1 Introduction

A popular hypothesis in macroeconomics and nance 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 rst 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 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 rst 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 rst 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 Mackowiak et al. (2018a) 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 Mackowiak et al. (2018b). 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

2 De ning foresight

This section de nes 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

It is also worth noting that foresight is a *directed* measure of information flow. To understand this, suppose that 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.

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

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

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

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

$$\hat{\Sigma}_T \equiv \mathrm{var}((y_{t+1},\dots,y_{t+T}) - E[(y_{t+1},\dots,y_{t+T})|I_t, y^t])$$
$$\Sigma_T \equiv \mathrm{var}((y_{t+1},\dots,y_{t+T}) - E[(y_{t+1},\dots,y_{t+T})|y^t]).$$

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

$$\lim_{T\to\infty} I((\varepsilon_{t+1},\dots,\varepsilon_{t+T}); I_t | \varepsilon^t) = 0 \tag{2}$$

for all possible information structures $\{I_t\}$.[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,

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

where $\sigma_u, \sigma_\varepsilon > 0$, $0 < \rho < 1$, and $\{u_t\}$ and $\{\varepsilon_t\}$ 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, $I_t = \text{span}(\varepsilon^t, u^t)$.

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

$$\hat{e}_{t+j|t} = y_{t+j} - E[y_{t+j}|I_t] = \sigma_u u_{t+j} + \sum_{k=1}^{j} \rho^{j-k} \sigma_\varepsilon \varepsilon_{t+k}.$$

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

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

From this we can see that $\wedge_T = Q_u Q_u^{\prime} + Q \ Q^{\prime}$

since $\quad_T = Q_\varepsilon Q_\varepsilon'$, 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 $(y_{t+1}; :::; y_{t+T})$. Plugging (4) and (6) into (1) and using the fact that $\ \ ^2_\varepsilon = \ \ ^2_u = \ $, we nd that

$$I((y_{t+1}; :::; y_{t+T}); (y^t; z^t)|y^t) = \quad \frac{1}{2} \ln \ (1 \quad r^2) + r^2 \ \ ^{2T} \ \ ;$$

Since $| \ | < 1$, 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*

$$\lim_{T \to \infty} I((y_{t+1}; :::; y_{t+T}); (\ ^t; u^t)|y^t) = \quad \frac{1}{2} \ln(1 \quad r^2);$$

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

$$r^2 \quad \frac{(1 + \ \ ^2 = \ ^2_u)}{(1 \quad \ ^2)}$$

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

$$\frac{1}{2} \ \ 1 + \ \ ^2 + \ \ ^2 = \ ^2_u \qquad \sqrt{(1 + \ \ ^2 + \ \ ^2 = \ ^2_u)^2 \quad 4 \ \ ^2} \ \ :$$

Notice that the quantity of foresight in this representation depends only on and the ratio $\ ^2 = \ ^2_u$. 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 with limiting values $\lim_{\ \to 0} F = 0$ *and* $\lim_{\ \to \infty} F = \quad \frac{1}{2} \ln(1 \quad r^2)$, *where*

$$r^2 \quad \frac{1 \quad (1 + \ \ ^2 = \ ^2_u)}{1 \quad \ ^2}$$

and

$$\frac{1}{2} \ \ 2 + \ \ ^2 = \ ^2_u \ \ =$$

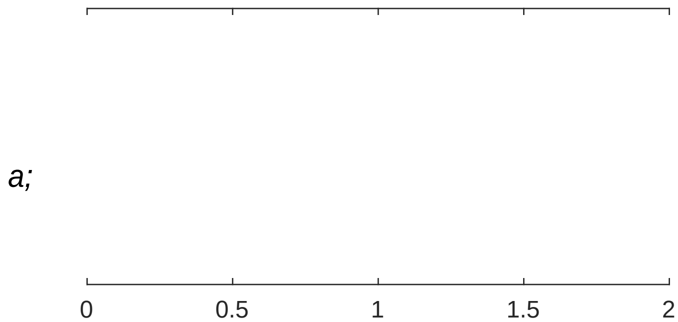

Figure 1: Quantity of foresight in the persistent-transitory representation. The circle shows the amount of information at the baseline parameter values $= 0.9$, ${}^2_u = 0.01$, and ${}^2 = 0.003$.

