Simple Models and Biased Forecasts*

Pooya Molavi†

March 28, 2022

This paper proposes a general framework in which agents are constrained to use simple time-series models to forecast economic variables and characterizes the resulting bias in the agents' forecasts. It considers agents who can only entertain state-space models with no more than d states, where d measures the agents' cognitive abilities. Agents' models are otherwise unrestricted a priori and disciplined endogenously by maximizing the fit to the true process. When the true process does not have a d-state representation, agents end up with misspecified models and biased forecasts. If the true process satisfies an ergodicity assumption, the bias manifests itself as persistence bias: a tendency to attend to the most persistent observables at the expense of less persistent ones. The bias causes agents' foreword-looking decisions to mimic the dynamics of backward-looking, persistent variables in the economy. It also dampens the response of agents' actions to shocks and leads to additional co-movements between various choices. The paper then proceeds to study the implications of the theory in the context of three calibrated workhorse macro models: the new-Keynesian, real business cycle, and Diamond–Mortensen–Pissarides models. In each case, constraining agents to use simple models brings the model's predictions more in line with the data, without adding any parameters other than the integer d.

*I am grateful to Marios Angeletos, Larry Christiano, Nico Crouzet, Ian Dew-Becker, Marty Eichenbaum, Timothy Feddersen, Laura Gáti, Yuriy Gorodnichenko, Lars Hansen, Ali Jadbabaie, Chen Lian, Annie Liang, Joshua Mollner, Emi Nakamura, Kris Nimark, Serena Ng, Alessandro Pavan, Giorgio Primiceri, Matt Rognlie, Jón Steinsson, Bruno Strulovici, Alireza Tahbaz-Salehi, Harald Uhlig, Olivier Wang, Iván Werning, Mike Woodford, and seminar participants at Cornell University, Northwestern University, UC Berkeley, and Saieh Family Fellows Conference for their helpful comments and discussions.

†Northwestern University, pmolavi@kellogg.northwestern.edu.

1 Introduction

Forecasting is hard, especially about the future—or so goes a famous saying. Yet rational-expectations models assume that agents can forecast the future as if they knew the true model of the economy. In reality, when faced with the difficult task of forecasting in a complex world, agents are bound to rely on simple models. The unrealism of the rational-expectations assumption would not be of great concern had the predictions of workhorse macro models been robust to alternative specifications of expectations. However, the answers to several important questions in macro, ranging from the determinacy of equilibrium in new-Keynesian models and the power of forward guidance to the consequences of government debt and the optimal path of fiscal policy, are highly sensitive to the specification of expectations.

This paper proposes a framework in which agents are constrained to forecast using simple time-series models and characterizes the resulting bias in their forecasts and actions. It considers the problem of an agent attempting to forecast future values of a set of observables based on their past realizations. The observables follow a stochastic process, which may not have a simple representation, whereas the agent can only entertain stochastic processes that can be represented by a low-dimensional state-space model. Consequently, the agent may end up with a low-dimensional approximation to the true process. This form of model misspecification leads to bias in the agent's forecasts and deviations in her actions from the rational-expectations benchmark. The paper characterizes this bias and studies its implications for the agent's actions, and ultimately, macroeconomic aggregates.

The bias in the agent's forecasts takes a straightforward form when the true process satisfies an ergodicity assumption. The bias then takes the form of *persistence bias*—a tendency to attend to the most persistent observables at the expense of less persistent ones. The agent forecasts the most persistent observables as well as an agent with rational expectations while treating less persistent observables as if they were i.i.d. This selective attention to persistent variables leads to the stickiness of expectations and actions. It also anchors the agent's foreword-looking decisions to sluggish, backward-looking observables and increases the co-movement between various actions.

The paper's sharp theoretical results offer a powerful toolbox, which helps one study how bounded rationality alters the predictions of standard macroeconomic models and the associated policy prescriptions. I illustrate the use of these tools in the context of three workhorse macro models: the new-Keynesian model, the real business cycle (RBC) model, and the Diamond–Mortensen–Pissarides (DMP) model. Constraining agents to simple models in the new-Keynesian model dampens the response of output and inflation to interest rate changes and significantly reduces the power of forward guidance. Doing so in the RBC model changes the behavior of aggregate consumption by making it act more like a stock variable, which responds sluggishly to productivity shocks. Constraining agents to simple models in a DMP model with correlated productivity and separation shocks leads to significant amplification and propagation of separation shocks and a negative co-movement between the unemployment rate and the number of vacancies in response to both shocks.

Section 2 presents the general framework. A population of agents observe a sequence of observables over time and use their past observations to forecast the future values of the observables. The observables follow a stationary and ergodic stochastic process that satisfies weak technical assumptions. However, the true process may be a complex object, which does not have a (finite-dimensional) parametric representation.

Agents do not have access to the true process. They instead rely on low-dimensional parametric representations of the true process when forming their forecasts. In particular, they use linear-Gaussian state-space models with at most d state variables. The set of models agents can entertain is otherwise flexible. For instance, it includes all ARMA models of appropriate order.

The only free parameter in the specification of agents' expectations is the dimension d of their models. Meanwhile, the matrices that parametrize agents' models are determined endogenously as they fit their models to the true process.

