-state value, $b_t \equiv B_t - B$. It is also convenient at this point to introduce the definitions $\sigma \equiv -u'(C)/(u''(C)C)$ and $\phi \equiv -2\Phi'(B) - \Phi''(B)B$, where $\sigma > 0$ is the consumer's intertemporal elasticity of substitution at the steady state, and $\phi > 0$ controls the elasticity of the interest rate with respect to savings at the steady state. Using this notation, we can state the appropriate LQ problem.

Lemma 1. *A purely quadratic LQ approximation to the nonlinear problem is one in which the consumer seeks to maximize the quadratic objective*

$$-\frac{1}{2}E\sum_{t=0}^{\infty}\beta^t\left\{c_t^2+\frac{\beta\sigma\phi}{C}b_t^2+\left(\frac{\beta-\delta}{\beta-\delta^2}\right)\frac{Y^2}{C^2}\left((1-\delta)y_t^2-2y_tx_t\right)\right\} \tag{8}$$

subject to the linear constraint $Cc_t + b_t = \beta^{-1}b_{t-1} + Yy_t$*, where* x_t *is an exponential*

[9]This type of LQ problem is the "correct LQ local approximation" in the language of Benigno and Woodford (2012); see that paper for more discussion on the importance of the second property.

average of current and future labor income,

$$x_t \equiv (1-\delta) \sum_{j=0}^{\infty} \delta^j y_{t+j}, \tag{9}$$

and the discounting parameter $0 < \delta < \beta$ *is*

$$\delta \equiv \frac{1}{2} \left(1 + \beta + \sigma\phi\beta^2 C - \sqrt{(1 + \beta + \sigma\phi\beta^2 C)^2 - 4\beta} \right).$$

Under this LQ formulation, certainty equivalence implies that the consumer's consumption and saving decisions can be decoupled from his choices regarding information.[10] Conditional on his information, the consumer's policy function takes the familiar permanent-income form[11]

$$c_t = \frac{1}{C} \left[(1-\delta)\beta^{-1} b_{t-1} + Y E_t[x_t] \right]. \tag{10}$$

The term $\beta^{-1} b_{t-1}$ is the total financial wealth the consumer has available for consumption at time t, and the second term is his optimal estimate of average current and future labor income. Based on the expression for x_t in (9), we can see that the consumer endogenously discounts future income at rate δ, which is less than β, due to the fact that interest rates are savings-elastic. In the limit as $\phi \to 0$, it is possible to show that $\delta \to \beta$. The scaling terms C and Y appear because we are approximating consumption and income in logs rather than levels.

The policy function (10) helps to clarify how the approach taken in this paper differs from the existing rational inattention literature. In that literature, the consumer is both uncertain about his current savings b_{t-1} and the average of his current and future labor income x_t. Therefore, up to the same level of approximation, the policy function of such an agent would depend only on his best estimates of *both* of these variables,

$$c_t = \frac{1}{C} \left[(1-\delta)\beta^{-1} E_t[b_{t-1}] + Y E_t[x_t] \right].$$

By contrast, we assume that the consumer perfectly knows his current and past income, and therefore his current savings, so $E_t[b_{t-1}] = b_{t-1}$. The relevant margin of uncertainty for him is not the past or present, but the future. He remembers his past

[10]This well-known result is originally due to Simon (1956) and Theil (1957); see Whittle (1983) for a somewhat more recent discussion.

[11]The intermediate steps are presented in Appendix (A).

income and freely observes his current income and savings account balance, but finds it costly to obtain *additional* information about his future income.

What remains is to specify the consumer's information choice problem. To do so, we first show that it is possible to rewrite the consumer's utility maximization problem as a tracking problem in terms of the target variable x_t.

Lemma 2. *Maximizing the quadratic objective* (8) *is equivalent to minimizing the loss function*

$$E\sum_{t=0}^{\infty}\beta^t(x_t - E_t[x_t])^2.$$

This result, which is something of a folk theorem in the information literature, has a straightforward interpretation: the consumer would like to choose an information structure that makes the discounted sum of errors in his estimate of average lifetime income as small as possible.[12] Due to the quadratic form of the objective in (8), the magnitude of these errors is judged on a mean-square basis. A notable feature of our proof of this result is that it does not invoke any assumptions regarding the law of motion of the exogenous labor income process, beyond that it is appropriately bounded. For example, we do not require it to be Markovian.

In minimizing the objective in Lemma (2), we allow the consumer to select his time-t information set $\mathcal{I}_t \supseteq \text{span}(y^t)$ subject to the constraint that the quantity of foresight it contains cannot exceed a finite amount $\kappa > 0$,

$$\lim_{T\to\infty} I((y_{t+1},\ldots,y_{t+T}),\mathcal{I}_t|y^t) \leq \kappa.$$

We close the model by allowing the consumer's income process to have arbitrary stationary dynamics, given by the Wold representation

$$y_t = h(L)\varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1). \tag{11}$$

To solve the consumer's information problem, it is helpful to rewrite the problem as one in which the consumer directly chooses an optimal forecast process rather than an optimal information set. Letting $\hat{x}_t \equiv E_t[x_t]$, we can set $\mathcal{I}_t = \text{span}(y^t, \hat{x}^t)$. This is without loss of generality because any information that is not contained in the current and past history of forecasts has no effect on the objective but is costly in terms of

[12]Versions of this result are appealed to by, among others, Woodford (2003) and Azeredo da Silveira and Woodford (2019).

the constraint. Using this observation, we can re-write the consumer's information problem as:

$$\min_{\{\hat{x}_t\}} E\sum_{t=0}^{\infty}\beta^t(x_t-\hat{x}_t)^2 \qquad \text{subject to} \tag{12}$$

(i) $\lim_{T\to\infty}((y_{t+1},\ldots,y_{t+T}),\hat{x}^t|y^t)\leq\kappa$

(ii) $E[(x_t-\hat{x}_t)\hat{x}_{t-j}]=0$ for all $j\geq 0$

(iii) $E[(x_t-\hat{x}_t)y_{t-j}]=0$ for all $j\geq 0$,

where $x_t=(1-\delta)(1-\delta L^{-1})^{-1}h(L)\varepsilon_t$, with $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$. The first constraint is the foresight constraint, after imposing $\mathcal{I}_t=\text{span}(y^t,\hat{x}^t)$. The second and third constraints are rationality constraints necessary to ensure that the optimal forecast equals the mathematical expectation of x_t with respect to $\mathcal{I}_t$; that is, they ensure that $\hat{x}_t=E[x_t|y^t,\hat{x}^t]$. One restriction that these constraints impose is that the the consumer can never "forget" any past information.

It is possible to obtain a closed-form solution to the consumer's problem, which we present in the following proposition.

Proposition 3. *The forecast process* $\{\hat{x}\}$ *given by*

$$\hat{x}_t=\left[(1-\delta)\frac{h(L)-e^{-2\kappa}\delta L^{-1}h(\delta)}{1-\delta L^{-1}}\right]\varepsilon_t+\left[\sqrt{e^{-2\kappa}(1-e^{-2\kappa})}(1-\delta)\frac{h(\delta)\delta}{1-\delta L}\right]v_t,$$

with $v_t \overset{iid}{\sim} N(0,1)$ *and* $\{v_t\}$ *independent of* $\{\varepsilon_t\}$*, solves problem* (12).

It is illustrative to consider how the forecast process in this proposition depends on the parameter κ, which controls the quantity of foresight available to the consumer. As $\kappa\to 0$, the consumer's forecast process converges to

$$\hat{x}_t=(1-\delta)\frac{h(L)-\delta L^{-1}h(\delta)}{1-\delta L^{-1}}\varepsilon_t=E[x_t|y^t],$$

which is the optimal forecast of x_t with no foresight, according to the well-known formula of Hansen and Sargent (1980). On the other hand, as $\kappa\to\infty$, the forecast process converges to

$$\hat{x}_t=(1-\delta)\frac{h(L)}{1-\delta L^{-1}}\varepsilon_t=x_t,$$

which is the perfect foresight solution. For intermediate values of κ, the optimal forecast places some weight on future disturbances, but is also subject to independent, purely expectational disturbances, captured by the process $\{v_t\}$.[13]

The solution in Proposition (3) is stated in terms of the consumer's optimal forecast of the target variable x_t. Additional insight into this solution can be gained from looking at the types of signal structures that can generate these forecasts. While there are generally many such signal structures, one that is particularly intuitive involves the consumer receiving a signal of the target variable x_t plus i.i.d. noise.

Corollary 3. *The optimal forecast process in* (1) *is consistent with the consumer having a time-t information set of the form* $\mathcal{I}_t = span(y^t, s^t)$ *with*

$$s_t = x_t + \sigma_v v_t$$

where $\{v_t\}$ *is orthonormal white noise, independent of* $\{\varepsilon_t\}$*, and*

$$\sigma_v^2 \equiv \left(\frac{e^{-2\kappa}}{1 - e^{-2\kappa}}\right)\frac{\delta^2 h(\delta)^2}{(1+\delta)^2}.$$

From the expression for σ_v^2 we can see that the variance of the noise component of the signal is monotonically decreasing in κ, with limiting values

$$\lim_{\kappa\to 0}\sigma_v^2 = \infty \quad \text{and} \quad \lim_{\kappa\to\infty}\sigma_v^2 = 0.$$

When the consumer has no information capacity, the signal becomes infinitely noisy, and therefore completely uninformative about the target variable x_t. As the consumer's capacity increases, the signal precision increases, eventually revealing the value of x_t perfectly. In interpreting this result, it is important to remember that the signal structure is *consistent* with the optimal forecasting behavior of the consumer, but it does not require us to interpret s_t as an objective random variable that might in principle be directly measured by an outside econometrician. The corollary only says that the consumer optimally forecasts future income "as if" he were receiving signals of this type.

Corollary (3) describes an "imperfect information" representation of the consumer's optimal information structure, in which the consumer solves a signal-extraction problem to form his forecast. Corollary (4) shows it is also possible

[13] For further discussion regarding the importance of purely expectational disturbances in models with exogenous information structures, see Chahrour and Jurado (2018).

to derive an equivalent "perfect information" representation, in which income is expressed as the sum of independent components with time-t disturbances that the consumer observes perfectly at each point in time.

Corollary 4. *The optimal forecast process in* (3) *is consistent with the consumer having a time-t information set $\mathcal{I}_t = span(\eta^t, u^t)$ with*

$$y_t = \sqrt{1 - e^{-2\kappa}} \left(\frac{L - \delta}{1 - \delta L} \right) h(L)\eta_t + e^{-\kappa} h(L) u_t,$$

where $\{u_t\}$ and $\{\eta_t\}$ are independent orthonormal Gaussian white noise processes.

One way to interpret this result is to imagine that the consumer chooses among arbitrary possible independent-component representations of income, subject to the foresight constraint. Corollary (4) says that the consumer endogenously compresses the information he receives into just two components, which optimally inform him about his future income. Both components inherit the dynamics of income through the term $h(L)$. However, the first component has an additional dynamic term which depends on the magnitude of the economic parameter δ.

So far we have characterized the solution to the consumer's foresight problem, but we have not explored how his optimal information choice affects his consumption and saving behavior. The simplest way to do this is through a numerical example. We assume that income follows the ARMA(1,1) process from Proposition (2), with the same parameter values we used to construct Figure (2) in Section (4). For the economic parameters in the model, we set

$$\beta = 0.95, \quad \sigma = 0.5, \quad \text{and} \quad \phi = 0.01,$$

and consider a range of different values for the informational parameter κ.

Figure (3) shows the impulse response functions associated with each of the two model disturbances, the income disturbance ε_t and the purely expectational disturbance v_t from Proposition (3). The horizontal axis measures the number of time periods since the disturbance has occurred, so negative values indicate periods before the disturbance has taken place.

