Econometrica Supplementary Material

SUPPLEMENT TO "MULTIVARIATE RATIONAL INATTENTION"

(*Econometrica*, Vol. 90, No. 2, March 2022, 907–945)

JIANJUN MIAO
Department of Economics, Boston University

JIERAN WU
Academy of Financial Research and College of Economics, Zhejiang University

ERIC R. YOUNG
Department of Economics, University of Virginia and Research Department, Federal Reserve Bank of Cleveland

APPENDIX B: PROOFS

PROOF OF PROPOSITION 1: FOR SIMPLICITY, we omit the time t subscript for all variables in the proof. By the singular-value decomposition of a positive semidefinite matrix, there exists an $n_x \times n_x$ orthogonal matrix U and a diagonal matrix Ψ such that $\Phi = U\Psi U'$. Let

$$\Psi = \begin{bmatrix} \widehat{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\widehat{\Psi} = \text{diag}(\varphi_1, \dots, \varphi_m)$ is an $m \times m$ diagonal matrix and $\{\varphi_i\}_{i=1}^m$ are the positive eigenvalues of Φ. Clearly, $\text{rank}(\Phi) = m \le n_x$. The matrix Φ can be factored into $\Psi = \Delta'\widehat{\Psi}\Delta$, where $\Delta = \begin{bmatrix} I_m & \mathbf{0}_{m\times(n_x-m)} \end{bmatrix}$. Let $C = \Delta U'$ and $V = \widehat{\Psi}^{-1}$, completing the proof. *Q.E.D.*

PROOF OF LEMMA 2: The assumption of $AA' + W \succ 0$ and $W \succeq 0$ ensures that $(A\Sigma A' + W)$ is invertible for $\Sigma \succ 0$. Compute the second derivative of $\mathcal{F}$:

$$\mathcal{F}''(\Sigma) = \Sigma^{-1} \otimes \Sigma^{-1} - \beta\big(A'\big(A\Sigma A' + W\big)^{-1}A\big) \otimes \big(A'\big(A\Sigma A' + W\big)^{-1}A\big).$$

By the property of the Kronecker product $\otimes$, it is sufficient to show that

$$\Sigma^{-1} - \sqrt{\beta}A'\big(A\Sigma A' + W\big)^{-1}A \succeq 0 \quad \text{for } \beta \in (0, 1],$$

with strict matrix inequality for $\beta \in (0, 1)$, or

$$\Sigma^{-1/2}\big(I - \sqrt{\beta}\Sigma^{1/2}A'\big(A\Sigma A' + W\big)^{-1}A\Sigma^{1/2}\big)\Sigma^{-1/2} \succeq 0 \quad \text{for } \beta \in (0, 1],$$

with strict matrix inequality for $\beta \in (0, 1)$. The last matrix inequality is equivalent to

$$I \succeq \sqrt{\beta}\Sigma^{1/2}A'\big(A\Sigma A' + W\big)^{-1}A\Sigma^{1/2}.$$

Jianjun Miao: miaoj@bu.edu
Jieran Wu: jw5ya@zju.edu.cn
Eric R. Young: ey2d@virginia.edu; eric.young@clev.frb.org

 https://doi.org/10.3982/ECTA18086

Thus, by the eigendecomposition theorem, we only need to show that the largest eigenvalue of the positive semidefinite matrix $\Sigma^{1/2}A'(A\Sigma A'+W)^{-1}A\Sigma^{1/2}$ does not exceed 1.[1]

Let $\sigma(X)$ denote the column vector of all eigenvalues of any n-dimensional square matrix X with elements ordered according to $\sigma_1(X) \leq \sigma_2(X) \leq \cdots \leq \sigma_n(X)$. We follow the convention that $\sigma(X) \geq \sigma(Y)$ if $\sigma_i(X) \geq \sigma_i(Y)$ for all i. We then have

$$\begin{aligned}\sigma\left(\Sigma^{1/2}A'\left(A\Sigma A'+W\right)^{-1}A\Sigma^{1/2}\right) &= \sigma\left(\left(A\Sigma A'+W\right)^{-1}A\Sigma A'\right)\\ &\leq \sigma\left(\left(A\Sigma A'+\epsilon I+W\right)^{-1}\left(A\Sigma A'+\epsilon I\right)\right), \qquad \text{(B.1)}\end{aligned}$$

for $\epsilon > 0$, where the equality follows from the fact that $\sigma(XY) = \sigma(YX)$ for any two square matrices X and Y (Horn and Johnson (2013), p. 65, Theorem 1.3.22) and the inequality follows from Theorem 5 of Wang, Xi, and Zhang (1999, p. 47).

Let $N \equiv \epsilon I + A\Sigma A'$. Since $A\Sigma A' \succeq 0$ and $\epsilon > 0$, we have $N \succ 0$. Since $W \succeq 0$, we have the decomposition $W = MM'$ for some $M \succeq 0$. By the matrix inversion lemma,

$$(W+N)^{-1} = \left(MIM'+N\right)^{-1} = N^{-1} - L,$$

where we define

$$L \equiv N^{-1}M\left(I+M'N^{-1}M\right)M'N^{-1} \succeq 0.$$

Then we have

$$\begin{aligned}\sigma\left(\left(A\Sigma A'+\epsilon I+W\right)^{-1}\left(A\Sigma A'+\epsilon I\right)\right) &= \sigma\left((W+N)^{-1}N\right)\\ &= \sigma\left(\left(N^{-1}-L\right)N\right) = \sigma(I-LN)\\ &= \sigma\left(N^{-\frac{1}{2}}N^{\frac{1}{2}}-LN\right)\\ &= \sigma\left(N^{-\frac{1}{2}}\left(I-N^{\frac{1}{2}}LN^{\frac{1}{2}}\right)N^{\frac{1}{2}}\right) = \sigma\left(I-N^{\frac{1}{2}}LN^{\frac{1}{2}}\right),\end{aligned}$$

where the last equality follows from Theorem 1.3.22 of Horn and Johnson (2013, p. 65). By Weyl's inequalities for eigenvalues of the sum of two symmetric matrices (Horn and Johnson (2013), p. 239, Theorem 4.3.1), the largest eigenvalue of $I - N^{\frac{1}{2}}LN^{\frac{1}{2}}$ does not exceed the sum of the largest eigenvalue of I and the largest eigenvalue of $-N^{\frac{1}{2}}LN^{\frac{1}{2}}$. Since $-N^{\frac{1}{2}}LN^{\frac{1}{2}} \preceq 0$, we have $\sigma_n(I - N^{\frac{1}{2}}LN^{\frac{1}{2}}) \leq 1$, where n denotes the dimension of Σ. Thus, $\sigma(I - N^{\frac{1}{2}}LN^{\frac{1}{2}}) \leq \mathbf{1}_n$, where $\mathbf{1}_n$ denotes the n-dimensional column vector of ones. It follows from (B.1) that

$$\sigma\left(\Sigma^{1/2}A'\left(A\Sigma A'+W\right)^{-1}A\Sigma^{1/2}\right) \leq \mathbf{1}_n$$

as desired. *Q.E.D.*

PROOF OF PROPOSITION 2: We prove that $J_t(\Sigma_{t-1})$ is strictly convex in Σ_{t-1} for $t = 0, 1, \ldots, T$ by backward induction. In the last period, it follows from (19) that $J_T(\Sigma_{T-1})$ is strictly convex in Σ_{T-1}. Suppose that $J_{t+1}(\Sigma_t)$ is strictly convex in Σ_t for any $t \leq T-1$.

[1]More precisely, let $\Sigma^{1/2}A'(A\Sigma A'+W)^{-1}A\Sigma^{1/2} = U\Psi U'$ be the (unitary) eigendecomposition, $UU' = I$. $\Psi \succeq 0$ is the diagonal matrix of eigenvalues. Then $I - \sqrt{\beta}\Sigma^{1/2}A'(A\Sigma A'+W)^{-1}A\Sigma^{1/2} = UU' - \sqrt{\beta}U\Psi U' = U(I-\sqrt{\beta}\Psi)U'$.

Then, by Lemma 2, the objective function in (20) is strictly convex. Since the constraint set is convex, we can verify that $J_t(\Sigma_{t-1})$ is strictly convex.