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

= 0:90
27]TJ/F3011.9552Tf128.5.610-3.3310d[(]TJ/F2711.9552Tf9.3590Td[(=)2780]TJ/F3011.9552Tf18.2780Td[(]TJ/F2711.9552Tf3.2520Td[(0,]T

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 span($^t; u^t$) = span($y^t; s^t$), where[8]

$$s_t = (1 \quad) \sum_{j=0}^{\infty} {}^{j} y_{t+j} + {}_{v} v_t; \qquad (7)$$

${}_v = (1 \quad)\ '' \ {}_u$

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 e ect 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 di erent signal that places more weight on future income,

$$\tilde{s}_t = (1 \quad \tilde{\ }) \sum_{j=0}^{\infty} \tilde{\ }^j y_{t+j} + \ _v v_t;$$

with $\tilde{\ } >$ (and $_v$ is de ned as before). The change in information from I_t = span$(y^t; s^t)$ to $\tilde{I}_t$ = 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 xed at (5).

We summarize this discussion with a proposition. It is not di cult 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* $\{I_t\}$*,* I_t *span*$(\ ^t; u^t)$*. Then* $I_t = span(y^t; s^t)$ *when*

$$y_t = \ y_{t-1} + \ _{\varepsilon}(\varepsilon_t \quad \varepsilon_{t-1})$$

$$s_t = (1 \quad) \sum_{j=0}^{\infty} \ ^j y_{t+j} + \ _v v_t;$$

where $\{\varepsilon_t\}$ *and* $\{v_t\}$ *are independent orthonormal Gaussian white noise processes,* $_{\varepsilon} \quad _u \sqrt{\quad =}$ *,* $_v \quad (1 \quad) \ _{\varepsilon} \ _u =$ *, and is de ned as in Proposition* (1)*.*

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

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

(i) *is monotonically increasing in* *with limiting values* $\lim_{\,!\,0}$ = 0 *and* $\lim_{\,!\,1}$ = *, where* *is de ned as in Corollary* (1)*.*

(ii) *is monotonically decreasing in* $^2 = {}^2_u$ *with limiting values* $\lim_{\,^2= \,^2_u \,!\,0}$ = *and* $\lim_{\,^2= \,^2_u \,!\,1}$ = 0*.*

The intuition behind part (i) of the corollary is that as 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 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 $^2 = {}^2_u$ 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 rst case, is large but the signal is uninformative because it becomes in nitely noisy, ${}^2_v = {}^2_{"} = {}^2_u = {}^2 \; ! \; 1$. In the second case, the signal is uninformative because it places no weight on future income, $!$ 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 = 0:56. Figure (2) illustrates the e ects of changing the parameters on the magnitude of . Each line shows what happens as the ratio $^2 = {}^2_u$ varies over the range of values on the horizontal axis. The di erent lines show these e ects for di erent values of . From the gure, we can see both how is monotonically increasing in and monotonically decreasing in $^2 = {}^2_u$. We can also see how the point at which each line crosses the vertical axis is equal to the corresponding value of .

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 $\sigma^2 = 0.003$.