Focusing on the left column, which depicts responses to the fundamental income disturbance, the top panel shows how the income disturbance affects income over time. The disturbance has no effect on income before it occurs, it has its largest

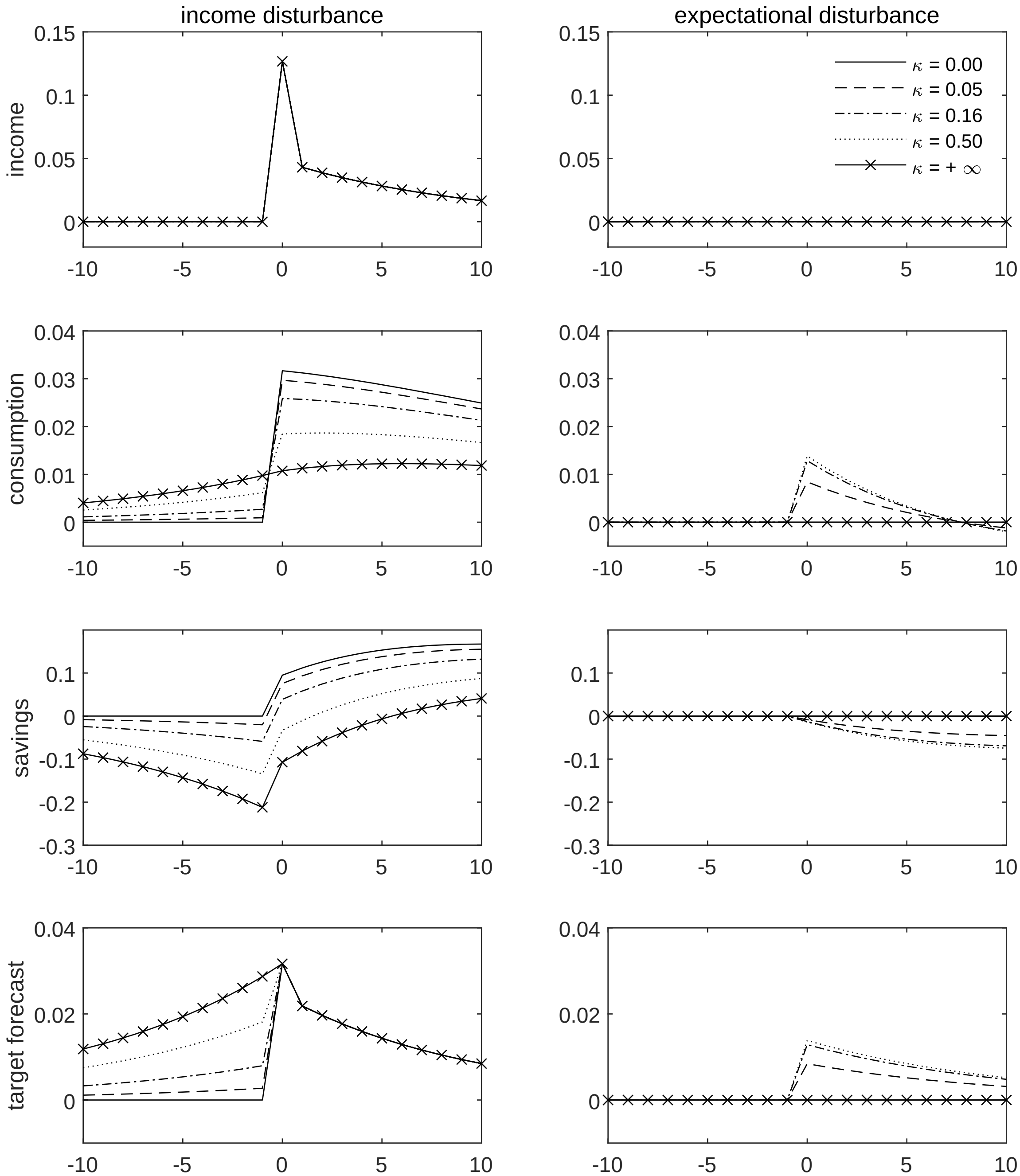

Figure 3: Impulse responses when $\beta = 0.95$, $\sigma = 0.5$, $\phi = 0.01$, and income follows the ARMA(1,1) process from Proposition (2) with $\rho = 0.9$, $\sigma_u^2 = 0.01$, and $\sigma_\eta^2 = 0.003$. The last row refers to the consumer's forecast of the target variable x_t, defined in (9), which is an exponential moving average of current and future income.

effect on impact, and then it has a smaller effect in subsequent periods as it decays at rate ρ. This response clearly does not depend on the value of the parameter κ, so all the lines lie on top of one another.

The response of consumption to the income disturbance depends on the information of the consumer, and therefore on the value of κ. When $\kappa = 0$, the consumer has no foresight; consumption does not respond to the income disturbance in advance, and then increases once the disturbance occurs. As κ increases, the consumer begins to respond to the disturbance in advance, by borrowing against his future income and reducing the increase in his consumption after the disturbance occurs. The degree to which the consumer is able to smooth his consumption response increases along with κ. The smoothest consumption profile is achieved when the consumer has perfect foresight.

The third panel in the first column shows how the consumer finances his chosen consumption path. He increases his debt by borrowing before the disturbance occurs, and then reduces his debt afterwards. The more foresight the consumer is allowed, the more he adjusts his asset position in order to smooth consumption. The only reason that consumption is not perfectly constant when $\kappa = \infty$ is that the elasticity of interest rates with respect to savings limits the consumer's desire to hold the extremely large asset positions required to achieve perfect consumption smoothing.

The bottom panel shows the response of $\hat{x}_t$, the consumer's forecast of the target variable x_t. With low levels of foresight, the consumer has only a small degree of confidence that a disturbance is going to occur in the future, so he only adjusts his forecast of lifetime income upwards by a small amount in anticipation of the shock. As the quantity of foresight increases, he becomes more confident that his lifetime income has increased and adjusts his forecast accordingly. Once the disturbance occurs, the consumer observes its realization and understands that income will evolve according to the dynamics in the top panel thereafter, so his forecasts no longer depend on κ.

Turning to the right column of Figure (3), the top panel shows that the expectational disturbance is independent of income. If the consumer has either no foresight or perfect foresight, then there are no independent disturbances to expectations, which is why the solid line and the cross-marked lines are both constant at zero for all the remaining panels in this column. For intermediate values of κ, the consumption response becomes positive on impact, and dies out gradually over time. The magnitude of the contemporaneous consumption response depends non-monotonically on κ, with an interior maximum. The consumer finances the higher consumption by borrowing against the future; however, since income never actually increases in the future, this means that at some point consumption must temporarily fall below its steady-state level, as can be seen in the figure around eight periods later.

Finally, we note that foresight increases the *persistence* of endogenous variables, holding fixed the persistence of income. Figure (4) plots the autocorrelation functions of consumption and the consumer's forecast of lifetime income. With higher levels of foresight, the consumer is better able to smooth consumption, which implies greater persistence.

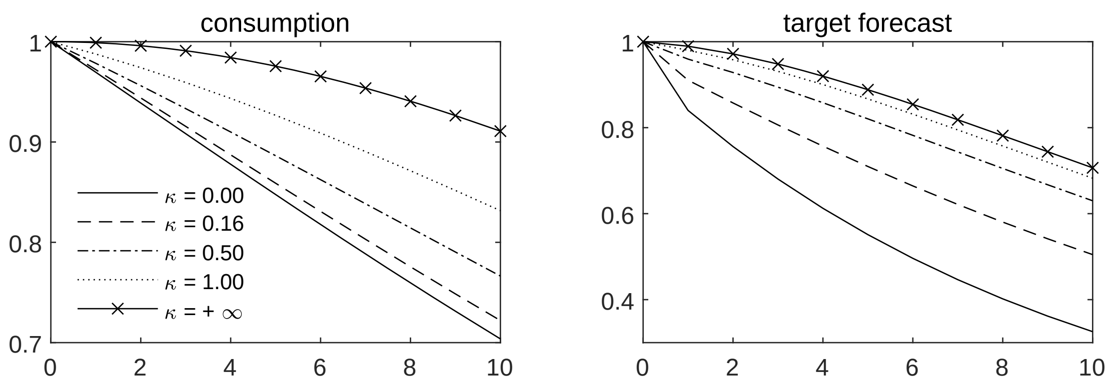

Figure 4: Autocorrelation functions when $\beta = 0.95$, $\sigma = 0.5$, $\phi = 0.01$, and income follows the ARMA(1,1) process from Proposition (2) with $\rho = 0.9$, $\sigma_u^2 = 0.01$, and $\sigma_\eta^2 = 0.003$. The right panel refers to the consumer's forecast of the target variable x_t, defined in (9), which is an exponential moving average of current and future income.

5.2 Comparison with sub-optimal foresight

Many of the results in Figure (3) would be qualitatively similar for any information structure with foresight, regardless of whether the type of foresight is optimal from the consumer's perspective or not. Here we perform two exercises to compare optimal foresight with sub-optimal foresight, as implied by the persistent-transitory representation from Section (3).

The first exercise is simply to plot the responses of model variables to the same two disturbances as in Figure (3) in two versions of the model: one in which the consumer has exogenous foresight of the type implied by the persistent-transitory representation (i.e. Proposition 2), and one in which the consumer has optimal foresight. In the first version, we use the same parameter values as in Section (3). We have seen that, at these values, the quantity of foresight is 0.16 nats. Therefore, in the version with optimal foresight, we set $\kappa = 0.16$ to ensure that the total quantity of foresight in both versions of the model is held constant.

Figure (5) shows the responses of income, consumption, savings, and expected lifetime income to the income and expectational disturbances. The second panel in the left column shows that consumption begins increasing earlier under optimal foresight, and does not exhibit a rapid run-up in the period just before the disturbance occurs. This result depends on the quantitative relationship between the parameters θ and δ. Under optimal foresight, Corollary (3) indicates that the consumer effectively constructs a signal which contains information about future income discounted at rate δ. However, with exogenous foresight of the type in Proposition (2), we saw that the consumer's signal discounts future income at rate θ. According to the parameter values in this example, $\theta < \delta$, which means that the exogenous signal is more informative about the near future than the consumer would optimally like. Because of this, the consumer's expectations of lifetime income mostly increase just before the income disturbance occurs, as can be seen in the bottom panel of the left column. (If instead $\delta < \theta$, then the reverse would be true, and the consumer's expectations of lifetime income would react more during the period of anticipation.)

One implication of information structures that are more informative about income far out into the future, is that expectational disturbances also have longer-lasting effects. In terms of the signal interpretation from Corollary (3), this is because it takes longer for the consumer to learn that increases in his signal were due to noise. Another consequence is that, while the overall volatility of consumption is lower under optimal foresight, the share of consumption volatility due to expectational disturbances is larger. Under the parameterization in this example, the share of the variance in log consumption that can be attributed to expectational disturbances is 1.9% with exogenous foresight, but 15.4% with optimal foresight. This is interesting because it demonstrates that an individual's optimal use of their limited information capacity implies their choices should also reflect substantial (ex-post) "mistakes" and that minimizing the effects of such noise is not the agent's objective.