Now we transform the dynamic programming problem (20) into an SDP representation. The matrix determinant lemma (Theorem 18.1.1 in Harville (1997)) implies that the preceding expression is equal to

$$\log\det\bigl(A_t\Sigma_t A_t' + W_t\bigr) - \log\det(\Sigma_t) = \log\det W_t - \log\det\bigl(\Sigma_t^{-1} + A_t' W_t^{-1} A_t\bigr)^{-1}. \tag{B.2}$$

Due to the monotonicity of the determinant function, we have

$$-\log\det\bigl(\Sigma_t^{-1} + A_t' W_t^{-1} A_t\bigr)^{-1} = \min_{\Pi_t \succ 0} -\log\det \Pi_t$$

subject to

$$\Pi_t \preceq \bigl(\Sigma_t^{-1} + A_t' W_t^{-1} A_t\bigr)^{-1}. \tag{B.3}$$

Apply the matrix inversion formula to rewrite (B.3) as

$$\Pi_t \preceq \Sigma_t - \Sigma_t A_t' \bigl(W_t + A_t \Sigma_t A_t'\bigr)^{-1} A_t \Sigma_t,$$

which is equivalent to

$$\begin{bmatrix} \Sigma_t - \Pi_t & \Sigma_t A_t' \\ A_t \Sigma_t & W_t + A_t \Sigma_t A_t' \end{bmatrix} \succeq 0, \tag{B.4}$$

by the Schur complement property. By (B.2) and the preceding derivations, we have

$$\log\det\bigl(A_t\Sigma_t A_t' + W_t\bigr) = \min_{\Pi_t \succ 0} -\log\det\Pi_t + \log\det W_t + \log\det(\Sigma_t)$$

subject to (B.4). Replacing $\log\det(A_t\Sigma_t A_t' + W_t)$ in (20) with the preceding minimized value, we obtain the representation in the proposition. *Q.E.D.*

PROOF OF PROPOSITION 3: We first consider the following static RI problem:

$$\min_{\Sigma} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\log\det(\Sigma_{-1}) - \frac{\lambda}{2}\log\det\Sigma \tag{B.5}$$

subject to $0 \prec \Sigma \preceq \Sigma_{-1}$, where Σ_{-1} is an exogenous prior covariance matrix. We can ignore the exogenous term $0.5\lambda\log\det(\Sigma_{-1})$ in the objective function. This is a convex problem. By SDP theory, define the Lagrangian as

$$\mathcal{L} = \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\log\det(\Sigma_{-1}) - \frac{\lambda}{2}\log\det\Sigma + \Lambda \bullet (\Sigma - \Sigma_{-1}),$$

where $\Lambda \succeq 0$ is the Lagrange multiplier. The Kuhn–Tucker conditions are necessary and sufficient for optimality:

$$\frac{\lambda}{2}\Sigma^{-1} = \Omega + \Lambda, \quad \Lambda \bullet (\Sigma - \Sigma_{-1}) = 0. \tag{B.6}$$

The following lemma presents the generalized reverse water-filling solution derived in the 2018 version of our paper. The condition here is weaker by allowing Ω to be symmetric.

LEMMA 3: *Suppose that Ω is symmetric and $\Sigma_{-1} \succ 0$. Perform the eigendecomposition*

$$\Sigma_{-1}^{\frac{1}{2}} \Omega \Sigma_{-1}^{\frac{1}{2}} = UDU',$$

where U is an orthogonal matrix and D is a diagonal matrix of eigenvalues $\{d_i\}$. Then the optimal solution to the static RI problem (B.5) is given by

$$\Sigma = \Sigma_{-1}^{\frac{1}{2}} U \left[\max\left(\frac{2}{\lambda} D, I \right) \right]^{-1} U' \Sigma_{-1}^{\frac{1}{2}}, \tag{B.7}$$

and the optimal information structure satisfies

$$C'V^{-1}C = \Sigma_{-1}^{-\frac{1}{2}} U \max\left(0, \frac{2}{\lambda} D - I \right) U' \Sigma_{-1}^{-\frac{1}{2}}.$$

The signal dimension is equal to the number of eigenvalues greater than $\lambda/2$ and decreases as λ increases.

PROOF: If $\Omega \succeq 0$, this result is the special case of Proposition 4 when $\rho = 0$ and Σ_{-1} is viewed as W. It follows from (44) that

$$\widehat{\Sigma}_i = \min(1, \widehat{\Sigma}_i^*), \qquad \widehat{\Sigma}_i^* = \frac{\lambda}{2d_i} \quad \text{for } d_i \geq 0. \tag{B.8}$$

Since the diagonal matrix $\text{diag}(\min(1, 0.5\lambda/d_i))_{i=1}^{n_x}$ can be equivalently written as $[\max(2D/\lambda, I)]^{-1}$ using the Matlab max operator, we obtain the desired result. If Ω is symmetric, we find that problem (B.19) still applies for $\rho = 0$ by inspecting the proof of Proposition 4. Thus, (B.8) holds for $d_i \geq 0$. For any eigenvalue $d_i < 0$, the objective in (B.19) decreases with $\widehat{\Sigma}_i$ so that constraint (B.20) binds for i when $\rho = 0$. Thus, the solution is $\widehat{\Sigma}_i = 1$. We can still write $\widehat{\Sigma}_i = [\max(2d_i/\lambda, 1)]^{-1}$ for any $d_i < 0$. Thus, we obtain (B.7). *Q.E.D.*

It follows from (B.6) and (B.7) that the Lagrange multiplier is given by

$$\Lambda = \Sigma_{-1}^{-\frac{1}{2}} U \max\left(\frac{\lambda}{2} I - D, 0 \right) U' \Sigma_{-1}^{-\frac{1}{2}}. \tag{B.9}$$

Next, we turn to the dynamic RI model. By backward induction, we claim that the first-order conditions (25) and (26) can be derived from solving the following sequence of static RI problems by taking the sequence of priors $\{\Sigma_{t|t-1}\}_{t=0}^{T}$ as given:

$$\min_{\Sigma_t} \text{tr}(\Theta_t \Sigma_t) - \frac{\lambda}{2} \log\det \Sigma_t \tag{B.10}$$

subject to

$$0 \prec \Sigma_t \preceq \Sigma_{t|t-1}, \tag{B.11}$$

where we define Θ_t in (29) for $t = 0, 1, \ldots, T$. In the last period T, $\Theta_T = \Omega_T$ and we immediately obtain (25) and (26) at T. Consider problem (B.10) at any $t < T$. In (29), we take $\Sigma_{t+1|t}$ as given and Λ_{t+1} as the Lagrange multiplier for (B.11) in the static problem

(B.10) at $t+1$. Then we take Θ_t as exogenous. Let Λ_t be the Lagrange multiplier for (B.11) in period t. The first-order conditions for this static problem give (25) and (26) at t. Notice that $\{\Sigma_{t|t-1}\}_{t=0}^{T}$ must also satisfy (14). In Appendix G, we develop an algorithm to compute solutions to dynamic RI problems based on the above sequence of static problems.

Viewing Θ_t as Ω and $\Sigma_{t|t-1}$ as Σ_{-1}, we apply Lemma 3 to the static problem (B.10) at any time t to obtain (30) and (31).[2] By (B.9), we have

$$\Lambda_t = \Sigma_{t|t-1}^{-\frac{1}{2}} U_t \max\left(\frac{\lambda}{2} I - D_t, 0\right) U_t' \Sigma_{t|t-1}^{-\frac{1}{2}}. \tag{B.12}$$

Substituting this expression into (29), we obtain (32):

$$\begin{aligned}
\Theta_t &= \Omega_t + \beta A_t' \left(\frac{\lambda}{2} \Sigma_{t+1|t}^{-1} - \Lambda_{t+1}\right) A_t \\
&= \Omega_t + \beta A_t' \Sigma_{t+1|t}^{-\frac{1}{2}} U_{t+1} \left(\frac{\lambda}{2} I - \max\left(\frac{\lambda}{2} I - D_{t+1}, 0\right)\right) U_{t+1}' \Sigma_{t+1|t}^{-\frac{1}{2}} A_t \\
&= \Omega_t + \beta A_t' \Sigma_{t+1|t}^{-\frac{1}{2}} U_{t+1} \min\left(D_{t+1}, \frac{\lambda}{2} I\right) U_{t+1}' \Sigma_{t+1|t}^{-\frac{1}{2}} A_t
\end{aligned}$$

for $t = 0, 1, \ldots, T-1$. *Q.E.D.*

PROOF OF PROPOSITION 4: The matrix determinant lemma implies that

$$\log\det(A\Sigma A' + W) - \log\det\Sigma = \log\det W - \log\det(\Sigma^{-1} + A'W^{-1}A)^{-1}.$$

Thus, problem (41) becomes

$$\min_{\Pi, \Sigma \succ 0} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}[\log\det W - \log\det\Pi] \tag{B.13}$$

subject to

$$\Pi = (\Sigma^{-1} + A'W^{-1}A)^{-1}, \tag{B.14}$$

$$A\Sigma A' + W \succeq \Sigma. \tag{B.15}$$

Recall that the symmetric matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$ admits an eigendecomposition $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}} = U\Omega_d U'$. Define matrices

$$\widehat{\Pi} = U'W^{-\frac{1}{2}}\Pi W^{-\frac{1}{2}}U, \qquad \widehat{\Sigma} = U'W^{-\frac{1}{2}}\Sigma W^{-\frac{1}{2}}U.$$

Then we can derive that

$$\Pi = W^{\frac{1}{2}}U\widehat{\Pi}U'W^{\frac{1}{2}}, \qquad \Sigma = W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}}, \qquad \operatorname{tr}(\Omega\Sigma) = \operatorname{tr}(\Omega_d\widehat{\Sigma}),$$

$$\log\det W - \log\det\Pi = -\log\det\widehat{\Pi}.$$

[2] Notice that it is not clear whether $\Theta_t \succeq 0$. But we do know Θ_t is symmetric so that we can apply Lemma 3.