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

$$C_t + B_t \le (1 + r_{t-1})B_{t-1} + Y_t;$$

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 $\lim_{t \to \infty} \prod_{s=0}^{t-1}(1 + r_s)^{-1} B_t \ge 0.$

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 *average of current and future labor income,*

$$x_t \quad (1 \quad) \sum_{j=0}^{\infty} {}^{j} y_{t+j}; \qquad (9)$$

and the discounting parameter $0 < \quad < \quad$ *is*

$$\frac{1}{2}\left[1 + \quad + \quad {}^2 C \quad \sqrt{(1 + \quad + \quad {}^2 C)^2 \quad 4 \quad} \right] :$$

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]

$$c_t = \frac{1}{C}\left[(1 \quad) \quad {}^{1} b_{t \; 1} + Y E_t[x_t] \right] : \qquad (10)$$

The term $\quad {}^{1} b_{t \; 1}$ is the total nancial 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 , which is less than , due to the fact that interest rates are savings-elastic. In the limit as $\quad \to 0$, it is possible to show that $\quad \to \quad$. 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,

$$c_t = \frac{1}{C}\left[(1 \quad) \quad {}^{1} E_t[b_{t \; 1}] + Y E_t[x_t] \right] :$$

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 nds 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 rst show that it is possible to rewrite the consumer's utility maximization problem as a tracking problem in terms of the target variable the constraint. Using this observation, we can re-write the consumer's information problem as:

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

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

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

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

where $x_t = (1-\beta)(1-\beta L^{-1})^{-1}h(L)\varepsilon_t$, with $\varepsilon_t \overset{iid}{\sim} N(0,1)$. The first constraint is the foresight constraint, after imposing $I_t = \text{span}(y^t, \hat{x}^t)$. 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 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*

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

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

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

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

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

which is the perfect foresight solution. For intermediate values of , the optimal 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* $I_t = span(\;^t; u^t)$ *with*

$$y_t = \sqrt{1 \quad e^{\;2}} \quad \frac{L}{1 \quad L} \quad h(L)\;_t + e \quad h(L)u_t;$$

where fu_tg *and* $f\;_tg$ *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 rst component has an additional dynamic term which depends on the magnitude of the economic parameter .

So far we have characterized the solution to the consumer's foresight problem, but we have not explored how his optimal information choice a ects 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

$$= 0{:}95; \qquad = 0{:}5; \quad \text{and} \qquad = 0{:}01;$$

and consider a range of di erent values for the informational parameter .

Figure (3) shows the impulse response functions associated with each of the two model disturbances, the income disturbance $''_t$ 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 a ects income over time. The disturbance has no e ect on income before it occurs, it has its largest

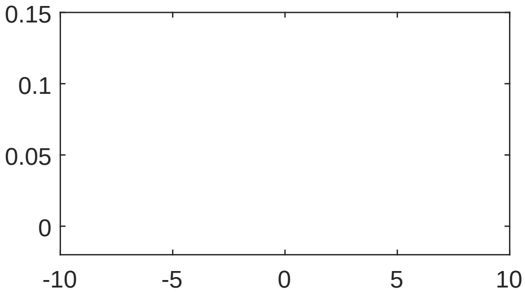

Figure 3: Impulse responses when $= 0{:}95$, $= 0{:}5$, $= 0{:}01$, and income follows the ARMA(1,1) process from Proposition (2) with $= 0{:}9$, ${}^{2}_{u} = 0{:}01$, and ${}^{2} = 0{:}003$. The last row refers to the consumer's forecast of the target variable x_t, de ned in (9), which is an exponential moving average of current and future income.

e ect on impact, and then it has a smaller e ect in subsequent periods as it decays at rate

rst 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 = 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 and . Under optimal foresight, Corollary (3) indicates that the consumer e ectively constructs a signal which ...

fomactive (b) -27 (utn) -420 (the) -33 (n (ear) -420futgure) -420 (tan,) -343 (the) -321 (consume
income disturbanceoccurt, aseseen inthe botomnpanel oftheleft
Iofonseand