The second comparison exercise we perform illustrates the type of misspecification errors that could arise from incorrectly assuming a sub-optimal form of foresight. The thought experiment is to suppose that, according to the data generating process, the

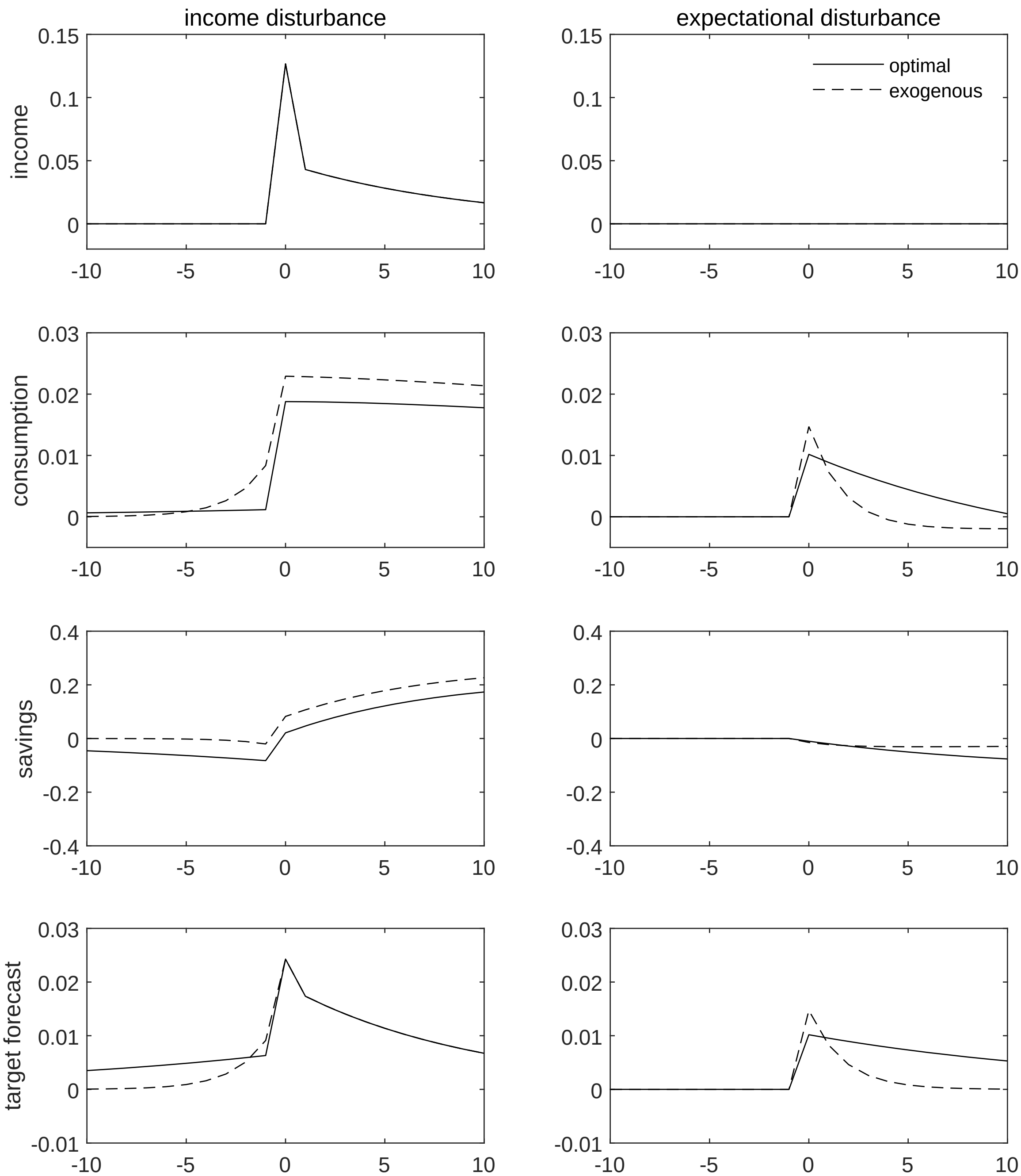

Figure 5: Impulse responses when $\beta = 0.95$, $\sigma = 0.5$, $\phi = 0.01$, $\kappa = 0.1625$, $\rho = 0.9$, $\sigma_\eta^2 = 0.003$, $\sigma_u^2 = 0.01$ and $h(L) = \sigma_\eta(1-\theta L)(1-\rho L)^{-1}$. The value of κ is chosen to keep the amount of foresight in both information structures the same. The last row refers to the consumer's forecast of the target variable x_t, defined in (9), which is an exponential moving average of current and future income.

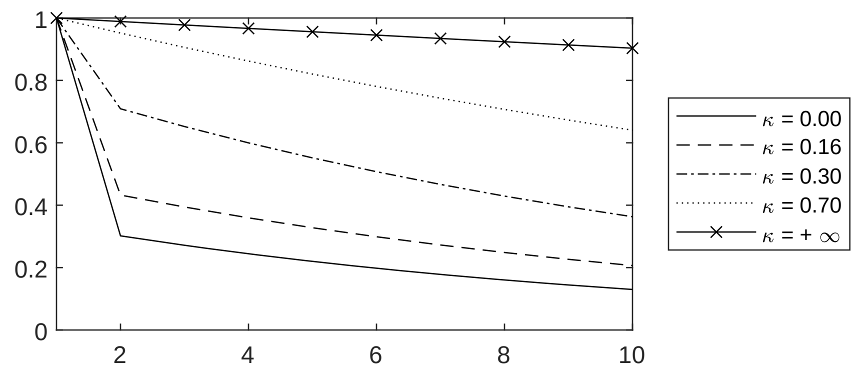

Figure 6: Misspecified estimates of the autocorrelation function of income.

consumer makes choices under optimal foresight; however, an outside econometrician attempts to fit a misspecified model in which the consumer has sub-optimal foresight of the type implied by the persistent-transitory representation. What we show is that, holding fixed the economic parameters, the only way the econometrician can match the additional persistence in the endogenous variables is by introducing additional (counter-factual) persistence in income.

Specifically, the econometrician chooses values of the three parameters in the persistent-transitory representation of income in the following way. First, he calibrates one to exactly match the variance of income. Then, he chooses the remaining two parameters to match the autocorrelation function of consumption as closely as possible. He does this by minimizing the distance between the empirical (true) and model-implied autocorrelations of consumption at one and ten periods (other ways of matching these autocorrelations have the same result). In this exercise, we assume that the econometrician knows the true values of the other model parameters, and that he observes autocovariances exactly. This second assumption is consistent with the econometrician basing his estimates on large samples from the data generating process.

Figure (6) plots the autocorrelation function of income estimated by the econometrician for different values of κ. When $\kappa = 0$, there is no foresight, and the econometrician's estimate exactly corresponds to the autocorrelation function of income in the data generating process. However, as κ increases, the econometrician mistakenly attributes the additional persistence in consumption to additional persistence in income. This highlights the danger of tying physical assumptions regarding income too closely to informational assumptions regarding foresight. In this case, the persistent-transitory representation provides insufficient degrees of freedom to correctly estimate these two independent sources of persistence. This is despite the fact that the subjective signal in both the econometrician's model and the data generating process has *exactly the same form*: "average future income plus i.i.d. noise" (cf. the signal in Proposition 2 and the signal in Corollary 3.)

6 Conclusion

Foresight is a common assumption in the literature on business cycles with technological news, asset pricing with long-run risks, or consumption choice with persistent and transitory components of income. In this paper we draw attention to this assumption and provide ways of comparing information structures in terms of the type and quantity of foresight they contain. A main result is Proposition (3), which generalizes the Hansen-Sargent formula to the case when agents can endogenously choose the type of foresight they have subject to an informational constraint.

The approach to endogenous foresight taken in this paper also suggests a number of possible applications. One would be to combine endogenous foresight with the typical assumption in the rational inattention literature that current and past exogenous variables are only imperfectly observed as well (although perhaps at a lower informational cost). Such a combination has been performed by Jurado (2020), but under the assumption that the cost of processing information about the past and the future is the same. Introducing asymmetric (but nonzero) costs of processing information about the past and future may be important for quantitatively reconciling the tension between empirical evidence suggesting strong responses to anticipated disturbances (e.g. Kurmann and Sims, 2017) with other evidence of slow adjustment to other economic developments (e.g. Carroll, 2003; Coibion and Gorodnichenko, 2015).

Other interesting applications require introducing endogenous foresight into a general equilibrium environment. This would permit an analysis of the interaction between foresight and economic policy, such as "forward guidance" regarding monetary policy. In such an environment, one advantage of shifting focus from endogenous hindsight to endogenous foresight is that it would allow us to avoid many of the conceptual challenges associated with market clearing and the presence of endogenous individual-level state variables faced by existing models of information choice. As a preliminary result in this direction, Appendix (B) illustrates how to apply our theory of foresight to the overlapping-generations equilibrium model of Gaballo (2016). The model simplifies many dimensions of the equilibrium analysis, a full treatment of which will require further work.

References

Azeredo da Silveira, R. and M. Woodford (2019). Noisy Memory and Over-Reaction to News. *AEA Papers and Proceedings 109*, 557–61.

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

Benigno, P. and M. Woodford (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory 147*(1), 1–42.

Carroll, C. D. (2003). Macroeconomic Expectations of Households and Professional Forecasters. *The Quarterly Journal of Economics 118*(1), 269–298.

Chahrour, R. and K. Jurado (2018). News or Noise? The Missing Link. *American Economic Review 108*(7), 1702–36.

Cochrane, J. H. (1994). Shocks. *Carnegie-Rochester Conference Series on Public Policy 41*(1), 295–364.

Coibion, O. and Y. Gorodnichenko (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review 105*(8), 2644–78.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Hoboken, NJ: Wiley-Interscience.

Friedman, M. and S. Kuznets (1945). *Income from Independent Professional Practice*. New York, NY: National Bureau of Economic Research.

Gaballo, G. (2016). Rational Inattention to News: The Perils of Forward Guidance. *American Economic Journal: Macroeconomics 8*(1), 42–97.

Gantmacher, F. R. and M. G. Krein (2002). *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems: Revised Edition.* Providence, RI: AMS Chelsea Publishing.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hansen, L. P. and T. J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control 2*, 7–46.

Hansen, L. P. and T. J. Sargent (2014). *Recursive Models of Dynamic Linear Economies.* Princeton, NJ: Princeton University Press.

Jurado, K. (2020). Rational Inattention in the Frequency Domain. Working Paper.

Kurmann, A. and E. Sims (2017). Revisions in utilization-adjusted tfp and robust identification of news shocks. Working Paper 23142, National Bureau of Economic Research.

Luo, Y. (2008). Consumption dynamics under information processing constraints. *Review of Economic Dynamics 11*(2), 366–385.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2018a). Dynamic Rational Inattention: Analytical Results. *Journal of Economic Theory 176*, 650–692.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2018b). Survey: Rational Inattention, a Disciplined Behavioral Model. Working Paper 13243, Center for Economic Policy Research.

Miao, J., J. Wu, and E. Young (2020). Multivariate Rational Inattention. Working paper.

Rozanov, Y. A. (1960). Spectral properties of multivariate stationary processes and boundary properties of analytic matrices. *Theory of Probability and its Applications 5*(4), 362–376.

Simon, H. A. (1956). Dynamic Programming Under Uncertainty with a Quadratic Criterion Function. *Econometrica 24*(1), 74–81.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy 49*, 317–356.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics 50*(3), 665–690.

Theil, H. (1957). A Note on Certainty Equivalence in Dynamic Planning. *Econometrica 25*(2), 346–349.

Veldkamp, L. L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton, NJ: Princeton University Press.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods*. Minneapolis, MN: University of Minnesota Press.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, pp. 25–58. Princeton, NJ: Princeton University Press.

Online Appendix to:
Optimal Foresight

Ryan Chahrour — Boston College

Kyle Jurado — Duke University

A Proofs

Proof of Proposition (1). Relative to the discussion in the text, what remains is to prove the expression for $\det \hat{\Sigma}_T$ given in (4), and then to show that $0 \leq r^2 \leq 1$. Regarding the first of these, notice that

$$\begin{aligned}
\det \hat{\Sigma}_T &= \det(I_T + C_\eta'(C_u C_u')^{-1} C_\eta) \det(C_u C_u') && \text{(matrix determinant lemma)} \\
&= \det\left(I_T + \frac{1}{\sigma_u^2} C_\eta C_\eta'\right) \sigma_u^{2T} && (C_u C_u' = \sigma_u^2 I_T) \\
&= \det\left(I_T + a A_T^{-1}\right) \sigma_u^{2T} && (C_\eta C_\eta' = \sigma_\eta^2 A_T^{-1}, \text{ with } a \text{ and } A_T \text{ defined below}) \\
&= \det\left(A_T + a I_T\right) \sigma_u^{2T} && \text{(matrix determinant lemma)} \\
&\equiv d_T \sigma_u^{2T} && (13)
\end{aligned}$$

where we have defined $a \equiv \sigma_\eta^2 / \sigma_u^2$ and

$$A_T = \begin{bmatrix} 1 & -\rho & & & \\ -\rho & 1+\rho^2 & -\rho & & \\ & -\rho & 1+\rho^2 & \ddots & \\ & & \ddots & \ddots & -\rho \\ & & & -\rho & 1+\rho^2 \end{bmatrix}.$$

Using the band structure of A_T, it follows that

$$A_T + aI_T = \begin{bmatrix} 1+a & -\rho & & & \\ -\rho & 1+\rho^2+a & -\rho & & \\ & -\rho & 1+\rho^2 & \ddots & \\ & & \ddots & \ddots & -\rho \\ & & & -\rho & 1+\rho^2+a \end{bmatrix}.$$

The determinant of this matrix for arbitrary $T \geq 1$, can be computed from the recurrence relation

$$d_T = (1 + \rho^2 + a)d_{T-1} - \rho^2 d_{T-2}$$

with $d_0 = 1$ and $d_1 = 1 + a$ (e.g. Gantmacher and Krein, 2002, p.67). The solution to this recurrence relation is

$$d_T = c_1 \lambda_1^T + c_2 \lambda_2^T, \tag{14}$$

where λ_1 and λ_2 are the two roots of the polynomial $\mathcal{P}(\lambda) = \lambda^2 - (1 + \rho^2 + a)\lambda + \rho^2$ and c_1 and c_2 are chosen to satisfy the initial conditions. Based on the definition of θ in Proposition (2), we can see that the two roots of $\mathcal{P}(\lambda)$ must be

$$\lambda_1 = \frac{\rho}{\theta} \quad \text{and} \quad \lambda_2 = \rho\theta.$$

Using these together with the initial conditions to determine c_1 and c_2, we find

$$c_1 = 1 - r^2 \quad \text{and} \quad c_2 = r^2, \quad \text{where} \quad r^2 \equiv \frac{\rho - \theta(1 + a)}{\rho(1 - \theta^2)}.$$

Plugging the expressions for λ_1, λ_2, c_1, and c_2 into (14) and then plugging the expression for d_T into (13), we arrive at the expression in (4).