Given $A = \rho I$, we can also show that equations (B.14) and (B.15) are equivalent to

$$\widehat{\Pi}^{-1} = \widehat{\Sigma}^{-1} + \rho^2 I, \tag{B.16}$$

$$\left(1 - \rho^2\right)\widehat{\Sigma} \preceq I. \tag{B.17}$$

Now the problem in (B.13) is equivalent to

$$\min_{\widehat{\Pi}, \widehat{\Sigma}} \operatorname{tr}(\Omega_d \widehat{\Sigma}) - \frac{\lambda}{2} \log \det \widehat{\Pi}$$

subject to (B.16) and (B.17). By the Hadamard inequality for positive definite matrices (Cover and Thomas (2006), Theorem 17.9.2),

$$\det \widehat{\Pi} \leq \prod_{i=1}^{n_x} \widehat{\Pi}_i,$$

where $\widehat{\Pi}_i$ is the diagonal element of $\widehat{\Pi}$. The equality holds if and only if $\widehat{\Pi}$ is diagonal. Thus, if diagonal elements of $\widehat{\Pi}$ are fixed, $\det \widehat{\Pi}$ is maximized by setting all off-diagonal entries to zero. As a result, the optimal solution for $\widehat{\Pi}$ must be diagonal. Let $\widehat{\Pi} = \operatorname{diag}(\widehat{\Pi}_i)_{i=1}^{n_x}$. By (B.16), $\widehat{\Sigma}$ is also diagonal and its diagonal elements are given by

$$\widehat{\Sigma}_i = \left(\widehat{\Pi}_i^{-1} - \rho^2\right)^{-1}, \quad i = 1, 2, \dots, n_x. \tag{B.18}$$

Thus, the problem is equivalent to

$$\min_{\widehat{\Pi}_i} \operatorname{tr}(\Omega_d \widehat{\Sigma}) - \frac{\lambda}{2} \sum_{i=1}^{n_x} \log \widehat{\Pi}_i$$

subject to (B.18) and

$$\left(1 - \rho^2\right)\widehat{\Sigma}_i \leq 1, \quad i = 1, \dots, n_x.$$

Equivalently, rewriting this problem in terms of $\widehat{\Sigma}_i$ using (B.18) yields

$$\min_{\widehat{\Sigma}_i > 0} \sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i + \frac{\lambda}{2} \sum_{i=1}^{n_x} \log\left(\rho^2 + \frac{1}{\widehat{\Sigma}_i}\right) \tag{B.19}$$

subject to

$$\left(1 - \rho^2\right)\widehat{\Sigma}_i \leq 1, \quad i = 1, \dots, n_x. \tag{B.20}$$

Since $\Omega \succeq 0$, $d_i \geq 0$ for all i. Consider two cases. First, let $|\rho| < 1$. If $d_i = 0$, then $\widehat{\Sigma}_i = 1/(1 - \rho^2)$. If $d_i > 0$, then we use the Kuhn–Tucker condition to show that

$$\widehat{\Sigma}_i = \min\left(\frac{1}{1 - \rho^2}, \widehat{\Sigma}_i^*\right), \tag{B.21}$$

where

$$\widehat{\Sigma}_i^* = \frac{1}{2\rho^2}\left(\sqrt{1 + \frac{2\rho^2 \lambda}{d_i}} - 1\right). \tag{B.22}$$

Using $\lim_{d_i\to 0}\widehat{\Sigma}_i^* = \infty$ and $\lim_{\rho\to 0}\widehat{\Sigma}_i^* = \lambda/(2d_i)$, we obtain the solution in the proposition.

Second, let $|\rho| \geq 1$ and $\Omega \succ 0$. Then all eigenvalues $d_i > 0$ and constraint (B.20) does not bind. The optimal solution to (B.19) is $\widehat{\Sigma}_i = \widehat{\Sigma}_i^*$. *Q.E.D.*

PROOF OF PROPOSITION 5: The optimal signal-to-noise ratio is given by

$$\begin{aligned}\Phi &= \Sigma^{-1} - \left(\rho^2\Sigma + W\right)^{-1} \\ &= W^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'W^{-\frac{1}{2}} - \left[\rho^2 W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}} + W\right]^{-1} \\ &= W^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'W^{-\frac{1}{2}} - W^{-\frac{1}{2}}U\left[\rho^2\widehat{\Sigma} + I\right]^{-1}U'W^{-\frac{1}{2}} \\ &= W^{-\frac{1}{2}}U\left(\widehat{\Sigma}^{-1} - \left[\rho^2\widehat{\Sigma} + I\right]^{-1}\right)U'W^{-\frac{1}{2}} \\ &= W^{-\frac{1}{2}}U\operatorname{diag}\left\{\max\left(0, \frac{2d_i}{\lambda}\left[1 - (1-\rho^2)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\}U'W^{-\frac{1}{2}},\end{aligned}$$

where the last equality follows from (B.21) and (B.22). The dimension of the signal is determined by the rank of the inside diagonal matrix, which is determined by the number of d_i such that

$$\frac{2d_i}{\lambda}\left[1 - (1-\rho^2)\widehat{\Sigma}_i^*\right] > 0.$$

Using equation (B.22) and Proposition 4, we obtain the desired result. *Q.E.D.*

PROOF OF PROPOSITION 6: Since $\operatorname{rank}(\Omega) = 1$, we have $\operatorname{rank}(W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}) = 1$. We claim that matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$ has a unique positive eigenvalue $d_1 \equiv \|W^{1/2}G'\|^2$ and an associated unit eigenvector $W^{\frac{1}{2}}G'/\|W^{1/2}G'\|$, where $\|\cdot\|$ denotes the Euclidean norm. To prove this claim, we verify that

$$\begin{aligned}W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}\frac{W^{\frac{1}{2}}G'}{\|W^{1/2}G'\|} &= \left(W^{\frac{1}{2}}G'\right)\left(W^{\frac{1}{2}}G'\right)'\frac{W^{\frac{1}{2}}G'}{\|W^{1/2}a\|} = \left(W^{\frac{1}{2}}G'\right)GW^{\frac{1}{2}}\frac{W^{\frac{1}{2}}G'}{\|W^{1/2}G'\|} \\ &= \left(W^{\frac{1}{2}}G'\right)\frac{\|W^{1/2}G'\|^2}{\|W^{1/2}G'\|} = \|W^{1/2}G'\|^2\frac{W^{\frac{1}{2}}G'}{\|W^{1/2}G'\|}.\end{aligned}$$

Thus, Ω_d has only one positive element $d_1 = \|W^{1/2}G'\|^2$ and other diagonal elements $d_i = 0$ for $i = 2, \dots, n_x$. Moreover, the optimal signal dimension is at most 1.

By Propositions 4 and 5, we have

$$\widehat{\Sigma}_1 = \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_1^*\right), \qquad \widehat{\Sigma}_i = \frac{1}{1-\rho^2}, \quad i = 2, \dots, n_x,$$

where

$$\widehat{\Sigma}_1^* = \frac{1}{2\rho^2}\left(\sqrt{1 + \frac{2\rho^2\lambda}{d_1}} - 1\right).$$

The optimal information structure $\{C, V\}$ satisfies

$$C'V^{-1}C = W^{-\frac{1}{2}} U \operatorname{diag}\left\{\max\left(0, \frac{2d_i}{\lambda}\left[1 - (1-\rho^2)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\} U' W^{-\frac{1}{2}}.$$

If $\lambda \geq 2d_1/(1-\rho^2)^2$, we can check that $\widehat{\Sigma}_i = 1/(1-\rho^2)$ for all i so that $\Sigma = W/(1-\rho^2)$ and no information is collected. There is only one positive element in the above inside diagonal matrix if $0 < \lambda < 2d_1/(1-\rho^2)^2$, which is

$$\frac{2d_1}{\lambda}\left[1 - (1-\rho^2)\widehat{\Sigma}_1^*\right] = \frac{d_1}{\lambda\rho^2}\left[1 + \rho^2 - (1-\rho^2)\sqrt{1 + \frac{2\rho^2\lambda}{d_1}}\right] > 0,$$

The optimal information structure corresponds to the positive eigenvalue's eigenvector and is given by

$$C' = W^{-\frac{1}{2}} \frac{W^{\frac{1}{2}} G'}{\|W^{1/2}G'\|} \quad \Longrightarrow \quad C = \frac{G}{\|W^{1/2}G'\|},$$

$$V^{-1} = \frac{d_1}{\lambda\rho^2}\left[1 + \rho^2 - (1-\rho^2)\sqrt{1 + \frac{2\rho^2\lambda}{d_1}}\right] > 0.$$

The optimal conditional covariance in the proposition follows from Proposition 4. In particular,

$$\Sigma = W^{\frac{1}{2}} U \begin{bmatrix} \widehat{\Sigma}_1^* & 0 \\ 0 & \dfrac{1}{1-\rho^2} I \end{bmatrix} U' W^{\frac{1}{2}}.$$

Partition $U = [U_1, U_2]$ conformably, where $U_1 = W^{\frac{1}{2}} G'/\|W^{1/2}G'\|$. Then we have $U_1U_1' + U_2U_2' = I$. Thus,

$$\Sigma = W^{\frac{1}{2}}\left[\frac{I}{1-\rho^2} - U_1U_1'\left(\frac{1}{1-\rho^2} - \Sigma_1^*\right)\right] W^{\frac{1}{2}}.$$

Simplifying yields the expression in the proposition. We can normalize C as $C = G$ so that the normalized optimal signal is given by

$$s_t = Gx_t + \|W^{1/2}G'\| v_t.$$

We then obtain (45). *Q.E.D.*

PROOF OF PROPOSITION 7: For the univariate case, we can write the RI problem as follows:

$$\min_{\{\Sigma_t\}} \sum_{t=0}^{\infty} \beta^t \left[\Sigma_t + \frac{\lambda}{2}\log\left(\frac{\Sigma_{t|t-1}}{\Sigma_t}\right)\right]$$

subject to $0 < \Sigma_t \le \Sigma_{t|t-1}$ for $t \ge 0$, $\Sigma_{0|-1}$ given, and $\Sigma_{t|t-1} = \rho^2\Sigma_{t-1} + W$ for $t \ge 1$. The first-order conditions are given by

$$\begin{aligned} &\frac{\lambda}{2}\Sigma_t^{-1} = 1 + \Lambda_t + \frac{\lambda}{2}\beta\rho^2\left(\rho^2\Sigma_t + W\right)^{-1} - \beta\rho^2\Lambda_{t+1}, \\ &\Lambda_t(\Sigma_{t|t-1} - \Sigma_t) = 0, \quad \Lambda_t \ge 0, \text{ for } t \ge 0. \end{aligned} \tag{B.23}$$

Case 1. $|\rho| < 1$. First, consider the steady state in which all variables are constant over time and the time subscripts are removed. If the no-forgetting constraint does not bind, then $\Lambda = 0$. Equation (B.23) becomes

$$\frac{\lambda}{2}\Sigma^{-1} = 1 + \frac{\lambda}{2}\beta\rho^2\left(\rho^2\Sigma + W\right)^{-1}. \tag{B.24}$$

Simplifying yields the quadratic equation in the proposition. Let the unique positive root be Σ^*. In the steady state, the no-forgetting constraint must hold so that $\Sigma \le \rho^2\Sigma + W$. This means $\Sigma \le W/(1-\rho^2)$. Thus, the steady-state solution is given by $\widehat{\Sigma}$ in the proposition.