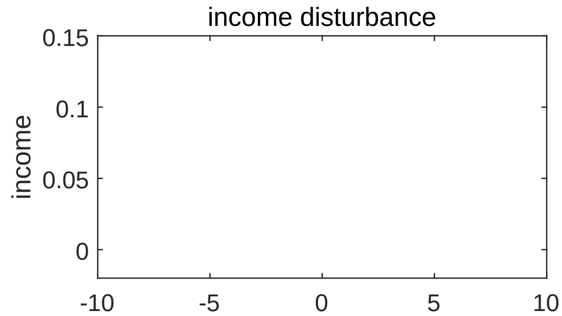

Figure 5: Impulse responses when = 0:95, = 0:5, = 0:01, = 0:1625, = 0:9, ^{2} = 0:003, $_u^2$ = 0:01 and $h(L)$ = (1 L)(1 L) ^{1}. The value of is chosen to keep the amount of foresight in both information structures the same. The last row refers to the consumer's forecast of the target variable x_t, de ned in (9), which is an exponential moving average of current and future income.

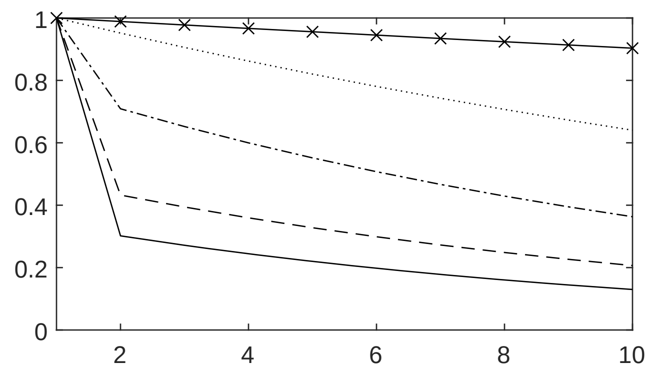

Figure 6: Misspecified estimates of the autocorrelation function of income.

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 . When = 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 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 insu cient 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 simpli es 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 uctuations. *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*

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 identi cation 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.

Mackowiak, B., F. Matejka, and M. Wiederholt (2018b). Dynamic Rational Inattention: Analytical Results. *Journal of Economic Theory 176*, 650{692.

Mackowiak, B., F. Matejka, and M. Wiederholt (2018a). 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 e ects 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 $\hat{}_T$ given in (4), and then to show that 0 r^2 1. Regarding the rst of these, notice that

$$\det \hat{}_T = \det(I_T + C'(C_u C_u')^{1} C\ ) \det(C_u C_u') \qquad \text{(matrix determinant lemma)}$$

$$= \det \quad I_T + \frac{1}{{}^2_u} C\ C' \quad {}^{2T}_u \qquad (C_u C_u' = {}^2_u I_T)$$

$$= \det \quad I_T + aA_T^{\ 1} \quad {}^{2T}_u \qquad (C\ C' = {}^2 A_T^{\ 1}, \text{ with } a \text{ and } A_T \text{ de ned below})$$

$$= \det (A_T + aI_T$$

alwhat r859 -7.892 Td [(u) -3859 + (al3uk) 2p1 Tf -03 y3x8859 -7.892d52 Tf 6.605 1.793276

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

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

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

$$d_T = c_1 \ {}_1^T + c_2 \ {}_2^T; \tag{14}$$

where $_1$ and $_2$ are the two roots of the polynomial $P(\ ) = {}^2 \quad (1 + {}^2 + a) \quad + {}^2$ and c_1 and c_2 are chosen to satisfy the initial conditions. Based on the de nition of in Proposition (2), we can see that the two roots of $P(\ )$ must be

$$_1 = - \quad \text{and} \quad _2 = \quad :$$

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

$$c_1 = 1 \quad r^2 \quad \text{and} \quad c_2 = r^2; \quad \text{where} \quad r^2 \quad \frac{(1 + a)}{(1 \quad ^2)}:$$