I assume that agents settle on d-state *pseudo-true* models, stochastic processes with d-state representations that minimize the Kullback–Leibler divergence rate (KLDR) from the true process. The KLDR generalizes the usual notion of KL divergence to (non-i.i.d.) stochastic processes. The choice of divergence is based on a Bayesian learning result: Agents who start with a full-support prior over the set of d-state models and update their beliefs using Bayes' rule will asymptotically forecast *as if* they were using a pseudo-true d-state model. This result holds irrespective of agents' preferences or the details of any decision problem they may be facing.

Section 3 characterizes the set of pseudo-true d-state models. I start the analysis by establishing a valuable *linear-invariance* property for the class of pseudo-true state-space models: A linear transformation of the vector of observables leads to an analogous transformation of the set of pseudo-true models. This property implies that the framing of agents' observations does not affect how they form their expectations. Agents' forecasts and actions only depend on the amount of information available to them and not on how it is presented. Neither are agents' expectations changed by augmenting the vector of observables with variables that are linear combinations of variables already in their information set. The linear-invariance property makes the framework immediately applicable in macro applications, even when there is no unique or obvious way of defining the vector of observables.

The linear-invariance result also illustrates a dichotomy in the agents' understanding of *intratemporal* (or cross-sectional) and *intertemporal* (or time-series) relationships among observables. Agents uncover all linear intratemporal relationships among observables, but they can only entertain intertemporal relationships mediated through a small number of persistent states. While arguably stark, this dichotomy brings into sharp focus the paper's premise that forecasting is difficult because it requires forecasters to recognize stochastic patterns that unfold over time.

It also sets this paper apart from the extensive literature that focuses on the difficulty of paying attention to a large cross-section of observables, such as in the rational inattention framework of Woodford (2003) and Sims (2003) and Gabaix (2014)'s model of sparsity.

I proceed to characterize agents' forecasts when they use pseudo-true d-state models, starting with the $d = 1$ case. Agents' forecasts given a pseudo-true 1-state model can be expressed in terms of four endogenous variables: perceived *persistence*, perceived *noise*, vector of *relative attention*, and vector of *relative sensitivity*. Persistence captures agents' belief about the persistence of the subjective state of the economy. Noise determines agents' belief about the noisiness of observables—when seen as signals of the subjective state. Observables that influence the agents' estimate of the subjective state by more have greater relative attention, and agents' forecasts of observables with larger relative sensitivity are more sensitive to changes in the agents' estimate of the subjective state.

These endogenous variables take simple forms when the true process satisfies an ergodicity condition, which I refer to as *exponential ergodicity*, and is satisfied in all my applications. The perceived persistence is then equal to the top eigenvalue of the autocorrelation matrix at lag one, the relative attention and relative sensitivity are equal to transformations of the corresponding eigenvector, and the perceived noise is zero. Agents' attention is focused on the most persistent component of the vector of observables when estimating the subjective state, and the perceived persistence of the subjective state is equal to the persistence of the most persistent component.

The remainder of Section 3 generalizes and marks the limits of the characterization result just discussed. I show that the result generalizes (under slightly stronger assumptions) to the $d > 1$ case: Agents constrained to d-state models track the d most persistent components of the vector of observables. I also show that the variance-covariance matrix of observables is identical under the pseudo-true and true models. In other words, agents constrained to simple models fully uncover the cross-sectional correlations in observables. Finally, I argue that without a richness assumption on observables, the true process may fail to be exponentially ergodic, and the perceived noise may be non-zero.

Section 4 develops the implications of the bias in agents' forecasts for their behavior. I do so by augmenting the general framework with a reduced-form specification of agents' actions, in which actions in each period linearly depend on forecasts of present discounted values of the observables. I show that agents' actions then exhibit three general properties: First, actions respond less to shocks than under rational expectations. Second, actions look as if they were responding to a small number of shocks. Third, agents' various actions co-move with each other more than under rational expectations.

Sections 5–7 of the paper illustrate the framework's versatility by applying it to three microfounded and calibrated macro models. As the first application, in Section 5, I consider a version of the standard new-Keynesian model in which agents are constrained to forecast using pseudo-true 1-state models. As in the rational expectations version of the model, the equilibrium has a simple linear representation, which can be computed analytically without going to the computer. The first result of this section is a characterization of the equilibrium in a regime without forward guidance.

I then use the equilibrium characterization to study the implications of bounded rationality for the conduct of monetary policy. Several new insights arise from the analysis. First, the monetary authority generically faces a trade-off between closing the output gap and achieving stable prices. The "divine coincidence" holds only in the knife-edge cases in which, in equilibrium, agents pay no attention to nominal interest rate or their inflation expectations are entirely insensitive to their estimate of the subjective state.

Second, conventional monetary policy and foreword guidance are less potent than under rational expectations. The stickiness of agents' expectations dampens the equilibrium response of output and inflation to interest rate changes. The fact that information has to be filtered through a low-dimensional model before agents can incorporate it into their forecasts, meanwhile, lowers the impact of forward guidance on output and inflation. Furthermore, the power of forward guidance is largely independent of the duration of guidance, unlike in the rational expectations version of the model.