Next, consider the transition dynamics. If $\Sigma_{0|-1} \ge \widehat{\Sigma}$, then we can verify that $\Sigma_t = \widehat{\Sigma}$ for all $t \ge 0$ is the solution. That is, Σ_t immediately jumps to the steady state. Since the problem is strictly convex, this is the unique solution. If $\Sigma_{0|-1} < \widehat{\Sigma}$, let t_0 be the first time when the no-forgetting constraint does not bind. Then we can verify that $\Sigma_t = \widehat{\Sigma}$ for $t \ge t_0$ satisfies the first-order conditions and no-forgetting constraints. Before time t_0, all no-forgetting constraints bind, $\Sigma_t = \Sigma_{t|t-1}$, $t \le t_0$. Thus, we have $\Sigma_t = \min(\Sigma_{t|t-1}, \widehat{\Sigma})$. By the uniqueness, this is the only solution.

Case 2. $|\rho| \ge 1$. Then Σ^* satisfies the no-forgetting constraint as $\Sigma^* < \rho^2\Sigma^* + W$. Thus, Σ^* is the steady-state solution. The rest of the proof is the same as in the previous case.

Finally, for the univariate case, we can write the optimal signal in the form $s_t = x_t + v_t$, where v_t is a Gaussian white noise with variance satisfying

$$V_t^{-1} = \Sigma_t^{-1} - \Sigma_{t|t-1}^{-1}.$$

All no-forgetting constraints bind before time t_0. During these periods, no signal is acquired. Starting from time t_0 on, the no-forgetting constraints never bind. We have $V_t^{-1} = \widehat{\Sigma}^{-1} - \Sigma_{t|t-1}^{-1}$. *Q.E.D.*

APPENDIX C: RI PROBLEMS WITH PERIOD-BY-PERIOD CAPACITY CONSTRAINTS

In this appendix, we study Problem 1 with period-by-period capacity constraints. As in the analysis of Section 2, we can show that the optimal information structure is determined by the following problem:

PROBLEM 6—Optimal information structure for Problem 1:

$$\min_{\{\Sigma_t\}_{t=0}^T} \sum_{t=0}^{T} \beta^t \operatorname{tr}(\Omega_t\Sigma_t)$$

subject to

$$\log\det\left(A_{t-1}\Sigma_{t-1}A'_{t-1} + W_{t-1}\right) - \log\det(\Sigma_t) \le 2\kappa, \tag{C.1}$$

$$\log\det(\Sigma_{-1}) - \log\det(\Sigma_0) \leq 2\kappa, \tag{C.2}$$

$$\Sigma_t \preceq A_{t-1}\Sigma_{t-1}A'_{t-1} + W_{t-1}, \tag{C.3}$$

$$\Sigma_0 \preceq \Sigma_{-1}, \tag{C.4}$$

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

Since the log-determinant function is concave, the constraint set may not be convex in $\{\Sigma_t\}_{t=0}^T$. Thus, the Kuhn–Tucker conditions may not be optimal. By dynamic programming, the value function satisfies the Bellman equation

$$J_t(\Sigma_{t-1}) = \min_{\Sigma_t \succ 0} \operatorname{tr}(\Omega_t\Sigma_t) + \beta J_{t+1}(\Sigma_t)$$

subject to (C.1) and (C.3) for $t \geq 1$. In the last period T, $J_{T+1}(\Sigma_T) \equiv 0$. In the initial period, we have

$$J_0(\Sigma_{-1}) = \min_{\Sigma_0 \succ 0} \operatorname{tr}(\Omega_0\Sigma_0) + \beta J_1(\Sigma_0)$$

subject to (C.2) and (C.4). Since $\log\det(A_{t-1}\Sigma_{t-1}A'_{t-1} + W_{t-1})$ is concave in Σ_{t-1}, the value function $J_t(\Sigma_{t-1})$ may not be convex for $t = 0, 1, \dots, T$. This can be easily seen for $J_T(\Sigma_{T-1})$ in the last period using the envelope theorem. For a univariate problem with $n_x = 1$, Σ_t is a scalar and we can rewrite (C.1) and (C.2) as linear scalar constraints so that $J_t(\Sigma_{t-1})$ is convex.

Nonconvexity poses substantial difficulty when solving the above dynamic programming problem. This issue does not arise when solving for the long-run golden-rule information structure.

PROBLEM 7—Golden-rule information structure for Problem 6:

$$\min_{\Sigma \succ 0} (1-\beta)\operatorname{tr}\big(A'PA\Sigma\big) + \operatorname{tr}(\Omega\Sigma) \tag{C.5}$$

subject to (42) *and*

$$\log\det\big(A\Sigma A' + W\big) - \log\det(\Sigma) \leq 2\kappa.$$

By Lemma 2, $\log\det(A\Sigma A' + W) - \log\det(\Sigma)$ is a convex function of Σ if $AA' + W \succ 0$. Thus, the above problem is a convex program under this assumption. This problem is the same as that in Sims (2003) except that there is a new term in (C.5) as discussed in Section 6. Notice that software CVX does not recognize that $\log\det(A\Sigma A' + W) - \log\det(\Sigma)$ is convex in Σ by its ruleset.

To apply CVX, we need to transform Problem 7 into a DCP. There are several ways to do it as discussed in Appendix F. For example, if $W \succ 0$, we can show that $\log\det(A\Sigma A' + W) - \log\det(\Sigma) = c(\Sigma)$, where $c(\Sigma)$ is a new function defined as

$$c(\Sigma) \equiv \min_{\Pi \succ 0} \; -\log\det\Pi + \log\det W$$

subject to

$$\begin{bmatrix} \Sigma - \Pi & \Sigma A' \\ A\Sigma & W + A\Sigma A' \end{bmatrix} \succeq 0. \tag{C.6}$$

Since the objective function is convex and the constraint is a linear matrix inequality, $c(\Sigma)$ is convex in Σ and can be added to the CVX atom library. We then transform Problem 7 into the following DCP:

$$\min_{\Sigma\succ 0}(1-\beta)\,\mathrm{tr}\big(A'PA\Sigma\big)+\mathrm{tr}(\Omega\Sigma) \tag{C.7}$$

subject to (42) and $c(\Sigma)\leq 2\kappa$. For tracking problems, the term $(1-\beta)\,\mathrm{tr}(A'PA\Sigma)$ does not appear in (C.7). We have used this method to numerically solve the pricing example in Section 4.

In an earlier version of our paper, we solve the following inverse problem as in the rate-distortion theory in the engineering literature:

$$R(D)\equiv\min_{\Sigma\succ 0}\frac{1}{2}\log\det\big(A\Sigma A'+W\big)-\frac{1}{2}\log\det(\Sigma) \tag{C.8}$$

subject to (42) and

$$(1-\beta)\,\mathrm{tr}\big(A'PA\Sigma\big)+\mathrm{tr}(\Omega\Sigma)\leq D.$$

The function $R(D)$ is decreasing and convex in D. Given any capacity $\kappa>0$, we can find D using this function and then solve the corresponding Σ. The earlier version of our paper also derives results similar to Propositions 4 and 5. We omit the details here.

APPENDIX D: INVERTIBILITY ASSUMPTION

In this appendix, we discuss how we can relax the assumption of the invertibility of W_t in Proposition 2. We then study an example for MA processes.

First, we consider the case in which the state transition matrix is invertible and present a different SDP representation.

PROPOSITION 8: *Suppose that $W_t\succeq 0$ is singular for some t and* $\mathrm{rank}(A_t)=n_x$ *for* $t=0,1,\ldots,T-1$. *Then the value function $J_t(\Sigma_{t-1})$ is strictly convex in Σ_{t-1} for $t=0,1,\ldots,T-1$ and satisfies the dynamic semidefinite program*:

$$\begin{aligned}J_t(\Sigma_{t-1})={}&\min_{\Psi_t\succ 0,\Sigma_t\succ 0}\mathrm{tr}(\Omega_t\Sigma_t)-\frac{\lambda}{2}(1-\beta)\log\det(\Sigma_t)\\&+\frac{\lambda\beta}{2}\big(2\log|\det A_t|-\log\det\Psi_t\big)+\beta J_{t+1}(\Sigma_t)\end{aligned} \tag{D.1}$$

subject to (17) *for* $t\geq 1$ *and* (18) *for* $t=0$, *and*

$$\begin{bmatrix}I-\Psi_t & M_t'\\ M_t & A_t\Sigma_tA_t'+W_t\end{bmatrix}\succeq 0, \tag{D.2}$$

where $W_t=M_tM_t'$ *with* $M_t\succeq 0$, $J_T(\Sigma_{T-1})$ *is strictly convex and satisfies* (19).