Plugging the expressions for $_1$, $_2$, 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 r^2 1, rst observe that 1 $(1 + a)=$. Multiplying both sides by , we nd ^{2} $(1 + a)$, which implies that r^2 1. To show that r^2 0, we need to prove that

$$(1 + a): \tag{15}$$

By the de nition of , = is the smaller root of the polynomial $P(z) = {}^2 z^2 \quad (1 + {}^2 + a)z + 1$. Since $P(0) = 1 > 0$, $P(1) = \quad a \quad 0$, and $P(z) > 0$ as $z \ ! \ 1$, it follows that the two roots of this polynomial satisfy $0 \quad z_1 \quad 1 \quad z_2$. Now notice that

$$P \quad \frac{1}{1 + a} \quad = \quad {}^2 \frac{a}{(1 + a)^2} \quad 0;$$

so it must also be true that $z_1 = \ = \ 1{=}(1 + a)$, which proves (15). □

Proof of Corollary (1). De ne the ratio $a \quad {}^2 = {}^2_u$ so that

$$r^2 = \frac{(1 + a)}{(1 \quad ^2)}:$$

Di erentiating with respect to ,

$$\frac{@r^2}{@} = \frac{(1 + a)(\quad ^0) + 2 \quad ^0 (\quad (1 + a))}{^2(1 \quad ^2)};$$

where ^{0} @ =@

Proof of Proposition (2). According to the persistent-transitory representation in (3), the autocovariance generating function of fy_tg is given by

$$g_y(z) = \frac{{}^2}{j1 \qquad zj^2} + {}^2_u = \frac{{}^2 + {}^2_u j1 \qquad zj^2}{j1 \qquad zj^2}:$$

Factoring the numerator,

$${}^2 + {}^2_u j1 \qquad zj^2 = {}^2$$

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

$$z_t = \frac{\sigma^2}{\sigma_u^2}\frac{\lambda/\beta}{(1-\lambda L)(1-\lambda\beta L^{-1})}y_t + \frac{\sqrt{\lambda/\beta}}{1-\lambda L}v_t,$$

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

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

where the second lines substitutes in the previous expression for z_t, and the third line defines $\sigma_v^2 \equiv (1-\lambda\beta)^2\sigma_\varepsilon^2\sigma_u^2/\sigma^2$. Because $0 < \lambda < 1$ and $y_t \in \mathcal{H}(y^t)$, 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 $a \equiv \sigma^2/\sigma_u^2$ so that

$$\lambda = \frac{1}{2}\left[1+\beta^2+a - \sqrt{(1+\beta^2+a)^2 - 4\beta^2}\right].$$

Differentiating with respect to β,

$$\frac{\partial\lambda}{\partial\beta} = \beta\left[1 - \frac{(1-\beta^2+a)}{\sqrt{(1+\beta^2+a)^2-4\beta^2}}\right] \geq 0,$$

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

$$\lim_{\beta\to 1}\lambda = \frac{1}{2}\left[2+a - \sqrt{(2+a)^2-4}\right].$$

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

$$\lim_{\beta\to 0}\lambda = \lim_{\beta\to 0}\beta\left(1 - \sqrt{\frac{(1+\beta^2+a)^2}{(1+\beta^2+a)^2-4\beta^2}}\right) = 0.$$

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

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

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

$$\lim_{a \to 0} \; = \frac{1}{2}\left(1 + {}^{2} \quad \sqrt{(1 \quad {}^{2})^{2}}\right) = \; :$$

To nd the limit as a approaches in nity, it is easiest to notice that

$$\frac{1}{\;} = \frac{1}{2}\left(1 + {}^{2} + a + \sqrt{(1 + {}^{2} + a)^{2} \quad 4 \, {}^{2}}\right) ;$$

since the polynomial $P(z) = \; z^2 \quad (1 + {}^2 + a)z + \;$ has reciprocal roots. Because the right side becomes in nite as a does, it follows that approaches zero.