Section 6 presents the application to the standard RBC model. The RBC model is an excellent case study in that it has only one exogenous shock and two state variables. Therefore, if agents are constrained to d-state models with $d \geq 2$, they recover the true process, and their expectations coincide with rational expectations. When $d = 1$, on the other hand, agents' models will be pseudo-true, and their forecasts will be biased. This prediction of the model distinguishes it from signal-extraction-type models, which revert to rational expectations when there is a single exogenous shock in the economy.

Constraining agents to one-dimensional models in the RBC model causes aggregate consumption to behave more like a stock variable. This prediction directly follows from the persistence bias in agents' expectations. Agents' estimate of the subjective state mostly depends on the value of the most persistent variable—the capital stock in equilibrium for the RBC model. Consequently, consumption—an almost purely forward-looking variable—almost perfectly comoves with the value of the capital stock. The anchoring of consumption to the capital stock makes consumption more sluggish and more volatile than under rational expectations.

For the last application, in Section 7, I study how the predictions of the DMP model change when agents are constrained to use simple models. The standard DMP model has difficulty generating realistic fluctuations in the unemployment rate, the number of vacancies, and job-finding rate, a fact known as the Shimer (2005) puzzle. I show that constraining agents to use simple models goes toward resolving this puzzle.

I consider a standard calibration of the DMP model with labor productivity and separation rate shocks. The equilibrium has a 3-state representation, so expectations of agents constrained to d-state models with $d \geq 3$ coincide with rational expectations. I instead consider agents constrained to use 1-state models. In equilibrium, agents' estimate of the subjective state closely tracks the evolution of the unemployment rate. Separation rate shocks increase the unemployment rate, thus making agents pessimistic about the state of the economy. The result is a *drop* in vacancy creation following an increase in separations and a negative co-movement between the unemployment rate and vacancies in response to the separation shock. Meanwhile, the stickiness of expectations slows the dynamics of the economy, thus improving the propagation mechanism of the model.

Related Literature. This paper belongs to the broad literature that studies the implications of bounded rationality in macroeconomics. Some recent contributions to this literature include Garcia-Schmidt and Woodford (2015), Farhi and Werning (2017), Bordalo, Gennaioli, and Shleifer (2018), Woodford (2018), Bhandari, Borovička, and Ho (2019), Da Silveira, Sung, and Woodford (2020), Afrouzi, Kwon, Landier, Ma, and Thesmar (2021), Bianchi, Ilut, and Saijo (2021), and Vimercati, Eichenbaum, and Guerreiro (2022). The paper's focus on biased forecasts that result from the agents' use of low-dimensional state-space models sets it apart from other papers in this literature.

A closely related literature studies macroeconomic implications of imperfect knowledge of payoff-relevant variables, either because of exogenous observation noise or stickiness of information, e.g., Lucas (1972), Mankiw and Reis (2002), and Angeletos and La'O (2009), or due to costly attention, e.g., Sims (2003), Woodford (2003), and Maćkowiak and Wiederholt (2009).[1] This paper abstracts from the difficulty of observing a large *cross-section* of variables and instead focuses on the difficulty of comprehending complex *time-series* relationships. The predictions of this framework also distinguish it from the literature mentioned above: In my model, agents fully uncover cross-sectional relationships among variables, but their expectations could deviate from rational expectations even if the economy has a single exogenous shock.

This paper is closest in spirit within behavioral macroeconomics to the learning literature, which goes back to Marcet and Sargent (1989a,b) and Sargent (1999).[2] The paper shares this literature's approach to modeling economic agents as econometricians who use estimated statistical models to make sense of their world. However, it deviates from the earlier works in that literature in several important ways. First, I assume that the agents' model of the economy is a state-space model, which is fully flexible except in its dimension. Second, I consider Bayesian agents and focus on the limit of when learning is complete, and agents have settled on some pseudo-true

[1] Other related papers in the literature on noisy and sticky information and rational inattention include Caplin and Dean (2013), Gabaix (2014, 2020), Alvarez, Lippi, and Paciello (2015), Angeletos and Lian (2016, 2018), Angeletos and Huo (2021), Angeletos and Sastry (2021) and Chahrour, Nimark, and Pitschner (2021).

[2] Other related papers in the learning literature include Cho, Williams, and Sargent (2002), Bullard and Mitra (2002), Marcet and Nicolini (2003), Orphanides and Williams (2005), Preston (2005), Bullard, Evans, and Honkapohja (2008), Adam and Marcet (2011), Eusepi and Preston (2011), Adam, Kuang, and Marcet (2012), Malmendier and Nagel (2016), Adam, Marcet, and Beutel (2017), Eusepi and Preston (2018a), and Gáti (2020). See Evans and Honkapohja (2012) for a textbook treatment and Eusepi and Preston (2018b) for a recent survey of the literature.

model. Finally, I accommodate the possibility that the agents' pseudo-true model differs from the true model due to misspecification.