To show that $0 \leq r^2 \leq 1$, first observe that $\theta \leq 1 \leq (1 + a)/\rho$. Multiplying both sides by $\rho\theta$, we find $\rho\theta^2 \leq \theta(1 + a)$, which implies that $r^2 \leq 1$. To show that $r^2 \geq 0$, we need to prove that

$$\rho \geq \theta(1 + a). \tag{15}$$

By the definition of θ, θ/ρ is the smaller root of the polynomial $\mathcal{P}(z) = \rho^2 z^2 - (1 + \rho^2 + a)z + 1$. Since $\mathcal{P}(0) = 1 > 0$, $\mathcal{P}(1) = -a \leq 0$, and $\mathcal{P}(z) > 0$ as $z \to \infty$, it follows that the two roots of this polynomial satisfy $0 \leq z_1 \leq 1 \leq z_2$. Now notice that

$$\mathcal{P}\left(\frac{1}{1 + a}\right) = -\rho^2 \frac{a}{(1 + a)^2} \leq 0,$$

so it must also be true that $z_1 = \theta/\rho \leq 1/(1 + a)$, which proves (15). □

Proof of Corollary (1)**.** Define the ratio $a \equiv \sigma_\eta^2/\sigma_u^2$ so that

$$r^2 = \frac{\rho - \theta(1 + a)}{\rho(1 - \theta^2)}.$$

Differentiating with respect to ρ,

$$\frac{\partial r^2}{\partial \rho} = \frac{(1 + a)(\theta - \rho\theta') + 2\theta\theta'\rho(\rho - \theta(1 + a))}{\rho^2(1 - \theta^2)},$$

where $\theta' \equiv \partial\theta/\partial\rho$. The denominator is non-negative, so let us focus on the numerator. The second term is non-negative, as we have shown in (15) and will establish in the proof of Corollary (2). As for the first term, notice that

$$\theta' = \frac{\theta(1-\rho^2+a)}{\rho\sqrt{(1+\rho^2+a)^2-4\rho^2}} \leq \frac{\theta}{\rho}\frac{1-\rho^2+a}{1+\rho^2+a} \leq \frac{\theta}{\rho}.$$

This means that $\theta - \rho\theta' \geq 0$, so the first term in the numerator is also non-negative, and $\partial r^2/\partial\rho \geq 0$. Since F is a monotonically increasing function of r^2, it follows that $\partial F/\partial\rho \geq 0$ as well.

For the limiting values with respect to ρ, we use the result in Corollary (2) that $\theta \to 0$ as $\rho \to 0$ to see that $r^2 \to 0$ and therefore $F \to 0$ in this case. Similarly, using the result that $\theta \to \bar{\theta}$ as $\rho \to 1$, it follows that $r^2 \to \bar{r}^2$ and $F \to -\frac{1}{2}\ln(1-\bar{r}^2)$ in this case.

Turning to the second part of the Corollary, we can use the fact that $\theta \to \rho$ as $a \to 0$ to see that

$$\lim_{a\to 0} r^2 = \frac{\rho-\rho}{1-\rho^2} = 0,$$

which implies that $F \to 0$ as well. For the second limit with respect to a, write

$$r^2 = \frac{\rho}{1-\theta^2} - \frac{\theta}{1-\theta^2} - \frac{\theta a}{1-\theta^2}.$$

As $a \to \infty$, we know from Corollary (2) that θ approaches zero, so the first term converges to ρ and the second term converges to zero. What remains is to prove that the third term converges to $-\rho$. To that end,

$$\begin{aligned}
\lim_{a\to\infty} -\theta a &= \lim_{a\to\infty} \frac{-2\rho a}{1+\rho^2+a+\sqrt{(1+\rho^2+a)^2-4\rho^2}} && \text{(θ and $1/\theta$ are reciprocal roots)}\\
&= \lim_{a\to\infty} \frac{-2\rho}{1+\sqrt{\dfrac{(1+\rho^2+a)^2}{(1+\rho^2+a)^2-4\rho^2}}} && \text{(L'Hopital's rule)}\\
&= -\rho.
\end{aligned}$$

Therefore, we have shown that $r^2 \to 0$ as $a \to \infty$, and so $F \to 0$ as well. □

Proof of Proposition (2)**.** According to the persistent-transitory representation in (3), the autocovariance generating function of $\{y_t\}$ is given by

$$g_y(z) = \frac{\sigma_\eta^2}{|1-\rho z|^2} + \sigma_u^2 = \frac{\sigma_\eta^2 + \sigma_u^2|1-\rho z|^2}{|1-\rho z|^2}.$$

Factoring the numerator,

$$\sigma_\eta^2 + \sigma_u^2|1-\rho z|^2 = \sigma_\eta^2\frac{\rho}{\theta}|1-\theta z|^2, \tag{16}$$

where θ is the root of the polynomial $\mathcal{P}(z) = \rho z^2 - (1+\rho^2+\sigma_\eta^2/\sigma_u^2)z + \rho$ that lies inside the unit circle,

$$\theta = \frac{1}{2\rho}\left(1+\rho^2+\sigma_\eta^2/\sigma_u^2 - \sqrt{(1+\rho^2+\sigma_\eta^2/\sigma_u^2)^2 - 4\rho^2}\right).$$

Defining $\sigma_\varepsilon^2 \equiv \sigma_\eta^2\rho/\theta$ and substituting (16) into the numerator of $g_y(z)$, we get

$$g_y(z) = \sigma_\varepsilon^2\frac{|1-\theta z|^2}{|1-\rho z|^2},$$

which is the autocovariance generating function of the ARMA(1,1) process stated in the proposition.

As for the signal process, first project z_t onto the space spanned by past, present, and future income,[14]

$$z_t = E[z_t|\ldots, y_{t+1}, y_t, y_{t-1}, \ldots] + \zeta_t = \frac{g_z(L)}{g_y(L)}y_t + \zeta_t, \tag{17}$$

where $g_z(z) = \sigma_\eta^2/|1-\rho z|^2$ is the autocovariance generating function of $\{z_t\}$, and $\{\zeta_t\}$ is independent of income. More explicitly, the projection weights are generated by the function

$$\frac{g_z(z)}{g_y(z)} = \frac{g_z(z)}{g_z(z)+\sigma_u^2} = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_u^2|1-\rho z|^2} = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\theta}{\rho}\frac{1}{|1-\theta z|^2},$$

where the last equality makes use of the factorization result (16). The autocovariance generating function of the error process $\{\zeta_t\}$ is given by

$$g_\zeta(z) = g_z(z) - \frac{g_{zy}(z)g_{yz}(z)}{g_y(z)} = \sigma_\eta^2\frac{\theta}{\rho}\frac{1}{|1-\theta z|^2}.$$

[14]See Whittle (1983) for a review of this and related projection results.

Substituting these two expressions into (17), we can write

$$z_t = \frac{\sigma_\eta^2}{\sigma_u^2} \frac{\theta/\rho}{(1-\theta L)(1-\theta L^{-1})} y_t + \frac{\sigma_\eta \sqrt{\theta/\rho}}{1-\theta L} v_t,$$

where v_t is orthonormal white noise. Lastly, define the new signal

$$\begin{aligned} s_t &\equiv (1-\theta)\frac{\sigma_u^2}{\sigma_\eta^2}\frac{\rho}{\theta}(1-\theta L) z_t \\ &= \frac{1-\theta}{1-\theta L^{-1}} y_t + (1-\theta)\frac{\sigma_u^2}{\sigma_\eta}\sqrt{\frac{\theta}{\rho}} v_t \\ &= (1-\theta)\sum_{j=0}^{\infty} \theta^j y_{t+j} + \sigma_v v_t, \end{aligned}$$

where the second lines substitutes in the previous expression for z_t, and the third line defines $\sigma_v^2 \equiv (1-\theta)^2 \sigma_\varepsilon^2 \sigma_u^2 / \sigma_\eta^2$. Because $0 < \theta < 1$ and $y_t \in (y^t)$, the transformation in the first line is such that $\text{span}(y^t, s^t) = \text{span}(y^t, z^t)$. □

Proof of Corollary (2)**.** Define the ratio $a \equiv \sigma_\eta^2/\sigma_u^2$ so that

$$\theta = \frac{1}{2\rho}\left(1+\rho^2+a-\sqrt{(1+\rho^2+a)^2-4\rho^2}\right).$$

Differentiating with respect to ρ,

$$\frac{\partial\theta}{\partial\rho} = \frac{\theta(1-\rho^2+a)}{\rho\sqrt{(1+\rho^2+a)^2-4\rho^2}} \geq 0,$$

which proves that θ is monotonically increasing in ρ. Regarding its limiting behavior as ρ approaches one, we have

$$\lim_{\rho\to1}\theta = \frac{1}{2}\left(2+a-\sqrt{(2+a)^2-4}\right).$$

As ρ approaches zero, we can use L'Hopital's rule,

$$\lim_{\rho\to0}\theta = \lim_{\rho\to0}\rho\left(1-\sqrt{\frac{(1+\rho^2+a)^2}{(1+\rho^2+a)^2-4\rho^2}}\right) = 0.$$

This completes the proof of the first part of the Corollary. For the second part, we can see from differentiating θ with respect to a that

$$\frac{\partial\theta}{\partial a} = \frac{1}{2\rho}\left(1-\sqrt{\frac{(1+\rho^2+a)^2}{(1+\rho^2+a)^2-4\rho^2}}\right) \leq 0,$$

so θ is monotonically decreasing in a. As a approaches zero, we have

$$\lim_{a\to 0}\theta = \frac{1}{2\rho}\left(1+\rho^2-\sqrt{(1-\rho^2)^2}\right) = \rho.$$

To find the limit as a approaches infinity, it is easiest to notice that

$$\frac{1}{\theta} = \frac{1}{2\rho}\left(1+\rho^2+a+\sqrt{(1+\rho^2+a)^2-4\rho^2}\right),$$

since the polynomial $\mathcal{P}(z) = \rho z^2 - (1+\rho^2+a)z+\rho$ has reciprocal roots. Because the right side becomes infinite as a does, it follows that θ approaches zero. □