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

$$V_0 = E_0 \sum_{t=0}^{\infty} {}^{t} \quad u(C) + \frac{1}{2}u''(C)C^2 c_t^2 \quad \frac{1}{2}u'(C) \quad b_t^2 + u'(C) \quad Y y_t + \frac{1}{2} Y y_t^2 \quad + O({}^3)$$

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 speci c terms will become clear in the proof of Lemma (2), but at this point it is su cient to observe that they do not a ect the consumer's rankings of alternative plans. □

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

$$c_t = E_t[c_{t+1}] + \quad b_t \qquad (21)$$

$$C c_t + b_t = \quad {}^{1} b_{t \quad 1} + Y y_t$$

Using this factorization, we can write (23) as

$$E_t[(1 \quad {}_1 L)(1 \quad {}_2 L) b_{t+1}] = Y E_t[y_{t+1} \quad y_t];$$

or as

$$q_t = \quad {}_2{}^{1} E_t[q_{t+1}] \quad {}_2{}^{1} Y E_t[y_{t+1} \quad y_t];$$

where $q_t \quad (1 \quad {}_1 L) b_t$. Because $j \ {}_2{}^{1} j < 1$, this can be solved forward and rewritten in terms of b_t to get

$$b_t = \quad {}_1 b_{t \ 1} \quad {}_2{}^{1} Y \sum_{j=0}^{\infty} \ {}_2{}^{j} E_t[y_{t+j+1} \quad y_{t+j}]: \tag{25}$$

Now, notice that

$$\sum_{j=0}^{\infty} \ {}_2{}^{j} E_t[y_{t+j+1} \quad y_{t+j}] = \quad {}_2 \sum_{j=1}^{\infty} \ {}_2{}^{j} y_{t+j} \quad y_t \quad \sum_{j=1}^{\infty} \ {}_2{}^{j} y_{t+j}$$

$$= \quad {}_2 y_t + (\ {}_2 \quad 1) \sum_{j=0}^{\infty} \ {}_2{}^{j} y_{t+j}:$$

Substituting this into (25), we get

$$b_t = \quad {}_1 b_{t \ 1} + Y y_t \quad (1 \quad {}_2{}^{1}) Y \sum_{j=0}^{\infty} \ {}_2{}^{j} E_t[y_{t+j}]: \tag{26}$$

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

$$Cc_t = (1 \quad {}_2{}^{1}) \quad {}^{1} b_{t \ 1} + (1 \quad {}_2{}^{1}) Y \sum_{j=0}^{\infty} \ {}_2{}^{j} E_t[y_{t+j}];$$

This is the consumption function from (10) when we de ne $\quad 1 = \ {}_2$. □

Proof of Lemma (2). We begin by nding a closed-form expression for the continuation utility V_t

To verify the expression in (27), let

$$u_t \quad \frac{1}{2} \quad c_t{}^2 + \frac{\quad}{C} b_t{}^2 + \quad \frac{\quad}{2} \quad \frac{Y^2}{C^2} \; (1 \quad) y_t^2 \quad 2 y_t x_t$$

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

$$b_t = -b_{t \;\; 1} + Y y_t \quad Y x_t \tag{29}$$

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

$$V_t = u_t + \quad V_{t+1}.$$

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

$$x_{t+1} = \quad {}^1 (x_t \quad (1 \quad) C$$ TJ/F30 11.9552 Tf 3.5563552 0 Td [(x)]TJ/F62

Expanding the right side of this equation, we nd that the equation is satis ed if and only if ! = 1= . Therefore, by (30),

$$E[V_0] = E[V_0\ ] \quad \frac{1}{\ } E[D_0]$$

Finally, since $E[V_0\ ]$ is independent of the consumer's choices, maximization of $E[V$

Therefore, $E[(x_t \quad \hat{x}_t)^2] = (1 \quad)^2 h(\;)^2 E[(z_t \quad \hat{z}_t)^2]:$ 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

$$E[x_t|y^t] = (1 \quad) \frac{h(L) \quad L^{\;1} h(\;)}{1 \quad L^{\;1}} "_t; \tag{33}$$