This paper also contributes to the literature that studies the properties of pseudo-true models. The term pseudo-true model originates in the pioneering work of Sawa (1978), who proposes the use of KL divergence as a model-selection criterion. Agents in the restricted-perceptions equilibrium of Bray (1982) and Bray and Savin (1986), Rabin and Vayanos (2010)'s model of the gambler's fallacy, the natural-expectations framework of Fuster, Laibson, and Mendel (2010) and Fuster, Hebert, and Laibson (2012), the Berk–Nash equilibrium of Esponda and Pouzo (2016, 2021), and the constrained rational expectations equilibrium of Molavi (2019) all use pseudo-true models to forecast payoff-relevant variables. However, despite this long history, surprisingly few general results on the properties of pseudo-true models have appeared in the literature. Such results are almost exclusively derived—with the notable exception of Rabin and Vayanos (2010)—in settings where the set of models is sufficiently restricted that the pseudo-true model can be estimated using OLS regressions, and the bias in agents' forecasts reduces to the omitted-variable bias.

The state-space models used in this paper are relatives of dynamic factor models, e.g., Stock and Watson (2011). However, the two are distinct mathematically and conceptually. Dynamic factor and state-space models offer two alternative representations of stochastic processes.[3] Each representation suggests a conceptually different decomposition of time-series data. Dynamic factor models decompose data into common factors and idiosyncratic disturbances, whereas for state-space models, the decomposition is into persistent and transitory components. The two approaches thus suggest two different simplifications of large time-series data: using a small number of common factors in the former case and a small number of persistent states in the latter case.

Finally, Molavi, Tahbaz-Salehi, and Vedolin (2021) use a closely related framework to study the implications of bounded rationality for asset prices and returns. They show that constraining the complexity of investors' models leads to return predictability and provides a parsimonious account of several puzzles in the asset-pricing literature.

2 General Framework

In this section, I present the environment and the main behavioral assumption of the paper.

2.1 The Environment

Time is discrete and is indexed by $t \in \mathbb{Z}$. There is a measure of identical agents, each observing a sequence of observables over time and using their past observations to forecast the future values of observables. I let $y_t \in \mathbb{R}^n$ denote the time-t value of the vector of observables, or simply the

[3] The sets of time series that can be represented by dynamic factor and state-space models are not nested. Instead, any finite dynamic factor model has a state-space representation, and any finite state-space model has a dynamic factor representation. See Forni and Lippi (2001) for a representation result for the (generalized) dynamic factor models.

observable. The observable follows a mean-zero stochastic process with distribution $\mathbb{P}$ and the corresponding expectation operator $\mathbb{E}[\cdot]$.

I make several technical assumptions on the true process $\mathbb{P}$. The process $\mathbb{P}$ is stationary ergodic with $\mathbb{E}[\|y_t\|^2] < \infty$. Furthermore, there exists a linear subspace $\mathcal{W}$ of $\mathbb{R}^n$ such that $\mathbb{P}(y_1, \ldots, y_t)$ is absolutely continuous with respect to the restriction of the Lebesgue measure to $\mathcal{W}^t$ for any t, with density $\mathbb{f}(y_1, \ldots, y_t)$.[4] Finally, I assume that the true process has finite entropy rate, i.e., $\lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[-\log \mathbb{f}(y_1, \ldots, y_t)\right] < \infty$. These assumptions are all quite weak. They are satisfied, for instance, if y_t follows a vector ARMA process with Gaussian innovations.

Agents have perfect information about the past realizations of the observable, with their time-t information set given by $\{y_t, y_{t-1}, \ldots\}$. However, they may use a misspecified model to map their information to their forecasts. This model misspecification leads to deviations in the agents' forecasts from those that arise in the rational-expectations benchmark.

2.2 Simple Models

As the main behavioral assumption of the paper, I assume that agents are bound to use simple state-space models to forecast the vector of observables. They can only entertain models of the form

$$\begin{aligned} z_t &= Az_{t-1} + w_t, \\ y_t &= B' z_t + v_t, \end{aligned} \tag{1}$$

where $z_t \in \mathbb{R}^d$ is a vector of (subjective) state variables, $A \in \mathbb{R}^{d\times d}$ is a convergent matrix, $w_t \in \mathbb{R}^d$ is i.i.d. $\mathcal{N}(0, Q)$, $B \in \mathbb{R}^{d\times n}$, $v_t \in \mathbb{R}^n$ is i.i.d. $\mathcal{N}(0, R)$, and B' denotes the transpose of B.[5] Formally, I define a *d-state model* as a stationary ergodic stochastic process for y_t that can be represented as in equation (1) with z_t a d-dimensional state variable. Whenever there is no risk of confusion, I use the term d-state model to refer both to the stochastic process for y_t and the parameters $\theta \equiv (A, B, Q, R)$ of its state-space representation. I let Θ_d denote the set of all d-state models, let P^θ denote the stationary distribution over $\{y_t\}_{t=-\infty}^{\infty}$ induced by model θ, and let $\mathcal{P}_d \equiv \left\{P^\theta : \theta \in \Theta_d\right\}$. With slight abuse of notation, I write $\Theta_d \subseteq \Theta_{d+1}$ to stress the fact that agents with a larger d can entertain a larger class of models.