Proof of Lemma (1)**.** We closely follow the strategy described in Benigno and Woodford (2012). In the consumption-saving model, the steady-state values of consumption C and savings B must satisfy the two equations

$$1 = \beta(1+r+\Phi(B)+\Phi'(B)B) \tag{18}$$

$$C = (r+\Phi(B))B+Y, \tag{19}$$

where Y is the (exogenous) steady-state level of income. We now seek to construct an LQ approximation around this steady state. First, we perform a second-order approximation to the consumer's lifetime utility function

$$V_0 \equiv E_0\sum_{t=0}^{\infty}\beta^t u(C_t) = E_0\sum_{t=0}^{\infty}\beta^t\left[u(C)+u'(C)C\left(c_t+\frac{1}{2}c_t^2\right)+\frac{1}{2}u''(C)C^2c_t^2\right]+\mathcal{O}(\epsilon^3) \tag{20}$$

where $\epsilon > 0$ is a real number that scales the degree of randomness in the model. The flow budget constraint implies that

$$E_0\sum_{t=0}^{\infty}\beta^t\lambda\left[(1+r+\Phi(B_{t-1}))B_{t-1}+Y_t-C_t-B_t\right] = 0,$$

where $\lambda = u'(C)$ is the steady-state Lagrange multiplier. Taking a second-order approximation to this expression in terms of the variables c_t, b_t, and y_t, we obtain

$$0 = \sum_{t=0}^{\infty}\beta^t u'(C)\left[-Cc_t-\frac{1}{2}Cc_t^2-b_t+Yy_t+\frac{1}{2}Yy_t^2+\beta\left(\frac{1}{\beta}b_t-\frac{1}{2}\phi b_t^2\right)\right]+\mathcal{O}(\epsilon^3),$$

where we have substituted in the steady-state value of the Lagrange multiplier, and made use of the steady-state relation (18). Rearranging this expression, we can write

$$\sum_{t=0}^{\infty}\beta^t u'(C)C\left[c_t+\frac{1}{2}c_t^2\right] = \sum_{t=0}^{\infty}\beta^t u'(C)\left[-\frac{1}{2}\beta\phi b_t^2+Yy_t+Yy_t^2\right]+\mathcal{O}(\epsilon^3).$$

We can then use this relation to eliminate the linear terms in (20), which delivers the purely quadratic approximation

$$V_0 = E_0 \sum_{t=0}^{\infty} \beta^t \left[u(C) + \frac{1}{2} u''(C) C^2 c_t^2 - \frac{1}{2} u'(C) \beta \phi b_t^2 + u'(C) \left(Y y_t + \frac{1}{2} Y y_t^2 \right) \right] + \mathcal{O}(\epsilon^3)$$

Finally, to arrive at the objective stated in the lemma, we divide this expression by $-u''(C)C^2 > 0$ and then add terms which are independent of the consumer's choices. The reason for including these specific terms will become clear in the proof of Lemma (2), but at this point it is sufficient to observe that they do not affect the consumer's rankings of alternative plans. □

Derivation of (10)**.** The optimality conditions associated with the LQ problem (1) are given by

$$c_t = E_t[c_{t+1}] + \beta \phi \sigma b_t \tag{21}$$

$$C c_t + b_t = \beta^{-1} b_{t-1} + Y y_t, \tag{22}$$

Substituting (22) into (21) to eliminate c_t, we obtain an expectational difference equation for $\{b_t\}$ of the form

$$E_t[A(L) b_{t+1}] = Y E_t[y_{t+1} - y_t], \tag{23}$$

where

$$A(L) \equiv 1 - (1 + \beta^{-1} + \sigma \phi \beta C) L + \beta^{-1} L^2.$$

This lag polynomial can be factored as

$$A(L) = (1 - \lambda_1 L)(1 - \lambda_2 L),$$

where λ_1, λ_2 are the two roots of the polynomial

$$\mathcal{P}(\lambda) = \lambda^2 - (1 + \beta^{-1} + \sigma \phi \beta C)\lambda + \beta^{-1}. \tag{24}$$

Notice that $\mathcal{P}(0) = \beta^{-1} > 0$, $\mathcal{P}(1) = -\sigma \phi \beta C < 0$, and $\mathcal{P}(\lambda) > 0$ for all large enough positive values of λ. It follows that there must be two real roots, satisfying $0 < \lambda_1 < 1 < \lambda_2$. Comparing the factorization with the original lag polynomial, we see that

$$\lambda_1 + \lambda_2 = 1 + \beta^{-1} + \sigma \phi \beta C, \quad \text{and} \quad \lambda_1 \lambda_2 = \beta^{-1}.$$

Using this factorization, we can write (23) as

$$E_t[(1-\lambda_1 L)(1-\lambda_2 L)b_{t+1}] = Y E_t[y_{t+1} - y_t],$$

or as

$$q_t = \lambda_2^{-1} E_t[q_{t+1}] - \lambda_2^{-1} Y E_t[y_{t+1} - y_t],$$

where $q_t \equiv (1-\lambda_1 L)b_t$. Because $|\lambda_2^{-1}| < 1$, this can be solved forward and rewritten in terms of b_t to get

$$b_t = \lambda_1 b_{t-1} - \lambda_2^{-1} Y \sum_{j=0}^{\infty} \lambda_2^{-j} E_t[y_{t+j+1} - y_{t+j}]. \tag{25}$$

Now, notice that

$$\begin{aligned}\sum_{j=0}^{\infty} \lambda_2^{-j} E_t[y_{t+j+1} - y_{t+j}] &= \lambda_2 \sum_{j=1}^{\infty} \lambda_2^{-j} y_{t+j} - y_t - \sum_{j=1}^{\infty} \lambda_2^{-j} y_{t+j} \\ &= -\lambda_2 y_t + (\lambda_2 - 1) \sum_{j=0}^{\infty} \lambda_2^{-j} y_{t+j}.\end{aligned}$$

Substituting this into (25), we get

$$b_t = \lambda_1 b_{t-1} + Y y_t - (1-\lambda_2^{-1}) Y \sum_{j=0}^{\infty} \lambda_2^{-j} E_t[y_{t+j}]. \tag{26}$$

Substituting this solution for b_t into (22) and solving for Cc_t, we obtain

$$Cc_t = (1-\lambda_2^{-1})\beta^{-1} b_{t-1} + (1-\lambda_2^{-1}) Y \sum_{j=0}^{\infty} \lambda_2^{-j} E_t[y_{t+j}],$$

This is the consumption function from (10) when we define $\delta \equiv 1/\lambda_2$. □

Proof of Lemma (2)**.** We begin by finding a closed-form expression for the continuation utility $V_t^*(b_{t-1})$ of a (hypothetical) consumer who enters period t with savings b_{t-1} and has *perfect foresight* from then on. Using the method of undetermined coefficients, we find that

$$V_t^*(b_{t-1}) = -\frac{1}{2}\left(\frac{1}{1-\delta}\right)\left[c_t^{*2} - \frac{\beta-\delta}{\beta-\delta^2}\frac{Y^2}{C^2} x_t^2\right], \tag{27}$$

where c_t^* is the consumption plan from time t onward under perfect foresight,

$$c_t^* = \frac{1}{C}\left[(1-\delta)\beta^{-1} b_{t-1} + Y x_t\right]. \tag{28}$$

To verify the expression in (27), let

$$u_t^* \equiv -\frac{1}{2}\left\{c_t^{*2} + \frac{\beta\sigma\phi}{C}b_t^{*2} + \left(\frac{\beta-\delta}{\beta-\delta^2}\right)\frac{Y^2}{C^2}\left((1-\delta)y_t^2 - 2y_t x_t\right)\right\}$$

denote the time-t utility flow under perfect foresight, where

$$b_t^* = \frac{\delta}{\beta}b_{t-1} + Yy_t - Yx_t \tag{29}$$

is the associated optimal savings plan. The continuation value V_t^* must satisfy the recursion

$$V_t^*(b_{t-1}) = u_t^* + \beta V_{t+1}^*(b_t^*).$$

By plugging the above expressions for u_t^* and conjectured $V_{t+1}^*(b_t^*)$ into the right side of this equation, repeatedly substituting in the policy functions (28) and (29), and using the fact that

$$x_{t+1} = \delta^{-1}(x_t - (1-\delta)y_t)$$

together with the definition of δ in Lemma (1), we arrive at the expression for $V_t^*(b_{t-1})$ in (27).

Next, we conjecture that the optimal continuation utility of the consumer with imperfect foresight $V_t(b_t)$ can be written in terms of the perfect foresight continuation utility $V_t^*(b_{t-1})$ and a discounted sum of forecast error variances,

$$V_t(b_{t-1}) = E_t V_t^*(b_{t-1}) - \omega D_t, \tag{30}$$

where ω is an undetermined coefficient, and

$$D_t = \frac{1}{2}E_t[(x_t - E_t x_t)^2] + \beta E_t D_{t+1}.$$

Letting

$$u_t \equiv -\frac{1}{2}E_t\left\{c_t^2 + \frac{\beta\sigma\phi}{C}b_t^2 + \left(\frac{\beta-\delta}{\beta-\delta^2}\right)\frac{Y^2}{C^2}\left((1-\delta)y_t^2 - 2y_t x_t\right)\right\}$$

denote the optimal time-t utility flow under imperfect foresight, (30) and the recursions for V_t^* and D_t imply that $V_t(b_{t-1}) = u_t + \beta E_t V_{t+1}(b_t)$ if and only if

$$u_t = E_t\left[u_t^* + \beta V_{t+1}^*(b_t^*) - \beta V_{t+1}^*(b_t)\right] - \omega\frac{1}{2}E_t[(x_t - E_t x_t)^2].$$

Expanding the right side of this equation, we find that the equation is satisfied if and only if $\omega = \delta^{-1}(C/Y)^2$. Therefore, by (30),

$$E[V_0] = E[V_0^*] - \delta^{-1}(C/Y)^2 E[D_0]$$

Finally, since $E[V_0^*]$ is independent of the consumer's choices, maximization of $E[V_0]$ is equivalent to minimization of $E[D_0]$, which is the loss function stated in the lemma.

□

Proof of Proposition (3)**.** The proof begins by reducing the consumer's problem (12) to a simpler problem which only involves predicting the part of the target variable that is unknown conditional on current and past income.

Lemma 3. *Choosing* $\{\hat{x}_t\}$ *to solve the consumer's problem* (12) *is equivalent to choosing* $\{\hat{z}_t\}$ *to solve the problem*

$$\min_{\{\hat{z}_t\}} E[(z_t - \hat{z}_t)^2] \qquad \textit{subject to} \tag{31}$$

(i) $\lim_{T\to\infty} I((\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}, \hat{z}^t | \varepsilon^t) \leq \kappa$

(ii) $E[(z_t - \hat{z}_t)\hat{z}_{t-j}] = 0$ *for all* $j \geq 0$

(iii) $E[(z_t - \hat{z}_t)\varepsilon_{t-j}] = 0$ *for all* $j \geq 0$,

where $z_t = \delta L^{-1}/(1 - \delta L^{-1})\varepsilon_t$, *and then setting*

$$\hat{x}_t = (1-\delta)\frac{h(L) - \delta L^{-1} h(\delta)}{1 - \delta L^{-1}}\varepsilon_t + (1-\delta)h(\delta)\hat{z}_t.$$

Proof. By exchanging expectation and summation, is easy to see that minimizing (12) is equivalent to minimizing $E[(x_t - \hat{x}_t)^2]$. We now show that this, in turn, is equivalent to minimizing $E[(z_t - \hat{z}_t)^2]$, when z_t is defined as in the lemma and $\hat{z}_t \equiv E_t[z_t]$. To see this, notice that

$$\begin{aligned} x_t - \hat{x}_t &= (x_t - E[x_t|y^t]) - (E_t[x_t] - E[x_t|y^t]) && \text{(definition of } \hat{x}_t\text{)} \\ &= (x_t - E[x_t|y^t]) - E_t[x_t - E[x_t|y^t]] && \text{(since span}(y^t) \subseteq \mathcal{I}_t\text{)} \\ &= (1-\delta)h(\delta)(z_t - \hat{z}_t), \end{aligned}$$

if we define

$$z_t \equiv \frac{1}{(1-\delta)h(\delta)}\left(x_t - E[x_t|y^t]\right). \tag{32}$$