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 I_t and rearrange to find the implied relationship between $\hat{x}_t$ and $\hat{z}_t$ stated in the lemma,

$$\hat{x}_t = E[x_t|y^t] + (1 \quad) h(\;) \hat{z}_t:$$

Proof. Observe that

$$= \lim_{T \to \infty} I((\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}); \hat{z}^t | \varepsilon^t) \quad \text{(foresight constraint)}$$

$$= \lim_{T \to \infty} I((z_t, \ldots, z_{t+T}); \hat{z}^t | \varepsilon^t) \quad (\text{since } \varepsilon_{t+1} = z_t = \quad z_{t+1})$$

$$I(z_t; \hat{z}^t | \varepsilon^t) \quad \text{(property of conditional information)}$$

$$= \frac{1}{2} \ln E[(z_t \quad E[z_t | \varepsilon^t])^2] \quad \frac{1}{2} \ln E[(z_t \quad \hat{z}_t)^2] \quad \text{(Gaussianity)}$$

$$= \frac{1}{2} \ln \frac{\;^2}{1 \quad ^2} \quad \frac{1}{2} E[(z_t \quad \hat{z}_t)^2]. \quad (\text{definition of } z_t)$$

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

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

Therefore

$$E[(z_t \quad \hat{z}_t)^2 = (1 \quad)^2 \frac{}{1 \quad ^2} + (1 \quad) \frac{^2}{1 \quad ^2} = \frac{^2}{1 \quad ^2} e^{\;2};$$

4Tf 4.732 -4.936[(t)]

so the conjectured solution attains the 552 Tf 6.r41 -32.611 Td [(so)].51 0 Td [( ).51 0 7n1n327(552 Tf

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

$$g(z) = \begin{matrix} 2 \\ 6 \\ 6 \\ 6 \\ 4 \end{matrix} \quad \frac{2}{}$$

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

$$\begin{bmatrix} {}_{t+1|t} \\ {}_{t+2|t} \\ {}_{t+3|t} \\ \end{bmatrix}$$

If we de ne $\bar{z}_t \quad \hat{z}_t \quad \hat{z}_{t\ \ 1}$, then $(y^t;\hat{z}^t) = (y^t;\bar{z}^t)$, since this de nition implies that $\hat{z}_t$ is a discounted sum of $\bar{z}_t;\bar{z}_{t\ \ 1};:::$. Now de ne the signal

$$s_t \quad (1$$

De ning $_t$ $w_{1;t}$ and u_t $w_{2;t}$, and substituting

$v''H) > 0$, 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*

$$\frac{1}{2}E_{it}\; h_{it}^2 \quad 2\; (p_{t+1} + \quad_{it} \quad p_t)\, h_{it}\;: \tag{38}$$

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

$$U_{it} \quad u\;\; w + P_t P_{t+1}^{\;1} \quad_{it}^{\;1} H_{it} \quad v(H_{it}):$$

Next, note that steady-state optimality requires that $u'(w + H) = v'(H)$. 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).

Speci cally, a quadratic approximation to U_{it} is

$$\begin{aligned} U_{it} = U &+ (\underbrace{u'H \quad v'H}_{=0})\,h_{it} \\ &+ \frac{1}{2}\left(u''H^2 \quad v''H^2 + (\underbrace{u'H \quad v'H}_{=0})\right) h_{it}^2 + (u'H + u''H)\,(p_t \quad_{it} \quad p_{t+1})h_{it} \\ &+ \frac{1}{2}\;\; u''H^2 + u'H\;\; (p_t \quad_{it} \quad p_{t+1})^2 + O(\;^3) \end{aligned} \tag{39}$$

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

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

$$E\;\; (p_{t+1} \quad E_{it}[p_{t+1}])$$

proof of Lemma (2). Substituting this optimality condition into the objective (38) and simplifying, we get

$$\frac{1}{2}E_{it}\ \ h_{it}^2 \quad 2 \ \ (p_{t+1} + \ \ _{it} \quad p_t)\, h_{it} \ \ = \ \ \frac{1}{2}\ \ ^2\,(E_{it}[p_{t+1}] \quad p_{t+1})^2 + \text{t.i.p.}$$