The integer d is a primitive that captures the agents' sophistication in modeling the stochastic process for the vector of observables, with larger values of d amounting to agents who can entertain more complex models. When d is sufficiently large, the agents' set of models is large enough to contain good approximations to any true process $\mathbb{P}$. But when d is small relative to the number of states required to model the true process, no model in the agents' set of models will provide a good approximation to $\mathbb{P}$. Agents then necessarily end up with a misspecified model of

[4] This assumption is weaker than the assumption that $\mathbb{P}(y_1, \ldots, y_n)$ is absolutely continuous with respect to the Lebesgue measure over $\mathbb{R}^{nt}$ since it allows for the possibility that the true process is degenerate. This additional level of generality will be useful in applications where the elements of y_t may be linearly dependent.

[5] A matrix is convergent if all of its eigenvalues are smaller than one in magnitude. Matrix A being convergent is sufficient for the agents' model to define a stationary ergodic process.

the true process and biased forecasts—regardless of which model in the set $\mathcal{P}_d$ they use to make their forecasts. Characterizing this bias is the focus of the next section of the paper.

My preferred rationale for the constraint on the number of states is to capture limits on the agents' cognitive abilities, but the constraint can also arise from the agents' rational fear of overfitting. Models with a large number of parameters and many degrees of freedom are prone to overfitting. Such concerns may lead rational agents to limit themselves to statistical models with a small number of parameters, especially if they only have a short time series to draw upon when estimating the parameters of their model. In the remainder of the paper, I abstract away from any issues arising from small samples and instead consider the long-run limit where the sampling error vanishes.

2.3 Pseudo-True Models

I assume that agents forecast using models in the family of d-state models that provide the best fit to the true process. I use the Kullback–Leibler divergence rate of process P^θ from the true process $\mathbb{P}$ as the measure of the fit of model θ.[6] The *Kullback–Leibler divergence rate* (KLDR) of P^θ from $\mathbb{P}$ is denoted by $\mathrm{KLDR}(\theta)$ and defined as follows. Recall that the true process is supported on a linear subspace $\mathcal{W}$ of $\mathbb{R}^n$. If P^θ is also supported on $\mathcal{W}$, then

$$\mathrm{KLDR}(\theta) \equiv \lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\log\left(\frac{\mathbb{f}(y_1,\ldots,y_t)}{f^\theta(y_1,\ldots,y_t)}\right)\right],$$

where $f^\theta(y_1,\ldots,y_t)$ denotes the density of P^θ with respect to the restriction of the Lebesgue measure to $\mathcal{W}^t$; if P^θ is not supported on $\mathcal{W}$, then $\mathrm{KLDR}(\theta) \equiv +\infty$.

The Kullback–Leibler divergence rate is the natural generalization of the notion of Kullback–Leibler (KL) divergence to stationary stochastic processes. In the i.i.d. case, the KL divergence of a candidate model from the true model captures the difficulty of rejecting the candidate model in favor of the true model using a likelihood-ratio test. That is why the KL divergence is commonly used as a measure of the fit of a model.[7] Similarly, the $\mathrm{KLDR}(\theta)$ captures the rate at which the power of a test for separating a stochastic process P^θ from the true process $\mathbb{P}$ approaches one as $t \to \infty$.[8] The KLDR is also tightly linked to asymptotics of Bayesian learning, as I discuss in the following subsection.

Model $\theta^* \in \Theta_d$ is a *pseudo-true* d-state model if $\mathrm{KLDR}(\theta^*) \le \mathrm{KLDR}(\theta)$ for all $\theta \in \Theta_d$. If the agents' set of models contains a model θ such that $f^\theta(y_1,\ldots,y_t) = \mathbb{f}(y_1,\ldots,y_t)$ almost everywhere and for all t, then any pseudo-true d-state model is observationally equivalent to the true process. The set $\mathcal{P}_d$ of distributions is then correctly specified. When no such d-state model exists, $\mathrm{KLDR}(\theta) > 0$ for any model $\theta \in \Theta_d$, and the set $\mathcal{P}_d$ is misspecified. I let Θ_d^* denote the set of all

[6] The mean-squared forecast error is another commonly used notion of fit. In Appendix B, I define the weighted mean-squared forecast error and show that it is equivalent to the Kullback–Leibler divergence rate under an appropriate choice of the weighting matrix.

[7] See, for instance, Hansen and Sargent (2008).

[8] See, for instance, Shalizi (2009).

pseudo-true d-state models, and let $\mathcal{P}_d^* \equiv \{P^\theta : \theta \in \Theta_d^*\}$. The following result shows that pseudo-true models are observationally equivalent to the true process when the set of models is correctly specified:

Theorem 1. *Suppose the set $\mathcal{P}_d$ of d-state models is correctly specified. Then any pseudo-true d-state model $P^{d*} \in \mathcal{P}_d^*$ is observationally equivalent to the true process $\mathbb{P}$.*

2.4 Learning Foundation

Pseudo-true models arise naturally as the long-run outcome of learning by Bayesian agents with misspecified priors. Consider an agent who starts with prior μ_0 with full support over the points in the set $\mathbb{R}^d \times \Theta_d$, each corresponding to an initial value of the subjective states, z_0, and a d-state model, θ, which describes how states and the observable co-evolve. Suppose the agent observes y_t over time and updates her belief using Bayes' rule. Let μ_t denote the agent's time-t Bayesian posterior over $\mathbb{R}^d \times \Theta_d$. Berk (1966)'s theorem establishes that, in the limit $t \to \infty$, the agent's posterior will assign probability one to the set of pseudo-true models.[9]