Therefore, $E[(x_t - \hat{x}_t)^2] = (1-\delta)^2 h(\delta)^2 E[(z_t - \hat{z}_t)^2]$. To verify that the law of motion for z_t in Lemma (3) is consistent with (32), we can use the formula of Hansen and Sargent (1980) to compute

$$E[x_t|y^t] = (1-\delta)\frac{h(L) - \delta L^{-1} h(\delta)}{1-\delta L^{-1}}\varepsilon_t, \tag{33}$$

and then substitute this expression and the definition of x_t into (32). We can also take conditional expectations on both sides of (32) with respect to $\mathcal{I}_t$ and rearrange to find the implied relationship between $\hat{x}_t$ and $\hat{z}_t$ stated in the lemma,

$$\hat{x}_t = E[x_t|y^t] + (1-\delta)h(\delta)\hat{z}_t.$$

Second, we show that the constraints of the original problem are satisfied if and only if the constraints of the reduced problem are satisfied. Notice that with Gaussian information structures, the constraint sets only depend on the linear spaces spanned by the relevant variables, and not on the variables themselves. And, in both cases, Gaussian information structures are optimal for the consumer because the objective is to minimize the error variance, and Gaussian processes maximize entropy (minimize foresight) for a given error variance. Therefore, all we need to show is that $\text{span}(\hat{x}^t, y^t) = \text{span}(\hat{z}^t, \varepsilon^t)$. But this follows from the fact that $\hat{x}_t \in \text{span}(\hat{z}^t, \varepsilon^t)$ and $y_t \in \text{span}(\hat{z}^t, \varepsilon^t)$ for all t, and conversely, $\hat{z}_t \in \text{span}(\hat{x}^t, y^t)$ and $\varepsilon_t \in \text{span}(\hat{x}^t, y^t)$ for all t as well. □

Using the relation between $\hat{x}_t$ and $\hat{z}_t$ stated in Lemma (3), the conjectured solution for $\hat{x}_t$ in (3) translates into a conjectured solution for $\hat{z}_t$ of the form

$$\hat{z}_t = \psi\frac{\delta L^{-1}}{1-\delta L^{-1}}\varepsilon_t + \sqrt{\psi(1-\psi)}\frac{\delta}{1-\delta L}v_t, \tag{34}$$

where we have defined the new parameter $\psi \equiv 1 - e^{-2\kappa}$ to simplify some of the following expressions. Therefore, what we need to show is that this solves the problem in Lemma (3). To do so, we first construct a lower bound on the objective function and show that the conjectured solution attains this lower bound. Then, we show that the conjecture is feasible by verifying that it satisfies all the constraints.

Lemma 4.

$$E[(z_t - \hat{z}_t)^2] \geq \left(\frac{\delta^2}{1-\delta^2}\right)e^{-2\kappa},$$

with equality when $\hat{z}_t$ is given by (34).

Proof. Observe that

$$\begin{aligned}
\kappa &= \lim_{T\to\infty} I((\varepsilon_{t+1},\ldots,\varepsilon_{t+T}),\hat{z}^t|\varepsilon^t) && \text{(foresight constraint)}\\
&= \lim_{T\to\infty} I((z_t,\ldots,z_{t+T}),\hat{z}^t|\varepsilon^t) && (\text{since } \varepsilon_{t+1} = z_t/\delta - z_{t+1})\\
&\geq I(z_t,\hat{z}^t|\varepsilon^t) && \text{(property of conditional information)}\\
&= \frac{1}{2}\ln E[(z_t - E[z_t|\varepsilon^t])^2] - \frac{1}{2}\ln E[(z_t-\hat{z}_t)^2] && \text{(Gaussianity)}\\
&= \frac{1}{2}\ln\left(\frac{\delta^2}{1-\delta^2}\right) - \frac{1}{2}E[(z_t-\hat{z}_t)^2]. && (\text{definition of } z_t)
\end{aligned}$$

By rearranging this inequality, we obtain the lower bound stated in the lemma. Under the conjectured solution,

$$z_t - \hat{z}_t = (1-\psi)\frac{\delta L^{-1}}{1-\delta L^{-1}}\varepsilon_t - \sqrt{\psi(1-\psi)}\frac{\delta}{1-\delta L}v_t. \tag{35}$$

Therefore

$$E[(z_t-\hat{z}_t)^2] = (1-\psi)^2\left(\frac{\delta^2}{1-\delta^2}\right) + \psi(1-\psi)\left(\frac{\delta^2}{1-\delta^2}\right) = \left(\frac{\delta^2}{1-\delta^2}\right)e^{-2\kappa},$$

so the conjectured solution attains the lower bound on the objective. □

Lastly, we need to show that the conjectured solution is feasible. First, observe that constraint (iii) is trivially satisfied by the conjecture, since the weights on $\{\varepsilon_t\}$ in (34) are zero for all ε_{t-j}, $j \geq 0$. Second, by combining (34) with (35), the cross autocovariance generating function of the forecast error and the forecast is given by

$$g_{z-\hat{z},\hat{z}}(z) = \psi(1-\psi)\frac{\delta^2}{|1-\delta z|^2} - \psi(1-\psi)\frac{\delta^2}{|1-\delta z|^2} = 0,$$

so constraint (ii) is satisfied as well. To show that the foresight constraint (i) is satisfied, we will need to make use of the following lemma.

Lemma 5. *The Wold representation of $\xi_t \equiv (\hat{z}_t, \varepsilon_t)'$ is given by $\xi_t = \Gamma(L)w_t$, where $\{w_t\}$ is orthonormal Gaussian white noise and*

$$\Gamma(z) = \begin{bmatrix} \sqrt{\psi}\dfrac{\delta}{1-\delta z} & 0 \\ \sqrt{\psi}\dfrac{z-\delta}{1-\delta z} & \sqrt{1-\psi} \end{bmatrix}.$$

Proof. Using the law of motion for $\hat{z}_t$ in (34), the autocovariance generating function of $\{\xi_t\}$ is given by

$$g_\xi(z) = \begin{bmatrix} \psi \dfrac{\delta^2}{|1-\delta z|^2} & \psi \dfrac{\delta z^{-1}}{1-\delta z^{-1}} \\ \psi \dfrac{\delta z}{1-\delta z} & 1 \end{bmatrix}.$$

Now notice that (i) all elements of $\Gamma(z)$ are rational with respect to z, (ii) $\Gamma(z)\Gamma(z)^* = g_\xi(z)$, where the asterisk denotes complex conjugate transposition, (iii) $\Gamma(z)$ is analytic in the unit circle (i.e. the Laurent expansions of each element have no negative powers of z), and (iv) $\Gamma(z)$ is full rank for all $|z|< 1$. Therefore, by Theorem 7 of Rozanov (1960), $\Gamma(z)$ is maximal. By Theorem 4 of that same paper, $\{w_t\}$ is fundamental with respect to $\{\xi_t\}$, which is to say that the representation $\xi_t = \Gamma(L)w_t$ is the Wold representation of $\{\xi_t\}$. □

Using Lemma (5), we can express the joint dynamics of $\hat{z}_t$ and ε_t as

$$\begin{aligned} \hat{z}_t &= \delta \hat{z}_{t-1} + \sqrt{\psi}\delta w_{1,t} \\ \varepsilon_t &= \delta \varepsilon_{t-1} - \sqrt{\psi}(\delta w_{1,t} - w_{1,t-1}) + \sqrt{1-\psi}(w_{2,t} - \delta w_{2,t-1}), \end{aligned}$$

where $w_t = (w_{1,t}, w_{2,t})$ is orthonormal white noise and $\psi = 1 - e^{-2\kappa}$. From this representation, we can compute the optimal j-step ahead forecast error of the income disturbance,

$$\zeta_{t+j|t} = \varepsilon_{t+j} - E_t[\varepsilon_{t+j}] = -\sqrt{\psi}\delta w_{1,t+j} + \sqrt{1-\psi} w_{2,t+j} + \sqrt{\psi}(1-\delta^2) \sum_{k=1}^{j-1} \delta^{j-1-k} w_{1,t+k}.$$

Stacking these up for $j = 1, \ldots, T$,

$$\begin{bmatrix} \zeta_{t+1|t} \\ \zeta_{t+2|t} \\ \zeta_{t+3|t} \\ \vdots \\ \zeta_{t+T|t} \end{bmatrix} = \underbrace{\sqrt{\psi}(1-\delta^2) \begin{bmatrix} \frac{-\delta}{1-\delta^2} & 0 & 0 & \cdots & 0 \\ 1 & \frac{-\delta}{1-\delta^2} & 0 & \cdots & 0 \\ \delta & 1 & \frac{-\delta}{1-\delta^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta^{T-2} & \delta^{T-3} & \delta^{T-4} & \cdots & \frac{-\delta}{1-\delta^2} \end{bmatrix}}_{Q_1} \begin{bmatrix} w_{1,t+1} \\ w_{1,t+2} \\ w_{1,t+3} \\ \vdots \\ w_{1,t+T} \end{bmatrix}$$

$$+ \underbrace{\begin{bmatrix} \sqrt{1-\psi} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{1-\psi} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{1-\psi} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{1-\psi} \end{bmatrix}}_{Q_2} \begin{bmatrix} w_{2,t+1} \\ w_{2,t+2} \\ w_{2,t+3} \\ \vdots \\ w_{2,t+T} \end{bmatrix}$$

From this we can compute the covariance matrix of forecast errors as $\Sigma_T = Q_1 Q_1' + Q_2 Q_2'$. Multiplying these matrices together, we find that

$$\Sigma_T = I_T - \psi(1-\delta^2) b_T b_T',$$

where $b_T = (1, \delta, \delta^2, \ldots, \delta^{T-1})'$. By the matrix determinant lemma,

$$\det \Sigma_T = 1 - \psi(1-\delta^2) b_T' b_T = 1 - \psi(1-\delta^2) \sum_{k=0}^{T-1} \delta^{2k} = 1 - \psi(1-\delta^{2T}).$$

Taking limits as $T \to \infty$, we find

$$\lim_{T\to\infty} I((\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}), \hat{z}^t | \varepsilon^t) = -\frac{1}{2}\ln(1-\psi) = -\frac{1}{2}\ln(e^{-2\kappa}) = \kappa.$$

This proves that constraint (i) is also satisfied. The conjectured solution achieves the lower bound on the objective, and is feasible; therefore, it is optimal. So far we have not formally established that the stated solution is *unique*; however, evidence from brute-force numerical solutions indicates that it is. □

Proof of Corollary (3)**.** From Proposition (3), we know that the consumer's optimal information set is $\text{span}(y^t, \hat{z}^t)$, where

$$\hat{z}_t = \psi \frac{\delta L^{-1}}{1-\delta L^{-1}} \varepsilon_t + \sqrt{\psi(1-\psi)} \frac{\delta}{1-\delta L} v_t.$$

If we define $\tilde{z}_t \equiv \hat{z}_t - \delta \hat{z}_{t-1}$, then $(y^t, \hat{z}^t) = (y^t, \tilde{z}^t)$, since this definition implies that $\hat{z}_t$ is a discounted sum of $\tilde{z}_t, \tilde{z}_{t-1}, \ldots$. Now define the signal

$$s_t \equiv \frac{(1-\delta)h(\delta)}{\psi(1-\delta^2)}(\tilde{z}_t - E[\tilde{z}_t|y^t]) + E[x_t|y^t]. \tag{36}$$

Since both $E[\tilde{z}_t|y^t]$ and $E[x_t|y^t]$ are elements of (y^t), this transformation of $\tilde{z}_t$ preserves the consumer's time-t information set, $(y^t, \hat{z}^t) = (y^t, s^t)$. Now we just need to verify that this definition of s_t coincides with the expression stated in the proposition.