PROOF: We can apply the same proof for Proposition 2 to show that $J_t(\Sigma_{t-1})$ is convex using the Bellman equation (20). Now we derive a different SDP representation. Since $W_t\succeq 0$, we have the decomposition $W_t=M_tM_t'$ with $M_t\succeq 0$. Since A_t is invertible, $A_t\Sigma_tA_t'$ is also invertible. Applying the matrix determinant lemma yields

$$\det\big(A_t\Sigma_tA_t'+W_t\big)=\det\big(I+M_t'\big(A_t\Sigma_tA_t'\big)^{-1}M_t\big)\det\big(A_t\Sigma_tA_t'\big).$$

Thus, we have

$$\begin{aligned}&\log\det\bigl(A_t\Sigma_tA_t'+W_t\bigr)-\log\det(\Sigma_t)\\&\quad=-\log\det\bigl(I+M_t'\bigl(A_t\Sigma_tA_t'\bigr)^{-1}M_t\bigr)^{-1}+\log\det\bigl(A_t\Sigma_tA_t'\bigr)-\log\det(\Sigma_t)\\&\quad=-\log\det\bigl(I+M_t'\bigl(A_t\Sigma_tA_t'\bigr)^{-1}M_t\bigr)^{-1}+2\log|\det A_t|.\end{aligned}$$

Due to the monotonicity of the determinant function, the last expression is equal to the optimal value of

$$\min_{\Psi_t}2\log|\det A_t|-\log\det\Psi_t$$

subject to

$$0\prec\Psi_t\preceq\bigl(I+M_t'\bigl(A_t\Sigma_tA_t'\bigr)^{-1}M_t\bigr)^{-1}. \tag{D.3}$$

Now use the matrix inversion lemma to get

$$\bigl(I+M_t'\bigl(A_t\Sigma_tA_t'\bigr)^{-1}M_t\bigr)^{-1}=I-M_t'\bigl(A_t\Sigma_tA_t'+M_tM_t'\bigr)M_t.$$

By the Schur complement property, (D.3) is equivalent to

$$\begin{bmatrix}I-\Psi_t & M_t'\\ M_t & A_t\Sigma_tA_t'+W_t\end{bmatrix}\succeq 0. \tag{D.4}$$

In sum, we have shown that

$$\log\det\bigl(A_t\Sigma_tA_t'+W_t\bigr)=\min_{\Psi_t\succ 0}2\log|\det A_t|-\log\det\Psi_t+\log\det(\Sigma_t)$$

subject to (D.4). Substituting this equation into (20) yields the desired result. *Q.E.D.*

To illustrate the application of this proposition, we consider the LQG control problem with VAR(p) state dynamics

$$x_t=A_1x_{t-1}+A_2x_{t-2}+\cdots+A_px_{t-p}+B_0u_t+\epsilon_t,$$

where $A_1,\ldots,$ and A_p are $n\times n$ matrices and ϵ_t is Gaussian white noise with covariance matrix $W_0\succ 0$. We transform the state dynamics into VAR(1) form:

$$\bar{x}_t=A\bar{x}_{t-1}+Bu_t+\bar{\epsilon}_t,$$

where $\bar{x}_t=[x_t',x_{t-1}',\ldots,x_{t-p+1}']'$, $\bar{\epsilon}_t$ is a Gaussian white noise with covariance matrix W, and

$$A=\begin{bmatrix}A_1 & A_2 & \ldots & A_{p-1} & A_p\\ I_n & 0 & \cdots & 0 & 0\\ 0 & I_n & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & 0 & 0\\ 0 & 0 & 0 & I_n & 0\end{bmatrix},\qquad B=\begin{bmatrix}B_0\\ 0\\ \vdots\\ 0\\ 0\end{bmatrix},$$

$$W = \begin{bmatrix} I_n \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} W_0 \begin{bmatrix} I_n & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Then the problem fits in our general LQG RI framework. Notice that the covariance matrix of $\bar{\epsilon}_t$ satisfies $W \succeq 0$ and it is singular. So the SDP representation in Proposition 2 does not apply. As long as A_p is invertible so that A is invertible, we can apply Proposition 8 to derive an SDP representation. Notice that this proposition can also be applied to solve models with ARMA(p, q) processes $(p > q)$ as shown in Section 5.1.2 once we derive a state space representation.

Next, we consider a weaker assumption introduced by Afrouzi and Yang (2019): $A_t A_t' + W_t$ is invertible, but neither W_t nor A_t is invertible. Then Lemma 2 and Proposition 2 show that the dynamic RI problem is still convex. But the SDP representations in Propositions 2 and 8 do not apply. The first-order conditions based methods can easily incorporate the weaker assumption.

For the value function based methods, we have two ways to handle this case. The first way is to apply the convex-concave procedure (CCP) in the mathematics literature (Lipp and Boyd (2016)). The idea is to transform the difference of two concave functions as a DCP form using a linear approximation of $\log\det(A_t \Sigma_t A_t' + W_t)$. Lipp and Boyd (2016) established the global convergence of this procedure. The second way is to notice that the dynamic RI problem can be viewed as a sequence of static RI problems (B.10) as shown in the proof of Proposition 3. Each static problem is a DCP. In Appendix G, we describe algorithms to implement both procedures.

We close this appendix by solving a univariate tracking problem with MA process.[3] Let the tracking variable y_t follow an MA(2) process

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2},$$

where ϵ_t is a Gaussian white noise with variance σ^2. Then it admits the state space representation $x_{t+1} = A x_t + \eta_{t+1}$ and $y_t = G x_t$, where $x_t = (y_t, \epsilon_t, \epsilon_{t-1})'$, $\eta_t = (\epsilon_t, \epsilon_t, 0)'$, $G = [1, 0, 0]$,

$$A = \begin{bmatrix} 0 & \theta_1 & \theta_2 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad W = \begin{bmatrix} \sigma^2 & \sigma^2 & 0 \\ \sigma^2 & \sigma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \Omega = G'G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We can check that A and W are not invertible, but $AA' + W$ is invertible. Notice that this assumption is also satisfied by general ARMA(p, q) processes $(p, q \geq 0)$ using Hamilton's (1994) representation like (55).

Now we solve a numerical example with parameter values: $\theta_1 = 0.8$, $\theta_2 = 0.5$, $\sigma^2 = 0.25$, $\lambda = 0.5$, and $\beta = 0.9$. We apply the above two methods to compute the steady-state posterior covariance matrix for x_t:

$$\Sigma = \begin{bmatrix} 0.1943 & 0.1297 & 0.0613 \\ 0.1297 & 0.1640 & -0.0368 \\ 0.0613 & -0.0368 & 0.1482 \end{bmatrix},$$

[3]Maćkowiak, Matějka, and Wiederholt (2018) used a different approach to solve this problem under the information-flow constraint with $\beta = 1$.

which is identical to the solution using the first-order conditions based method discussed in our paper and in Afrouzi and Yang (2019). The steady-state optimal signal is one dimensional and takes the form

$$s_t = 0.9320y_t + 0.3176\epsilon_t + 0.1748\epsilon_{t-1} + v_t,$$

where v_t is a Gaussian white noise with variance 0.6051.

APPENDIX E: INFINITE-HORIZON CASE

We study the following infinite-horizon problem with discounted information costs at time 1:

$$\min_{\{\Sigma_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1}\left[\operatorname{tr}(\Omega\Sigma_t) + \frac{\lambda}{2}\left(\log\det\left(A\Sigma_{t-1}A' + W\right) - \log\det\Sigma_t\right)\right] \tag{E.1}$$

subject to

$$\Sigma_t \preceq A\Sigma_{t-1}A' + W, \quad t = 1, 2, \dots, \Sigma_0 \text{ given.} \tag{E.2}$$

Define the value function as $\mathcal{V}(\Sigma_0)$. By the dynamic programming principle (Stokey and Lucas with Prescott (1989) and Miao (2014)), it satisfies the Bellman equation

$$\mathcal{V}(\Sigma_0) = \min_{\Sigma\in\Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\log\det\left(A\Sigma_0A' + W\right) - \log\det\Sigma\right] + \beta\mathcal{V}(\Sigma),$$

where

$$\Gamma(\Sigma_0) \equiv \left\{\Sigma \succ 0 : \Sigma \preceq A\Sigma_0A' + W\right\}. \tag{E.3}$$

To convert this problem into a DCP, we study an auxiliary problem. Define

$$J(\Sigma_0) \equiv \mathcal{V}(\Sigma_0) - \frac{\lambda}{2}\log\det\left(A\Sigma_0A' + W\right).$$

Then it satisfies the Bellman equation:

$$J(\Sigma_0) = \min_{\Sigma\in\Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta\log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] + \beta J(\Sigma). \tag{E.4}$$

Let $\Sigma = h(\Sigma_0)$ be an associated optimal policy function. The policy function h generates a sequence of optimal covariance matrices $\{\Sigma_t\}_{t=1}^{\infty}$ through $\Sigma_t = h(\Sigma_{t-1})$, $t \geq 1$. Notice that the above problem is not a bounded discounted dynamic programming problem. We use the method of successive approximations (VFI) to analyze it.

Define the value function

$$f_0(\Sigma_0) \equiv \min_{\Sigma\in\Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det(\Sigma). \tag{E.5}$$

Because the constraint set in (E.3) is convex and the log-determinant function is strictly concave, the problem in (E.5) is a convex program and hence $f_0(\Sigma_0)$ is also strictly convex.