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

Consistent with the baseline analysis in Section III of Gaballo (2016), we take $^2\ \ !\ \ 1$, so that $_{it}$ is not informative about the value of the aggregate state variable $_t$. Given Lemma (A.2), the agent's foresight problem can therefore be written as

$$\min_{I_{it}} E[(p_{t+1} \quad E[p_{t+1} j I_{it}])^2] \quad \text{s.t} \quad \lim_{T\ !\ 1} I((p_{t+1}; : : : ; p_{t+T}); I_{it} j p^t) \qquad (40)$$

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 $w_t \overset{iid}{\ } N(0;1)$. Then the forecast process*

$$E[p_{t+1} j I_{it}] = \frac{h(L) \quad e^{\ \ 2}\, h(0)}{L} w_t + \text{p}\overline{e^{\ \ 2}\ (1 \quad e^{\ \ 2}\ )}\, h(0) v_{it};$$

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

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

$$\min_{I_{it}} E[(w_{t+1} \quad E[w_{t+1} j I_{it}])^2] \quad \text{s.t} \quad \lim_{T\ !\ 1} I((w_{t+1}; : : : ; w_{t+T}); I_{it} j w^t) \qquad ; \qquad (41)$$

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

$$\begin{aligned} p_{t+1} \quad E[p_{t+1} j I_{it}] &= (p_{t+1} \quad E[p_{t+1} j p^t]) \quad (E[p_{t+1} j I_{it}] \quad E[p_{t+1} j p^t]) \\ &= (p_{t+1} \quad E[p_{t+1} j p^t]) \quad E[(p_{t+1} \quad E[p_{t+1} j p^t]) j I_{it}]; \end{aligned}$$

since I_{it} span(p^t). Therefore, it is equivalent to treat $p_{t+1} \quad E[p_{t+1} j p^t]$ as the target variable. Moreover, we can use the Wold representation of $f p_t g$ to write

$$p_{t+1} \quad E[p_{t+1} j p^t] = \frac{h(L)}{L} w_t \quad \frac{h(L) \quad h(0)}{L} w_t = h(0) w_{t+1}:$$

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

$$E[w_{t+1}|I_{it}] = \rho w_{t+1} + \sqrt{\rho(1-\rho)}\, v_{it} \equiv \hat{z}_{it}, \tag{42}$$

where $\rho \equiv 1 - e^{-2\kappa}$. 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

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

This is because

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

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

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

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 $\{v_{it}\}$

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

$$\sigma_v^2 = \frac{\sigma_e^{2}}{1-\sigma_e^{2}}\, h(0)^2.$$

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

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

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

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

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

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

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 $\{I_t\}$ such that $I_t = \text{span}(y^t, x^t)$, where $\{y_t\}$ and $\{x_t\}$ are n_y and n_x dimensional vector processes related by the state-space structure

$$y_t = A x_t \qquad x_t = B x_{t-1} + C e_t. \tag{43}$$

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 Ω_T and $\hat{\Omega}_T$ 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

$$\lim_{T\to\infty} I((y_{t+1},\ldots,y_{t+T}); I_t | y_{it}$$

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,

$$I((y_{t+1}, \ldots, y_{t+T}); I_t | y^t) = \quad 1$$

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

$$\begin{bmatrix} _{t+1|t} \\ _{t+2|t} \\ _{t+3|t} \\ \vdots \\ _{t+T|t} \end{bmatrix} = \begin{bmatrix} D & 0 & 0 & & 0 \\ AC & D & 0 & & 0 \\ ABC & AC & D & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ AB^{T\ 2}C & & & & \end{bmatrix}$$

```
function f = foresight(A,B,C)
% -------------------------------------------------------------------------
% Numerically compute the amount of foresight in the state-space model
```

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.