This result offers an "as if" interpretation of pseudo-true d-state models. One can assume that every agent has a subjective prior—which may be different from the true distribution—and updates her belief in light of new information using Bayes' law. By Berk's theorem, any such agent whose prior is supported on the set of d-state models will forecast the observable in the long run *as if* she were using a pseudo-true d-state model. Focusing on pseudo-true models instead of Bayesian posteriors allows me to do away with the sampling variance in the agents' posteriors and their forecasts and instead focus on the asymptotic bias resulting from misspecification.[10]

Note that the set of pseudo-true d-state models is independent of the agents' preferences. Instead, it only depends on the number of states agents can entertain and the true stochastic process. The independence of the agents' pseudo-true models from their preferences is evident given the "as if" interpretation discussed above: Two agents who start with identical priors, observe the same sequence of observations, and update their beliefs using Bayes' rule will end up with identical posteriors at any point in time—irrespective of their preferences. Berk's theorem goes a step further by establishing that, in the long run, the posterior only depends on the support of the prior (not its other details) and the distribution of observations (not their realizations).

The independence of the agents' pseudo-true models from their preferences has a significant consequence: The set of pseudo-true d-state models is generically disjoint from the set of d-state models that maximize the agents' payoffs. However, this disparity is a feature, not a bug, of a positive theory of bounded rationality. While finding the payoff-maximizing model requires

[9] While Berk (1966) only covers the case of i.i.d. observations and parametric models, the result has seen been extended much more generally. Bunke and Milhaud (1998) and Kleijn and Van Der Vaart (2006) substantially extend Berk (1966) by providing conditions for the weak convergence of posterior distributions and considering infinite-dimensional models. Shalizi (2009)'s extension of Berk's theorem covers the case of non-i.i.d. observations and hidden Markov models.

[10] One can alternatively consider agents who estimate the parameters of their d-state models using a quasi-maximum-likelihood estimator. Such agents too will asymptotically forecast *as if* they relied on pseudo-true d-state models. See, for instance, Theorem 2 of Douc and Moulines (2012).

knowledge of the true process, one arrives at the set of pseudo-true models simply by following Bayes' rule—no knowledge of the true process is necessary.[11] Following Bayes' rule would have led agents to the truth had their model been correctly specified, but it can lead them astray in the presence of model misspecification.

3 Pseudo-True Subjective Beliefs

In this section, I characterize the subjective beliefs of agents who use pseudo-true d-state models. As a preliminary step, I establish a useful invariance property for the class of pseudo-true models, which is of independent interest.

3.1 Linear Invariance

There are no constraints on the agents' set of models other than the bound on the number of subjective state variables. Formally, matrices A, B, Q, and R of representation (1) are unrestricted, other than the minimal restrictions required for (1) to define a proper stationary ergodic stochastic process.[12] This flexibility in the agents' set of models enables them to capture any linear intratemporal relationship among observables by the appropriate choice of matrices A, B, Q, and R. It thus results in a crucial linear-invariance property for pseudo-true d-state models.

Theorem 2 (linear invariance)**.** *Let $\tilde{y}_t = T y_t$ denote a linear transformation of y_t, and let $\widetilde{\mathbb{P}}$ denote the probability distribution over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$ induced by $\mathbb{P}$ and T. Let $\mathcal{P}_d^*$ denote the set of pseudo-true d-state stationary distributions when $\mathbb{P}$ is the true process, and let $\widetilde{\mathcal{P}}_d^*$ denote the corresponding set when $\widetilde{\mathbb{P}}$ is the true process. If T is a full-rank matrix, then the set of probability distributions over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$ induced by $\mathcal{P}_d^*$ and T coincides with $\widetilde{\mathcal{P}}_d^*$.*[13]

The theorem establishes that the framing of agents' observations does not affect how they form their expectations. Agents' models and forecasts only depend on the amount of information available to them, not how it is presented. For instance, whether agents observe the nominal interest rate and the inflation rate or the real interest rate and the inflation rate is immaterial for how they form their expectations. Likewise, agents' expectations are not affected by augmenting the vector of observables with linear combinations of variables already in the agents' information set.

[11] An analogy with the ordinary least squares (OLS) estimation with omitted variables is instructive. Consider an agent who can only entertain models of the form $y_i = \beta x_i + \varepsilon_i$, with ε_i i.i.d. and normally distributed, and is interested in the causal effect of x on y. In the presence of omitted variables, the OLS estimate (which coincides with the pseudo-true model) will generally be different from the linear-Gaussian model that maximizes the agent's payoff, given her preference for estimating the true causal effect. Nevertheless, finding the payoff-maximizing model requires the agent to know the joint distribution of the independent, dependent, and omitted variables—an impossible ask.

[12] In particular, the eigenvalues of A need to be smaller than one in magnitude and matrices Q and R need to be positive semidefinite.