Define the Bellman operator $\mathbf{B}$ on the set of functions of positive semidefinite matrices:

$$\mathbf{B}(f)(\Sigma_0) \equiv \min_{\Sigma\in\Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta\log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] + \beta f(\Sigma).$$

Iterating this operator, we can construct a sequence of functions:

$$f_k(\Sigma_0) = \mathbf{B}^k(f_0)(\Sigma_0), \quad k \geq 1. \tag{E.6}$$

By induction and Lemma 2, each function $f_k(\cdot)$ is strictly convex and is obtained by solving a DCP problem. Let the corresponding optimal policy function be $\Sigma = h_k(\Sigma_0)$.

Say a sequence of matrices $\{\Sigma_t\}_{t=1}^{\infty}$ is feasible if $\Sigma_t \in \Gamma(\Sigma_{t-1})$ for each $t \geq 1$.

PROPOSITION 9: *Suppose that* $W \succeq 0$, $AA' + W \succ 0$, $\beta \in (0,1)$, *and* $\Omega \succeq 0$. *For any* $\Sigma_0 \succ 0$, *if there is a feasible sequence of matrices* $\{\Sigma_t\}_{t=1}^{\infty}$ *such that the objective in* (E.1) *is finite*, *then* $f_k(\Sigma_0)$ *increases monotonically to a finite strictly convex limit function* $J(\Sigma_0)$ *as* $k \to \infty$, *which satisfies* (E.4). *Moreover,* $h_k(\Sigma_0)$ *converges to* $h(\Sigma_0)$ *pointwise on any compact set as* $k \to \infty$.

PROOF: We first show that $f_1(\Sigma_0) \geq f_0(\Sigma_0)$. For any $\Sigma \in \Gamma(\Sigma_0)$, let $\Sigma^* \in \Gamma(\Sigma)$ be the optimal solution that attains the value $f_0(\Sigma)$. Then since $\Sigma^* \preceq A\Sigma A' + W$, we have

$$\log\det(A\Sigma A' + W) \geq \log\det(\Sigma^*).$$

It follows that

$$\begin{aligned}
&\operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det(A\Sigma A' + W) - \log\det(\Sigma)\right] + \beta f_0(\Sigma) \\
&= \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det(A\Sigma A' + W) - \log\det(\Sigma)\right] \\
&\quad + \beta\left[\operatorname{tr}(\Omega\Sigma^*) - \frac{\lambda}{2}\log\det(\Sigma^*)\right] \\
&\geq \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det(\Sigma) \geq f_0(\Sigma_0),
\end{aligned} \tag{E.7}$$

where we have used the fact that $\operatorname{tr}(\Omega\Sigma^*) \geq 0$ as $\Omega \succeq 0$ and $\Sigma^* \succ 0$. Minimizing the expression on the first line of (E.7) over $\Sigma \in \Gamma(\Sigma_0)$ yields $f_1(\Sigma_0) \geq f_0(\Sigma_0)$.

It is easy to see that $\mathbf{B}(f) \geq \mathbf{B}(g)$, if $f \geq g$. Thus, we can show that $f_{k+1}(\Sigma_0) \geq f_k(\Sigma_0)$ by induction. By assumption, for any $\Sigma_0 \succ 0$, there is a feasible sequence of matrices $\{\Sigma_t\}_{t=1}^{\infty}$ such that the objective in (E.1) is finite. Thus, the increasing sequence $\{f_k(\Sigma_0)\}$ is bounded above and has a finite limit. Let the limit function be $J(\Sigma_0)$. To show J satisfies (E.4), notice that

$$f_k(\Sigma_0) = \mathbf{B}(f_{k-1})(\Sigma_0) \leq \mathbf{B}(J)(\Sigma_0).$$

On the other hand,

$$J(\Sigma_0) \geq f_k(\Sigma_0) = \mathbf{B}(f_{k-1})(\Sigma_0).$$

Taking limits on the above two inequalities yields $J(\Sigma_0) = \mathbf{B}(J)(\Sigma_0)$.

By induction and Lemma 2, each function $f_k(\Sigma_0)$ is strictly convex and hence the policy function h_k is unique. The limit function J is convex. Since $J = \mathbf{B}(J)$ and the objective function in (E.4) is strictly convex, J is also strictly convex. Thus, the policy function h is also unique. Since f_k is continuous, $f_k(\Sigma_0)$ converges to $J(\Sigma_0)$ uniformly on any compact set. By Theorem 3.8 of Stokey and Lucas with Prescott (1989), $h_k(\Sigma_0)$ converges to $h(\Sigma_0)$ pointwise. *Q.E.D.*

APPENDIX F: EQUILIBRIUM STICKY PRICES

In this appendix, we derive the equilibrium solution for the model in Section 5.1.2 and provide a numerical algorithm to solve the equilibrium. We focus on the steady-state equilibrium. Suppose that the equilibrium aggregate price level can be approximated by a stationary ARMA process: $p_t = \Psi(\mathbf{L})\epsilon_{at}$, where Ψ is given by

$$\Psi(z) = \frac{b_0 + b_1 z + b_2 z^2 + \cdots + b_m z^m}{1 - a_1 z - a_2 z^2 - \cdots - a_r z^r}, \tag{F.1}$$

and z is in the unit circle on the complex space. We solve for an equilibrium with $r \geq m+1$.

As discussed in Section 5.1.2, we can construct a state space representation for firm j:

$$x_{jt} = A x_{j,t-1} + \eta_{jt}, \tag{F.2}$$

$$p^*_{jt} = G x_{jt}, \quad s_{jt} = C_j x_{jt} + v_{jt}, \tag{F.3}$$

where A and G are given in (55) and (56), and $\eta_{jt} = [\epsilon_{jt}, \epsilon_{at}, \epsilon_{at}, 0, \ldots, 0]'$ and v_{jt} are independent Gaussian white noise processes with covariance matrices W and V_j. Assume that v_{jt} satisfies $\int_0^1 v_{jt}\, dj = 0$. Notice that $W \succeq 0$ and $V_{jt} \succ 0$ by our construction. In particular, the $(1, 1)$ entry of W is σ_i^2, the $(2, 2)$, $(2, 3)$, $(3, 2)$, and $(3, 3)$ entries are σ_a^2, and all other entries are zero. We can easily check that W is singular and A is nonsingular.

We solve for the symmetric steady-state information structure under RI for which the posterior covariance matrix Σ for x_{jt} and (C, V) are the same for each firm j. The optimal price under RI for firm j is given by

$$p_{jt} = \mathbb{E}\big[p^*_{jt} | s^t_j\big] = G\mathbb{E}\big[x_{jt} | s^t_j\big] = G\widehat{x}_{jt}. \tag{F.4}$$

The Kalman filter gives

$$\widehat{x}_{jt} = (I - KC)A\widehat{x}_{j,t-1} + K s_{jt}, \tag{F.5}$$

where the Kalman gain is given by

$$K = \big(A\Sigma A' + W\big)C'\big[C\big(A\Sigma A' + W\big)C' + V\big]^{-1}.$$

Using the matrix inversion lemma, we can show that

$$KC = \big(A\Sigma A' + W\big)C'\big[C\big(A\Sigma A' + W\big)C' + V\big]^{-1}C = I - \Sigma\big(A\Sigma A' + W\big)^{-1}, \tag{F.6}$$

which is independent of C and V.

Assume that all eigenvalues of $(I - KC)A$ lie in the unit circle. Using the lag operator $\mathbf{L}$, we can rewrite (F.5) as

$$\widehat{x}_{jt} = X(\mathbf{L})K s_{jt}, \tag{F.7}$$

where

$$X(z) \equiv \big[I - (I - KC)Az\big]^{-1}.$$

It follows from (F.3) and (F.7) that

$$\widehat{x}_{jt} = X(\mathbf{L})KC x_{jt} + X(\mathbf{L})K v_{jt}.$$

Assuming that all eigenvalues of A are in the unit circle, we can rewrite (F.2) as

$$x_{jt} = (I - A\mathbf{L})^{-1}\eta_{jt}.$$

It follows from the preceding two equations that

$$\widehat{x}_{jt} = X(\mathbf{L})KC(I - A\mathbf{L})^{-1}\eta_{jt} + X(\mathbf{L})Kv_{jt}.$$

Aggregating across $j \in [0, 1]$ yields

$$\int_0^1 \widehat{x}_{jt}\, dj = X(\mathbf{L})KC(I - A\mathbf{L})^{-1}M\epsilon_a, \tag{F.8}$$

where $M \equiv [0, 1, 1, 0, \dots, 0]'$ is a $(r + 2)$-dimensional vector and we have used the assumptions

$$\int_0^1 v_{jt}\, dj = 0, \qquad \int_0^1 \epsilon_{jt}\, dj = 0.$$

It follows from (F.4) and (F.8) that the aggregate price level satisfies

$$p_t = \int_0^1 p_{jt} dt = G\int_0^1 \widehat{x}_{jt}\, dj = GX(\mathbf{L})KC(I - A\mathbf{L})^{-1}M\epsilon_a.$$

Given the conjectured form of the equilibrium aggregate price $p_t = \Psi(\mathbf{L})\epsilon_{at}$, we obtain the equilibrium condition:

$$\Psi(z) = GX(z)KC(I - Az)^{-1}M, \tag{F.9}$$

where

$$X(z)KC = \left[I - (I - KC)Az\right]^{-1}KC$$

is independent of (C, V) by (F.6). Equation (F.9) is a functional equation for the coefficients $(a_1, a_2, \dots, a_r, b_0, b_1, \dots, b_m)$. The solution determines the equilibrium pricing function Ψ.

We use the following algorithm to solve for these coefficients.[4]

Step 0. Initialize $k \geq 2$. Let $\{z_1, \dots, z_N\}$ be an evenly spaced grid on $(-1, 1)$ for some integer N.