First, we compute the forecast error of $\tilde{z}_t$ based on current and past income. By definition,

$$\tilde{z}_t = \psi(1-\delta L)\frac{\delta L^{-1}}{1-\delta L^{-1}}\varepsilon_t + \sqrt{\psi(1-\psi)}\delta v_t.$$

Therefore,

$$\tilde{z}_t - E[\tilde{z}_t|y^t] = \psi\left[(1-\delta L)\frac{\delta L^{-1}}{1-\delta L^{-1}}\right]_- \varepsilon_t + \sqrt{\psi(1-\psi)}\delta v_t,$$

where $[f(z)]_-$ indicates the principal part of the Laurent series expansion of $f(z)$ around $z = 0$. Since

$$\begin{aligned}(1-\delta z)\frac{\delta z^{-1}}{1-\delta z^{-1}} &= \sum_{j=1}^{\infty}\delta^j z^{-j} - \delta^2\sum_{j=0}^{\infty}\delta^j z^{-j} \\ &= (1-\delta^2)\frac{\delta z^{-1}}{1-\delta z^{-1}} - \delta^2,\end{aligned}$$

we can write

$$\left[(1-\delta L)\frac{\delta L^{-1}}{1-\delta L^{-1}}\right]_- = (1-\delta^2)\frac{\delta L^{-1}}{1-\delta L^{-1}}.$$

Therefore,

$$\tilde{z}_t - E[\tilde{z}_t|y^t] = \psi(1-\delta^2)z_t + \sqrt{\psi(1-\psi)}\delta v_t.$$

Substituting this into (36) and using (32) to substitute out $E[x_t|y^t]$, we find that $s_t = x_t + \sigma_v v_t$, where σ_v^2 is defined as in the proposition. □

Proof of Corollary (4)**.** Income is related to the vector $\xi_t = (\hat{z}_t, \varepsilon_t)'$ by the linear transformation

$$y_t = \begin{bmatrix} 0 & h(L) \end{bmatrix} \xi_t$$

In Lemma (5), we derived the Wold representation of the vector as $\xi_t = \Gamma(L)w_t$. Substituting this into the previous expression, we get

$$y_t = \begin{bmatrix} \sqrt{\psi}\frac{L-\delta}{1-\delta L}h(L) & \sqrt{1-\psi}h(L) \end{bmatrix} w_t.$$

Defining $\eta_t \equiv w_{1,t}$ and $u_t \equiv w_{2,t}$, and substituting $\psi = 1 - e^{-2\kappa}$, we obtain the representation of the income process stated in the proposition. Moreover, since $\text{span}(w^t) = \text{span}(\xi^t)$ by definition of $\{w_t\}$, it follows that $\text{span}(\eta^t, u^t) = \text{span}(\hat{z}^t, y^t)$. □

B Relation to Gaballo (2016)

In this section, we show how our theory of endogenous foresight provides a formal justification for the private signal structure assumed exogenously in Gaballo (2016). The first step is to show that, up to an appropriate linear-quadratic approximation of the agent's objective function, the target variable is next period's aggregate price. The second step is to derive the solution to the optimal foresight problem with this target variable. The third step is to verify that the agent's optimal forecast can be generated by a private signal of the form assumed in Gaballo (2016), which is "next period aggregate price plus i.i.d. noise."

According to the model, overlapping generations of "young" agents work and save to finance their consumption when they become "old". Because agents only plan for two periods, their problem does not involve optimization over infinite sequences, which greatly simplifies the task of determining the optimal target variable. Slightly generalizing the functional forms in Gaballo (2016), the utility maximization problem of the agent is

$$\max_{H_{it}, C_{it+1}} E_{it}\left[u(C_{it+1}) - v(H_{it})\right] \quad \text{s.t} \quad P_{t+1}C_{it+1} = P_{t+1}\omega + RP_t\Theta_{it}^{-1}H_{it}, \tag{37}$$

where P_t is the price of a private consumption good, ω is an exogenous and fixed net tax expenditure, $R = 1$ is the nominal return on the risk free asset, $Q_t \equiv P_t\Theta_{it}^{-1}H_{it}$ is the quantity of the risk free asset, which are financed by labor supplied at the stochastic wage Θ_{it}^{-1}. This wage has independent aggregate and idiosyncratic components,

$$\Theta_{it}^{-1} = \Theta_t^{-1}\Xi_{it}^{-1},$$

where $\theta_t \equiv \ln\Theta_t \overset{\text{iid}}{\sim} N(0, \sigma_\theta^2)$ and $\xi_{it} \equiv \ln\Xi_{it} \overset{\text{iid}}{\sim} N(0, \sigma_\xi^2)$. When making choices, the agent treats the random variables P_t P_{t+1}, and Θ_{it}^{-1} as exogenous, and costlessly observes P_t and Θ_{it}.

Using letters without time subscripts to represent non-stochastic steady-state values, and defining $\eta_{it} \equiv \ln(H_{it}/H)$, $p_t \equiv \ln(P_t/P)$, and $\gamma \equiv (u' + u''H)/(u''H - v''H) >$

0, the following lemma presents a purely quadratic approximation to the agent's objective.

Lemma A.1. *A purely quadratic LQ approximation to the nonlinear problem* (37) *is one in which the agent seeks to maximize the quadratic objective*

$$-\frac{1}{2}E_{it}\left[\eta_{it}^2 - 2\gamma\left(p_{t+1}+\theta_{it}-p_t\right)\eta_{it}\right]. \tag{38}$$

Proof. Substituting the constraint into the objective to eliminate C_{it+1}, the agent's objective is $E_{it}U_{it}$, where

$$U_{it} \equiv u\left(\omega + P_t P_{t+1}^{-1}\Theta_{it}^{-1}H_{it}\right) - v(H_{it}).$$

Next, note that steady-state optimality requires that $u'(\omega+H) = v'(H)$. We can use this optimality condition to eliminate the linear terms in the quadratic approximation of the agent's objective without following the more complicated steps in Benigno and Woodford (2012), as we did in the proof of Lemma (1).

Specifically, a quadratic approximation to U_{it} is

$$\begin{aligned} U_{it} = U + &\underbrace{(u'H - v'H)}_{=0}\eta_{it} \\ &+ \frac{1}{2}\left(u''H^2 - v''H^2 + \underbrace{(u'H - v'H)}_{=0}\right)\eta_{it}^2 + (u'H + u''H)\left(p_t - \theta_{it} - p_{t+1}\right)\eta_{it} \\ &+ \frac{1}{2}\left(u''H^2 + u'H\right)\left(p_t - \theta_{it} - p_{t+1}\right)^2 + \mathcal{O}(\epsilon^3) \end{aligned} \tag{39}$$

The quadratic term in the last line is independent of the agent's policy variables. Removing this term and dividing by $-(u''H^2 - v''H^2)$ gives the desired result. □

Lemma A.2. *Maximizing the quadratic objective* (38) *is equivalent to minimizing the loss function*

$$E\left[(p_{t+1} - E_{it}[p_{t+1}])^2\right].$$

Proof. Optimal labor choice requires that

$$\eta_{it} = \gamma\left(E_{it}[p_{t+1}] + \theta_{it} - p_t\right).$$

Because of the overlapping generations structure, we do not need to conjecture anything about the continuation value of the problem in future periods, as we did in the proof of Lemma (2). Substituting this optimality condition into the objective (38) and simplifying, we get

$$-\frac{1}{2}E_{it}\left[\eta_{it}^2 - 2\gamma\left(p_{t+1}+\theta_{it}-p_t\right)\eta_{it}\right] = -\frac{1}{2}\gamma^2\left(E_{it}[p_{t+1}]-p_{t+1}\right)^2 + \text{t.i.p.}$$

where t.i.p. indicates terms that are independent of the agent's policy variables. Removing this term and dividing by $-\frac{1}{2}\gamma^2$ gives the desired result. □

Consistent with the baseline analysis in Section III of Gaballo (2016), we take $\sigma_\xi^2 \to \infty$, so that Θ_{it} is not informative about the value of the aggregate state variable Θ_t. Given Lemma (A.2), the agent's foresight problem can therefore be written as

$$\min_{\mathcal{I}_{it}} E[(p_{t+1}-E[p_{t+1}|\mathcal{I}_{it}])^2] \quad \text{s.t} \quad \lim_{T\to\infty} I((p_{t+1},\ldots,p_{t+T}),\mathcal{I}_{it}|p^t) \le \kappa \tag{40}$$

The solution to this problem is presented in the following Lemma.

Lemma A.3. *Let $p_t = h(L)w_t$ denote the Wold representation of the equilibrium price process, with $w_t \overset{iid}{\sim} N(0,1)$. Then the forecast process*

$$E[p_{t+1}|\mathcal{I}_{it}] = \frac{h(L)-e^{-2\kappa}h(0)}{L}w_t + \sqrt{e^{-2\kappa}(1-e^{-2\kappa})}h(0)v_{it},$$

with $v_{it} \overset{iid}{\sim} N(0,1)$ and $\{v_{it}\}$ independent of $\{w_t\}$, solves problem (40).

Proof. First, we show that solving problem (40) is equivalent to solving

$$\min_{\mathcal{I}_{it}} E[(w_{t+1}-E[w_{t+1}|\mathcal{I}_{it}])^2] \quad \text{s.t} \quad \lim_{T\to\infty} I((w_{t+1},\ldots,w_{t+T}),\mathcal{I}_{it}|w^t) \le \kappa, \tag{41}$$

where $\{w_t\}$ are the Wold innovations in $\{p_t\}$. The left side of the constraint is the same, since $\text{span}(p^t) = \text{span}(w^t)$ for all t by definition of $\{w_t\}$. With respect to the objective, note that

$$\begin{aligned} p_{t+1}-E[p_{t+1}|\mathcal{I}_{it}] &= (p_{t+1}-E[p_{t+1}|p^t]) - (E[p_{t+1}|\mathcal{I}_{it}]-E[p_{t+1}|p^t]) \\ &= (p_{t+1}-E[p_{t+1}|p^t]) - E[(p_{t+1}-E[p_{t+1}|p^t])|\mathcal{I}_{it}], \end{aligned}$$

since $\mathcal{I}_{it} \supseteq \text{span}(p^t)$. Therefore, it is equivalent to treat $p_{t+1}-E[p_{t+1}|p^t]$ as the target variable. Moreover, we can use the Wold representation of $\{p_t\}$ to write

$$p_{t+1}-E[p_{t+1}|p^t] = \frac{h(L)}{L}w_t - \frac{h(L)-h(0)}{L}w_t = h(0)w_{t+1}.$$

Diving by $h(0)$ does not affect the optimal choice, which means it is also equivalent to treat w_{t+1} as the target variable. This establishes that solving (40) is equivalent to solving (41).

Second, we conjecture that

$$E[w_{t+1}|\mathcal{I}_{it}] = \psi w_{t+1} + \sqrt{\psi(1-\psi)}\upsilon_{it} \equiv \hat{z}_{it}, \tag{42}$$

where $\psi \equiv 1 - e^{-2\kappa}$. To verify this conjecture, we show that it attains a lower bound on the objective function, and it is feasible. The lower bound on the objective function is

$$E[(w_{t+1} - E[w_{t+1}|\mathcal{I}_{it})^2] \geq e^{-2\kappa}.$$

This is because

$$\begin{aligned}\kappa &\geq \lim_{T\to\infty} I((w_{t+1},\dots,w_{t+T}),\mathcal{I}_{it}|w^t)\\ &\geq I(w_{t+1},\mathcal{I}_{it}|w^t)\\ &= \frac{1}{2}\ln E[(w_{t+1} - E[w_{t+1}|w^t])^2) - \frac{1}{2}\ln E[(w_{t+1} - E[w_{t+1}|\mathcal{I}_{it})^2].\end{aligned}$$

Using the fact that $E[w_{t+1}|w^t] = 0$ and rearranging delivers the stated lower bound. Moreover, the conjecture in (42) attains this lower bound, since

$$E[(w_{t+1} - \hat{z}_{it})^2] = (1-\psi)^2 + \psi(1-\psi) = 1 - \psi = e^{-2\kappa}.$$

To verify that the conjecture is feasible, notice first that under the conjecture, the following two rationality restrictions are satisfied:

$$\begin{aligned}E[(w_{t+1} - \hat{z}_{it})w_{t-j}] &= 0 \quad \text{for all } j \geq 0\\ E[(w_{t+1} - \hat{z}_{it})\hat{z}_{i,t-j}] &= 0 \quad \text{for all } j \geq 0.\end{aligned}$$

The first holds by definition of $\{w_t\}$ and because the innovations are independent of $\{\upsilon_{it}\}$, and the second follows from observing that the cross autocovariance generating function between $\{w_{t+1} - \hat{z}_{it}\}$ and $\{\hat{z}_{it}\}$ is zero.