[13] The distribution induced over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$ by the distribution P over $\{y_t\}_{t=-\infty}^{\infty}$ and the mapping $T : y_t \mapsto \tilde{y}_t$ is the pushforward distribution $P \circ T^{-1}$ defined as $P \circ T^{-1}\left(\{\tilde{y}_t\}_{t=-\infty}^{\infty} \in Y\right) \equiv P\left(\{y_t\}_{t=-\infty}^{\infty} : \{T y_t\}_{t=-\infty}^{\infty} \in Y\right)$ for any set $Y \subseteq \mathbb{R}^{n\mathbb{Z}}$. The set of distributions over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$ induced by the set of distributions $\mathcal{P}$ over $\{y_t\}_{t=-\infty}^{\infty}$ and the mapping $T : y_t \mapsto \tilde{y}_t$ is given by $\mathcal{P} \circ T^{-1} \equiv \{P \circ T^{-1} : P \in \mathcal{P}\}$.

The dichotomy in the agents' cognitive abilities—capable of observing all the relevant variables and uncovering all linear *intratemporal* relationships among them, yet thoroughly constrained in the complexity of *intertemporal* relationships they can entertain—is arguably stark. Nevertheless, it highlights the paper's premise that forecasting is challenging because it requires forecasters to recognize stochastic patterns that unfold over time. This dichotomy also allows me to abstract away from the cognitive costs of acquiring information and the mistakes individuals make when information is presented differently. Furthermore, it makes the framework immediately portable across different applications, thanks to the linear-invariance result.[14]

The result also showcases the endogeneity of agents' expectations. Since the parameters of the agents' model are determined endogenously by maximizing the fit to the true distribution, they covary with the true distribution. This feature of pseudo-true d-state models, which rational-expectations models share, makes the framework particularly suited to counterfactual analysis in macroeconomics, where policy changes can result in changes in the distribution of payoff-relevant variables.

The linear invariance result allows me to focus on non-degenerate processes. Define the lag-l autocovariance matrix of the observable under the true process as follows:

$$\Gamma_l \equiv \mathbb{E}[y_t y_{t-l}']. \tag{2}$$

The true process is degenerate if Γ_0 is singular. Whenever Γ_0 is singular, there is some lower-dimensional vector $\tilde{y}_t$ with $\mathbb{E}[\tilde{y}_t \tilde{y}_t']$ non-singular and some full-rank matrix T such that $y_t = T\tilde{y}_t$. By Theorem 2, the set of pseudo-true models when the observable is given by y_t can be found by first finding the set of pseudo-true models given $\tilde{y}_t$ and then transforming that set by T. Restricting attention to non-singular true processes allows me to restrict agents to the set of models under which the subjective variance-covariance matrix of the vector of observables is non-singular.[15] In the remainder of the paper, I assume that the variance-covariance matrix Γ_0 is non-singular, and agents can only entertain subjective models with non-singular variance-covariance matrices.

3.2 The One-Dimensional Case

I start the analysis of the agents' pseudo-true models by considering the case where agents can only entertain one-dimensional models. In this case, a complete characterization of the agents' forecasts is possible. The insights from the single-state case generalize to the d-state case, as I discuss later in this section.

[14] Rabin (2013) calls for the use of portable extensions of existing models in behavioral economics, and economic theory more generally. The current framework can be seen as a portable extension of the rational-expectations benchmark, which spans, by varying a single parameter, d, the range between full rationality and a severe form of serial-correlation misperception where agents perceive serially-correlated variables as independent over time.

[15] Whenever the true variance-covariance matrix Γ_0 is non-singular, any subjective model with a singular variance-covariance matrix is dominated in terms of fit to the true process by every subjective model with a non-singular variance-covariance matrix. Therefore, no subjective model with a singular variance-covariance matrix can be a pseudo-true model.

The agents' pseudo-true 1-state forecasts depend on the true process only through the autocorrelations of the vector of observables. Define the lag-l *autocorrelation matrix* of the observable under the true process as follows:

$$C_l \equiv \frac{1}{2}\Gamma_0^{\frac{-1}{2}}\left(\Gamma_l + \Gamma_l'\right)\Gamma_0^{\frac{-1}{2}}. \tag{3}$$

The notion of autocorrelation matrices is a natural generalization of the notion of autocorrelation functions. When the observable y_t is a scalar, C_l reduces to the usual autocorrelation function at lag l. When the observable is an n-dimensional vector, on the other hand, C_l is an $n \times n$ real symmetric matrix with eigenvalues inside the unit circle.[16]

Autocorrelation matrices capture the extent of serial correlation in the vector of observables. Let $\rho(C_1)$ denote the spectral radius of matrix C_1.[17] When $\rho(C_l)$ is close to zero for all l, the process is close to being i.i.d., whereas when $\rho(C_1)$ is close to one for small l, then the process is close to being unit root.

With the definition of autocorrelation matrices at hand, I can now state the general characterization result for the $d = 1$ case:

Theorem 3. *Under any pseudo-true* 1*-state model, the agents' s-period ahead forecast is given by*

$$E_t^{1*}[y_{t+s}] = a^{*s}(1-\eta^*)q^*p^{*\prime}\sum_{\tau=0}^{\infty} a^{*\tau}\eta^{*\tau}y_{t-\tau}, \tag{4}$$

where a^* and η^* are scalars in the $[-1, 1]$ and $[0, 1]$ intervals, respectively, that maximize $\lambda_{max}(\Omega(a, \eta))$, the largest eigenvalue of the $n \times n$ real symmetric matrix

$$\Omega(a,\eta) \equiv -\frac{a^2(1-\eta)^2}{1-a^2\eta^2}I + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\sum_{\tau=1}^{\infty} a^\tau \eta^{\tau-1}C_\tau,$$

and $p^* = \Gamma_0^{\frac{-1}{2}}u$ and $q^* = \Gamma_0^{\frac{1}{2}}u$, where u is an eigenvector of $\Omega(a^*, \eta^*)$ with eigenvalue $\lambda_{max}(\Omega(a^*, \eta^*))$, which is normalized so that $u'u = 1$.