Step 1. Given a positive integer k, set $r = k$ and $m = k - 1$. Initialize the polynomial coefficients $c \equiv (a_1, a_2, \dots, a_r, b_0, b_1, \dots, b_m)$.

Step 2. Given r, m, and c, compute the values $\{\Psi(z_i)\}_{i=1}^N$, where $\Psi(z)$ is the pricing function given by (F.1).

Step 3. Derive the state space representation in (F.2) and (F.3). Compute the steady-state information structure (C, V, Σ) for the individual RI problem with $\Omega = G'G$. To help convergence, we can use either the golden-rule solution with $\beta = 1$ or the steady-state solution with $\beta \in (0, 1)$ in the previous iteration as the initial guess for the current iteration. The golden-rule solution can be reliably solved using the CVX software.

Step 4. Compute the updated pricing function values

$$\Psi^+(z_i) \equiv GX(z_i)KC(I - Az_i)^{-1}M, \quad i = 1, 2, \dots, N.$$

[4] We have applied the toolbox, Ztran, developed by Han, Tan, and Wu (2019).

Find the updated polynomial coefficients $c^+ \equiv (a_1^+, a_2^+, \dots, a_{r^+}^+, b_0^+, b_1^+, \dots, b_{m^+}^+)$ such that the implied rational function $\Psi^+(z)$ fits the set of values $\{\Psi^+(z_i)\}_{i=1}^N$. Here r^+ and m^+ are the maximal integers such that $a_{r^+}^+ \neq 0$, $b_{m^+}^+ \neq 0$, $r^+ \leq k$, and $r^+ \geq m^+ + 1$.

Step 5. Set $c := c^+$, $r := r^+$, and $m := m^+$. Repeat Steps 2–4 until the relative difference between $\{\Psi^+(z_i)\}_{i=1}^N$ and $\{\Psi(z_i)\}_{i=1}^N$ is within some prespecified tolerance level $\epsilon_1 > 0$.

Step 6. If there is no convergence in Step 5, set $k := k + 1$ and go to Step 1. Otherwise, let the solution obtained in Step 5 be $\Psi^*(z)$. Find a rational function $\hat{\Psi}(z)$ for an ARMA(r, m) process that fits the values $\{\Psi^*(z_i)\}_{i=1}^N$ without the upper bound k restriction on the orders r and m. Check whether the distance between the MA(∞) representations (or the impulse response functions) for the ARMA processes implied by $\hat{\Psi}(z)$ and $\Psi^*(z)$ is within some prespecified tolerance level $\epsilon_2 > 0$. If so, then stop; otherwise, set $k := k + 1$ and go to Step 1.

APPENDIX G: NUMERICAL METHODS

In this appendix, we present our numerical methods to solve for the golden-rule information structure, the steady state, and the transition dynamics for the infinite-horizon RI problem. Our methods also work for the finite-horizon case by suitably modifying the terminal conditions. We have developed a Matlab toolbox to implement these methods.

Our toolbox focuses on the infinite-horizon version of dynamic RI Problem 3. For the LQG control problem, we have $\Omega = F'(R + \beta B'PB)F$. For the pure tracking problem, we have $\Omega = G'G$. Our toolbox works under the assumption that $W \succeq 0$ and $AA' + W \succ 0$. This assumption ensures that dynamic RI problems are convex and the first-order conditions are sufficient for optimality.

G.1. *Golden-Rule Solution*

Solving the golden-rule Problem 5 is simple because it is a static convex program. We simply derive a suitable SDP representation and then apply the CVX software. Specifically, if $W \succ 0$, we use Proposition 2 to derive

$$\min_{\Pi \succ 0, \Sigma \succ 0} (1 - \beta) \operatorname{tr}\bigl(A'PA\Sigma\bigr) + \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}[\log\det W - \log\det \Pi] \tag{G.1}$$

subject to (42) and

$$\begin{bmatrix} \Sigma - \Pi & \Sigma A' \\ A\Sigma & A\Sigma A' + W \end{bmatrix} \succeq 0. \tag{G.2}$$

If A is invertible but W is not, we use Proposition 8 to derive

$$\min_{\Psi, \Sigma \succ 0} (1 - \beta) \operatorname{tr}\bigl(A'PA\Sigma\bigr) + \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2} \log\det \Psi \tag{G.3}$$

subject to (42) and

$$\begin{bmatrix} I - \Psi & M' \\ M & A\Sigma A' + W \end{bmatrix} \succeq 0, \tag{G.4}$$

where $W = MM'$ with $M \succeq 0$.

If neither A nor W is invertible, but $AA' + W$ is invertible, we can apply the CCP algorithm:

ALGORITHM 1—Golden rule: CCP:
Step 1. Guess $\Sigma^{(0)} \succ 0$.
Step 2. Use CVX to solve the linearly convexified problem

$$\min_{\Sigma \succ 0} (1-\beta)\operatorname{tr}\big(A'PA\Sigma\big) + \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\big[g\big(\Sigma;\Sigma^{(0)}\big) - \log\det(\Sigma)\big] \tag{G.5}$$

subject to (42), where

$$g\big(\Sigma;\Sigma^{(0)}\big) \equiv \log\det\big(A\Sigma^{(0)}A' + W\big) + \operatorname{tr}\big(A'\big(A\Sigma^{(0)}A' + W\big)^{-1}A\big(\Sigma - \Sigma^{(0)}\big)\big).$$

Let $\Sigma^{(1)}$ denote the solution.
Step 3. If $\Sigma^{(1)}$ is close to $\Sigma^{(0)}$ up to a prespecified tolerance level, then stop. Otherwise, replace $\Sigma^{(0)}$ by $\Sigma^{(1)}$ and go to Step 2.

The CCP algorithm applies to general optimization problems involving the difference of convex (or concave) functions and is globally convergent to the optimum if this difference function is convex (or concave). For example, we can apply it to all dynamic RI problems studied in our paper.

G.2. *Value Function Based Methods*

To solve for the steady state and transition dynamics starting from any initial prior covariance matrix $\Sigma_{0|-1}$, we first consider the following basic VFI algorithm:

ALGORITHM 2—Basic VFI:
Step 1. Given any $\Sigma_0 \succ 0$, iteratively solve $f_k(\Sigma_0)$ using Bellman equations defined in Appendix E for $k = 0, 1, \ldots$, until convergence at iteration K.
Step 2. Given $\Sigma_{0|-1}$ at $t = 0$, use CVX to solve the following problem:

$$\min_{\Sigma_0 \succ 0} \operatorname{tr}(\Omega\Sigma_0) + \frac{\lambda}{2}\big[\beta\log\det\big(A\Sigma_0A' + W\big) - \log\det(\Sigma_0)\big] + \beta f_K(\Sigma_0)$$

subject to $\Sigma_0 \prec \Sigma_{0|-1}$.
Step 3. Starting from Σ_0 obtained in Step 2, iteratively solve the Bellman equation (E.4) with $J(\Sigma_0)$ replaced by $f_K(\Sigma_{t-1})$ to obtain Σ_t for $t = 1, 2, \ldots$, until Σ_t converges to a steady state.

Notice that we need to use the procedure in the previous subsection to transform all optimization problems in the algorithm into a DCP using an SDP representation. As is well known, the VFI method is slow, but reliable as Proposition 9 guarantees the convergence of the value function. If we just solve for the steady state, we can speed up the algorithm by starting with a good initial guess in Step 1. For example, we can take the golden-rule solution as Σ_0. After getting convergence of $f_k(\Sigma_0)$, we jump to Step 3 directly.

The second way to speed up the algorithm is to use the envelope condition (28) to replace the value function in (20). We then consider the following problem:

$$\min_{\Sigma_t \succ 0} \operatorname{tr}(\Omega\Sigma_t) + \frac{\lambda}{2}\big[\beta\log\det(\Sigma_{t+1|t}) - \log\det(\Sigma_t)\big] - \beta\operatorname{tr}\big(A'\Lambda_{t+1}A\Sigma_t\big) \tag{G.6}$$

subject to $\Sigma_t \preceq \Sigma_{t|t-1}$ for $t \geq 0$, where $\Sigma_{t|t-1} = A\Sigma_{t-1}A' + W$, $\Sigma_{0|-1}$ is exogenously given, and $\Lambda_{t+1} \succeq 0$ is the Lagrange multiplier for the no-forgetting constraint in period $t+1$. We can check that the system of first-order conditions for this problem is the same as the infinite-horizon version of (25), (26), and (27). We can then focus on the solution to the above problem. We first use the following algorithm to solve for the steady state:

ALGORITHM 3—Steady state: Modified VFI using the envelope condition:
Step 1. Start with a guess for $\Lambda^{(0)} \succeq 0$ and $\Sigma^{(0)} \succ 0$.
Step 2. Use CVX to solve the static problem:

$$\min_{\Sigma \succ 0} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] - \beta \operatorname{tr}\left(A'\Lambda^{(0)}A\Sigma\right)$$

subject to $\Sigma \preceq A\Sigma^{(0)}A' + W$. Let $\Sigma^{(1)}$ be the solution and $\Lambda^{(1)}$ denote the Lagrange multiplier for the no-forgetting constraint.
Step 3. If $\Sigma^{(1)}$ and $\Lambda^{(1)}$ are close to $\Sigma^{(0)}$ and $\Lambda^{(0)}$ up to a prespecified tolerance level, then stop. Otherwise, replace $\Sigma^{(0)}$ and $\Lambda^{(0)}$ by $\Sigma^{(1)}$ and $\Lambda^{(1)}$, and go to Step 2.