Finally, we need to show that the the foresight constraint is satisfied with equality. To do this, first note that the Wold representation of the vector process $(\hat{z}_{it}, w_t)$ under the conjecture is

$$\begin{bmatrix}\hat{z}_{it}\\ w_t\end{bmatrix} = \begin{bmatrix}\sqrt{\psi} & 0\\ \sqrt{\psi}L & \sqrt{1-\psi}\end{bmatrix}\begin{bmatrix}\upsilon_{1t}\\ \upsilon_{2t}\end{bmatrix},$$

where $\{\upsilon_t\}$ is a two-dimensional orthonormal Gaussian white noise process. Since $\mathcal{I}_{it} = \text{span}(\hat{z}_i^t, w^t)$, we can use this representation to compute the optimal j-step-ahead forecast errors

$$\zeta_{t+1|t} \equiv w_{t+j} - E[w_{t+j}|\mathcal{I}_{it}] = \begin{cases} \sqrt{1-\psi}\upsilon_{2t+1} & j = 1 \\ \sqrt{\psi}\upsilon_{1t+j-1} + \sqrt{1-\psi}\upsilon_{2t+j} & j > 1. \end{cases}$$

Stacking these up for $j = 1, \ldots, T$,

$$\begin{bmatrix} \zeta_{t+1|t} \\ \zeta_{t+2|t} \\ \zeta_{t+3|t} \\ \vdots \\ \zeta_{t+T|t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \sqrt{\psi} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\psi} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{\psi} & 0 \end{bmatrix}}_{Q_1} \begin{bmatrix} \upsilon_{1,t+1} \\ \upsilon_{1,t+2} \\ \upsilon_{1,t+3} \\ \vdots \\ \upsilon_{1,t+T} \end{bmatrix} + \underbrace{\sqrt{1-\psi}I_T}_{Q_2} \begin{bmatrix} \upsilon_{2,t+1} \\ \upsilon_{2,t+2} \\ \upsilon_{2,t+3} \\ \vdots \\ \upsilon_{2,t+T} \end{bmatrix}$$

The covariance matrix associated with these forecast errors is

$$\Sigma_T = Q_1 Q_1' + Q_2 Q_2' = \text{diag}(1-\psi, 1, \ldots, 1),$$

so $\det \Sigma_T = 1 - \psi$ for all values of T. From this it follows that

$$\lim_{T\to\infty} I((w_{t+1}, \ldots, w_{t+T}), \mathcal{I}_{it}|w^t) = -\frac{1}{2}\ln\det\Sigma_T = -\frac{1}{2}\ln(1-\psi) = \kappa.$$

Therefore, we have proven that the conjecture (42) solves problem (41). This implies that

$$E[p_{t+1}|\mathcal{I}_{it}] = h(0)\hat{z}_{it} + E[p_{t+1}|p_t] = h(0)\hat{z}_{it} + \frac{h(L)-h(0)}{L}w_t$$

solves problem (40). Plugging the conjecture (42) into this expression and simplifying gives the desired result. □

Finally, we show that the optimal forecast from Lemma (A.3) is consistent with the signal structure assumed in Gaballo (2016), where each period, in addition to observing p_t, the agent receives a private signal of the form "p_{t+1} plus i.i.d. noise."

Lemma A.4. *The optimal forecast process in* (A.3) *is consistent with the agent having a time-t information set of the form* $\mathcal{I}_t = \text{span}(p^t, s^t)$, *with*

$$s_{it} = p_{t+1} + \sigma_v v_{it},$$

where $\{v_{it}\}$ is orthonormal white noise, independent of $\{p_t\}$, and

$$\sigma_v^2 = \left(\frac{e^{-2\kappa}}{1 - e^{-2\kappa}}\right) h(0)^2.$$

Proof. We know from the proof of Lemma (A.3) that the optimal forecast of w_{t+1} is

$$E[w_{t+1}|\mathcal{I}_{it}] = \psi w_{t+1} + \sqrt{\psi(1-\psi)} v_{it} \equiv \hat{z}_{it},$$

and $\mathcal{I}_{it} = \text{span}(\hat{z}_i^t, w^t)$. Define the private signal

$$s_{it} = \frac{1}{\psi} h_0 \hat{z}_{i,t} + \sum_{j=0}^{\infty} h_{j+1} w_{t-j}.$$

Rescaling $\hat{z}_{it}$ and adding lags of the innovation process $\{w_t\}$ does not change the information set, since $\text{span}(w^t) \subset \mathcal{I}_{it}$. Therefore, $\mathcal{I}_{it} = \text{span}(s_i^t, w^t)$. But then, by substituting in the known law of motion for $\hat{z}_{it}$, it follows that

$$s_{it} = \frac{h(L)}{L} w_t + \sqrt{\frac{1-\psi}{\psi}} h(0) v_{it} = p_{t+1} + \sigma_v v_{it},$$

where σ_v is defined as in the statement of the Lemma. □

C Foresight in state-space models

This section presents a numerical algorithm that can be used to compute the quantity of foresight for a general class of information structures. The algorithm computes the quantity of foresight in an information structure $\{\mathcal{I}_t\}$ such that $\mathcal{I}_t = \text{span}(y^t, x^t)$, where $\{y_t\}$ and $\{x_t\}$ are n_y and n_x dimensional vector processes related by the state-space structure

$$y_t = Ax_t \qquad\qquad x_t = Bx_{t-1} + Ce_t. \tag{43}$$

The n_e dimensional random vector e_t is i.i.d. over time with distribution $N(0, I_{n_e})$.

Separately computing the determinants of the matrices Σ_T and $\hat{\Sigma}_T$ and dividing them, as we did in Section (3), can be numerically unstable. A preferable option is make use of the fact that, with Gaussian random variables, information depends only on the closed linear spaces spanned by each set of random variables; it is independent of the choice of bases in those spaces. This implies that

$$\lim_{T\to\infty} I((y_{t+1},\ldots,y_{t+T}),\mathcal{I}_t|y^t) = \lim_{T\to\infty} I((\varepsilon_{t+1},\ldots,\varepsilon_{t+T}),\mathcal{I}_t|y^t), \tag{44}$$

where ε_t is the n_ε dimensional disturbance in the Wold representation of the process $\{y_t\}$. By definition, it is i.i.d. over time with distribution $N(0, I_{n_\varepsilon})$, and its current and past values at each point in time form an orthonormal basis for span(y^t). The equality in (44) says that the amount of information about the future values of the process $\{y_t\}$ is the same as the amount of information about the future values of the disturbances $\{\varepsilon_t\}$.

The reason this is helpful is because, without foresight, the disturbance ε_{t+j} is, by definition, completely unforecastable for any $j > 0$. Therefore, the covariance matrix of forecast errors without foresight reduces to the identity matrix, which has a determinant of one. Combining (1) and (44), we can express conditional mutual information in terms of the determinant of one matrix,

$$I((y_{t+1}, \ldots, y_{t+T}), \mathcal{I}_t | y^t) = -\frac{1}{2} \ln \det \tilde{\Sigma}_T,$$

where $\tilde{\Sigma}_T \equiv \text{var}((\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}) - E[(\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}) | \mathcal{I}_t])$.

It is well known that, given a state-space structure of the form in (43), it is possible to relate $\{\varepsilon_t\}$ to $\{e_t\}$ through a state-space structure of the form (see, e.g. the discussion in ch. 8 of Hansen and Sargent 2014),

$$\varepsilon_t = \underbrace{\begin{bmatrix} V^{-1/2}AB & V^{-1/2}A \end{bmatrix}}_{\tilde{A}} \tilde{x}_{t-1} + \underbrace{V^{-1/2}AC}_{\tilde{D}} e_t \tag{45}$$

$$\tilde{x}_t = \underbrace{\begin{bmatrix} B & 0_{n_x} \\ KAB & B - KA \end{bmatrix}}_{\tilde{B}} \tilde{x}_{t-1} + \underbrace{\begin{bmatrix} C \\ KAC \end{bmatrix}}_{\tilde{C}} e_t,$$

where $V \equiv APA'$, $K \equiv BPA'V^{-1}$, and P solves the Riccati equation

$$P = BPB' + CC' - BPA'(APA')^{-1}APB'.$$

Using this system, the j-step-ahead forecast error is

$$\zeta_{t+j|t} = \varepsilon_{t+j} - E[\varepsilon_{t+j} | y^t, x^t] = \tilde{D} e_{t+j} + \sum_{k=1}^{j-1} \tilde{A} \tilde{B}^{T-1-k} \tilde{C} e_{t+k}.$$

Stacking these up for $j = 1, \ldots, T$,

$$\begin{bmatrix} \zeta_{t+1|t} \\ \zeta_{t+2|t} \\ \zeta_{t+3|t} \\ \vdots \\ \zeta_{t+T|t} \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{D} & 0 & 0 & \cdots & 0 \\ \tilde{A}\tilde{C} & \tilde{D} & 0 & \cdots & 0 \\ \tilde{A}\tilde{B}\tilde{C} & \tilde{A}\tilde{C} & \tilde{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{A}\tilde{B}^{T-2}\tilde{C} & \tilde{A}\tilde{B}^{T-3}\tilde{C} & \tilde{A}\tilde{B}^{T-4}\tilde{C} & \cdots & \tilde{D} \end{bmatrix}}_{\tilde{Q}} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+2} \\ \varepsilon_{t+3} \\ \vdots \\ \varepsilon_{t+T} \end{bmatrix}.$$

From this we can see that $\tilde{\Sigma}_T = \tilde{Q}\tilde{Q}'$, so to approximate the limit of $-\frac{1}{2} \ln \det \tilde{\Sigma}_T$ as $T \to \infty$, we can compute this quantity for successively larger values of T, until the incremental change in information, $-\frac{1}{2} \ln(\det \tilde{\Sigma}_T / \det \tilde{\Sigma}_{T-1})$, becomes acceptably close to zero. (One limitation of this algorithm is that it could run into problems if information converges at a slower rate than the growth in the computational burden of taking the determinant of a $T \times T$ matrix. For all the examples we have considered, however, this has not been an issue.) What follows is a simple Matlab function that implements this algorithm.

```
function f = foresight(A,B,C)
% -------------------------------------------------------------------------
% Numerically compute the amount of foresight in the state-space model
%   y(t) = A*x(t)   x(t) = B*x(t-1) + C*e(t)    e(t)~N(0,I)
% -------------------------------------------------------------------------

% State-space form for innovations
P     = dare(B',A',C*C',0);
V     = A*P*A.';
K     = B*P*A'/V;
sqrtV = chol(V,'lower');
A1    = sqrtV\[A*B,A];
B1    = [B,zeros(size(B));K*A*B,B-K*A];
C1    = [C;K*A*C];
D1    = sqrtV\A*C;

% Iterate to convergence
tol  = 1e-6;
err  = 1;
Tmax = 1e3;
Q    = D1;
f    = -1/2*log(det(Q*Q.'));
B1p  = 1;
T    = 2;
ny   = size(A,1);
ne   = size(C,2);
while err > tol && T < Tmax
    Q    = [Q,zeros(ny*(T-1),ne);A1*B1p*C1,Q(end-ny+1:end,:)];
    fnew = -1/2*log(det(Q*Q.'));
    err  = abs(fnew-f);
    f    = fnew;
    B1p  = B1*B1p;
    T    = T + 1;
end
end
```

References

Benigno, P. and M. Woodford (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory 147*(1), 1-42.

Gaballo, G. (2016). Rational inattention to news: The perils of forward guidance. *American Economic Journal: Macroeconomics 8*(1), 42-97.

Gantmacher, F. R. and M. G. Krein (2002). *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems: Revised Edition.* Providence, RI: AMS Chelsea Publishing.

Hansen, L. P. and T. J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control 2*, 7-46.

Hansen, L. P. and T. J. Sargent (2014). *Recursive Models of Dynamic Linear Economies.* Princeton, NJ: Princeton University Press.

Rozanov, Y. A. (1960). Spectral properties of multivariate stationary processes and boundary properties of analytic matrices. *Theory of Probability and its Applications 5*(4), 362-376.

Whittle, P. (1983). *Prediction and Regulation by Linear Least-Square Methods.* Minneapolis, MN: University of Minnesota Press.