If we use the golden-rule solution as the initial guess, it takes about 5 seconds for this algorithm to get convergence for the pricing example studied in Section 4. We next use the following algorithm to compute the transition dynamics:

ALGORITHM 4—Transition dynamics: Backward-forward shooting using the envelope condition:
Step 1. Fix a large T. Let Λ_{T+1} be the steady-state Lagrange multiplier. Guess $\{\Sigma_t \succ 0\}_{t=0}^{T}$.
Step 2. Compute $\Sigma_{t+1|t} = A\Sigma_t A' + W$ for $t = 0, 1, \ldots, T$. Use CVX to solve problem (G.6) backward to obtain $\{\Sigma_t^*\}_{t=0}^{T}$ and $\{\Lambda_t\}_{t=0}^{T}$. In each period t, we take Λ_{t+1} obtained in period $t+1$ as given.
Step 3. Update $\{\Sigma_t\}_{t=0}^{T} := \{\Sigma_t^*\}_{t=0}^{T}$ and go to Step 2. Iterate until convergence.

Notice that, for all algorithms to solve for the transition dynamics, we need to check whether T is large enough such that Σ_T indeed reaches the steady state. The third way to increase speed is to notice that the dynamic RI problem can be viewed as a sequence of static RI problems as established in the proof of Proposition 3. We use the following algorithm to compute the steady state. This algorithm applies to the weaker invertibility assumption of $AA' + W$ with no extra effort of deriving an SDP representation and thus it is our preferred algorithm.

ALGORITHM 5—Steady state: Modified VFI based on a sequence of static RI problems:
Step 1. Start with a guess for $\Theta \succeq 0$ and $\Sigma_p \succ 0$.
Step 2. Use CVX to solve the following static problem:

$$\min_{\Sigma \succ 0} \operatorname{tr}(\Theta\Sigma) - \frac{\lambda}{2} \log\det(\Sigma)$$

subject to $\Sigma \preceq \Sigma_p$. Let Σ^* and Λ^* denote the solution for the posterior covariance matrix and the Lagrange multiplier for the no-forgetting constraint.

Step 3. Compute the updated value: $\Sigma_p^* = A\Sigma^* A' + W$ and

$$\Theta^* = \Omega + \frac{\beta\lambda}{2} A'\Sigma_p^{*-1} A - \beta A'\Lambda^* A.$$

Step 4. If Θ^* and Σ_p^* are close to Θ and Σ_p within a prespecified tolerance level, then stop. Otherwise, replace Θ and Σ_p by Θ^* and Σ_p^* and go to Step 2.

The following algorithm computes the transition dynamics.

ALGORITHM 6—Transition dynamics: Backward-forward shooting based on a sequence of static RI problems:
Step 1. Fix a large T. Take Λ_{T+1} and $\Sigma_{T+1|T}$ as their steady-state values.
Step 2. Start with a guess for $\Sigma_{t+1|t}$ for $t = 0, 1, \dots, T-1$.
Step 3. Use CVX to solve the sequence of static problems (B.10) backward starting from time T. At time t, let

$$\Theta_t = \Omega + \frac{\beta\lambda}{2} A'\Sigma_{t+1|t}^{-1} A - \beta A'\Lambda_{t+1} A,$$

where Λ_{t+1} is obtained in period t. The solution to time-t problem gives the posterior covariance matrix Σ_t and the Lagrange multiplier Λ_t for the no-forgetting constraint.
Step 4. Compute $\Sigma_{t+1|t}^* = A\Sigma_t A' + W$ forward for $t = 0, 1, \dots, T-1$. If $\{\Sigma_{t+1|t}^*\}_{t=0}^{T-1}$ and $\{\Sigma_{t+1|t}\}_{t=0}^{T-1}$ are close enough within a prespecified tolerance level, then stop. Otherwise, replace $\{\Sigma_{t+1|t}\}_{t=0}^{T-1}$ by $\{\Sigma_{t+1|t}^*\}_{t=0}^{T-1}$ and go to Step 3.

For the pricing example in Section 4, it takes about 72 seconds for this algorithm to converge starting from the initial prior $\Sigma_{0|-1} = 0.5W$.

G.3. *First-Order Conditions Based Methods*

In this subsection, we use the first-order conditions characterized in Proposition 3 to solve for the steady state and the transition dynamics. The steady state can be solved by brute force fixed-point iteration as described in Section 3.2 or in Afrouzi and Yang (2019). Specifically, start with an initial guess for Σ and Θ. Use (36) to solve for U, D, and Σ_p. Then use (34) and (35) to solve for the updated Σ and Θ. Iterate until convergence of Σ and Θ. Recently, Afrouzi and Yang have developed a Julia toolbox to solve for the transition dynamics. Here we propose the following algorithm that works for both finite- and infinite-horizon RI problems.

ALGORITHM 7—Transition dynamics: Backward-forward shooting based on first-order conditions:
Step 1. Fix a sufficiently large T and set Θ_T to its steady-state value. Guess $\{\Sigma_t \succ 0\}_{t=0}^{T-1}$.
Step 2. Compute

$$\Sigma_{t+1|t} = A\Sigma_t A' + W, \quad t = 0, 1, \dots, T-1, \tag{G.7}$$

and perform the eigendecomposition as in (30).
Step 3. Compute $\{\Theta_t\}_{t=0}^{T-1}$ backward given Θ_T using (32). Compute the updated sequence $\{\Sigma_t^*\}_{t=0}^{T-1}$ forward given $\Sigma_{0|-1}$ using (31).

Step 4. If the difference between $\{\Sigma_t^*\}_{t=0}^{T-1}$ and $\{\Sigma_t\}_{t=0}^{T-1}$ under some norm is smaller than a prespecified tolerance level, then stop. Otherwise, replace $\{\Sigma_t\}_{t=0}^{T-1}$ by $\{\Sigma_t^*\}_{t=0}^{T-1}$ and go to Step 2.

It takes about 0.05 seconds for this algorithm implemented in Matlab to converge for the pricing example in Section 4. This method is as fast as the Julia toolbox of Afrouzi and Yang (2019). Unlike the VFI method, both their method and our shooting method do not guarantee convergence as a formal convergence proof is unavailable.

G.4. *Discussion*

As is well known in the mathematics and economics literature on dynamic optimization problems, both dynamic programming and first-order conditions (Euler equations) are used to characterize solutions. Numerical methods based on both value functions and first-order conditions have been widely developed in the literature. These methods often complement each other and no single method can universally dominate the others. For example, methods based on first-order conditions are typically much faster, but may be sensitive to initial values. Methods based on value functions are slower, but are more reliable because a convergence result is often available. More importantly, the value function based methods are flexible to incorporate many occasionally binding constraints and nonsmooth objectives such as pricing problems with menu costs.

To illustrate this point, suppose that there is an additional technological constraint on information processing so that the entropy conditional on observing a history of signals s^t satisfies

$$H\left(x_t|s^t\right) \geq L \text{ for some } L > 0, \quad t \geq 0.$$

As entropy measures the amount of uncertainty, the above constraint means that there is a limit on the decision maker's ability to reduce uncertainty. For our LQG RI model, the above constraint can be written as

$$\log\det(\Sigma_t) \geq l \quad \text{or some } l. \tag{G.8}$$

We now impose this constraint in the pricing example of Section 4. The new problem cannot be solved using the first-order conditions in Proposition 3. But we can incorporate this constraint easily using our dynamic programming formulation.

Consider an example with $l = -0.01$ and other parameter values in (47). Using the modified VFI method based on either the envelope condition or a sequence of static RI problems, it takes about 5 seconds to get convergence to the steady-state posterior covariance matrix:

$$\Sigma = \begin{bmatrix} 0.9916 & -0.0017 \\ -0.0017 & 0.9984 \end{bmatrix}.$$

We can check that the entropy constraint (G.8) binds in the steady state. Compared with (48), the steady-state posterior variances of the two shocks are higher.

REFERENCES

AFROUZI, H., AND C. YANG (2019): "Dynamic Rational Inattention and the Phillips Curve," working paper, Columbia University. [13,14,21,22]

COVER, T. M., AND J. A. THOMAS (2006): *Elements of Information Theory* (second Ed.). Wiley. [6]

HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton NJ: Princeton University Press. [13]
HAN, Z., F. TAN, AND J. WU (2019): "Analytic Policy Function Iteration," Working Paper, St. Louis University. [17]
HARVILLE, D. A. (1997): *Matrix Algebra From a Statistician's Perspective*, Vol. 1. Springer. [3]
HORN, R. A., AND C. R. JOHNSON (2013): *Matrix Analysis* (second Ed.). NY: Cambridge University Press. [2]
LIPP, T., AND S. BOYD (2016): "Variations and Extension of the Convex–Concave Procedure," *Optimization and Engineering*, 17, 263–287. [13]
MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2018): "Dynamic Rational Inattention: Analytical Results," *Journal of Economic Theory*, 176, 650–692. [13]
MIAO, J. (2014): *Economic Dynamics in Discrete Time*. Cambridge, MA: the MIT Press. [14]
SIMS, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690. [10]
STOKEY, N. L., AND R. E. JR. LUCAS, with E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Cambridge MA: Harvard University Press. [14,15]
WANG, B., B. XI, AND F. ZHANG (1999): "Some Inequalities for Sum and Product of Positive Semidefinite Matrices," *Linear Algebra and its Applications*, 293, 39–49. [2]

Co-editor Giovanni L. Violante handled this manuscript.

Manuscript received 4 February, 2020; final version accepted 20 August, 2021; available online 1 September, 2021.