The endogenous variables a^*, η^*, p^*, and q^* have intuitive meanings. The scalar a^* captures the perceived *persistence* of the vector of observables. When $a^* = 0$, agents perceive y_t as i.i.d. Whereas when $a^* = 1$, they believe that the observable follows a unit-root process. The scalar η^* captures the perceived *noise* in the agents' observations. When η^* is small, agents believe recent observations to be highly informative of the value of the subjective state. As a result, their expectations respond more to recent observations and discount old observations more. The vector p^* determines the agents' *relative attention* to different components of the vector of observables. When p_i^* is larger than p_j^*, agents put more weight on $y_{i,t-\tau}$ relative to $y_{j,t-\tau}$ for all τ when forming their estimate of the subjective state. Finally, the vector q^* captures the *relative sensitivity* of the agents' forecasts of different observables to changes in their estimate of the subjective state.

[16] See Lemma A.2 of the appendix for a proof.

[17] The spectral radius $\rho(X)$ of matrix X denotes the maximum among the magnitudes of eigenvalues of X.

When q_i^* is larger than q_j^*, then a change in the estimated value of the state at time t leads agents to change their forecast of $y_{i,t+s}$ by more than their forecast of $y_{j,t+s}$ for all s.

A few remarks about the characterization result are in order. First, Theorem 3 does not rule out the possibility that $|a^*| = 1$ and $\eta^* > 0$, in which case the pseudo-true 1-state model would not be stationary ergodic. However, the following result establishes that any pseudo-true 1-state model inherits the stationarity and ergodicity of the true process:

Theorem 4. *Let P^{1*} denote a pseudo-true 1-state model given true distribution $\mathbb{P}$. If $\mathbb{P}$ is stationary ergodic, then so is P^{1*}.*

Second, the theorem significantly reduces the computational complexity of finding the set of pseudo-true models. The set of all d-state models is a manifold of dimension $2nd$, which is never compact and does not admit a global parameterization for any $n > 1$—even if $d = 1$.[18] Additionally, the KLDR is a non-convex function of $\theta = (A, B, Q, R)$. Theorem 3 analytically concentrates out all but two of the parameters of agents' models, thus reducing an optimization problem over a $2n$-dimensional non-compact manifold to a problem over a two-dimensional compact square.

Third, although much easier than the problem of KLDR minimization over the space of d-state models, the problem of maximizing $\lambda_{\max}(\Omega(a, \eta))$ over (a, η) is still non-convex. Consequently, solving it requires the use of numerical global optimization methods. However, the problem can be solved efficiently in any application, regardless of the dimension of the vector of observables.

The non-convexity of the problem also makes an analytical solution elusive without further assumptions on the true process. I thus proceed by imposing an additional assumption on the true process $\mathbb{P}$, which permits a closed-form characterization of the set of pseudo-true 1-state models.

3.3 Exponential Ergodicity and Incomplete Information

The optimization problem in Theorem 3 has an intuitive closed-form solution given a class of true stochastic processes that arise naturally in many applications including those studied in Sections 5–7. The appropriate class turns out to be the following:

Definition 1. The stationary ergodic process $\mathbb{P}$ is *exponentially ergodic* if $\rho(C_l) \leq \rho(C_1)^l$ for all $l \geq 1$, where $\rho(C_l)$ denotes the spectral radius of the autocorrelation matrix C_l.[19]

Exponential ergodicity is stronger than ergodicity. While ergodicity requires the serial correlation at lag l to decay to zero as $l \to \infty$, exponential ergodicity requires the rate of decay to be faster than $\rho(C_1)$. However, many standard processes are exponentially ergodic. For instance,

[18] Although the set of d-state models can be parameterized using matrices A, B, Q, and R, these matrices are in general not identified: For any model θ, there exist a continuum of other models $\tilde{\theta}$ such that $P^\theta = P^{\tilde{\theta}}$. See Gevers and Wertz (1984) for more on identification and parameterization of state-space models.

[19] The term "exponential ergodicity" has been used to refer to a property of Markov chains, where the effect of initial condition on the current distribution of the state decays exponentially fast—see, for instance Meyn and Tweedie (1993). The definition used in this paper is mathematically distinct from the one in the context of Markov chains, but it captures the analogous idea that the serial correlation in the variables decays exponentially fast.

the vector of observables follows an exponentially ergodic process if it is a linear combination of n independent AR(1) shocks.

The following result characterizes the agents' pseudo-true 1-state forecasts when the true process is exponentially ergodic. It links the agents' forecasts to the eigenvalues and eigenvectors of the autocorrelation matrix at lag one:

Theorem 5. *If the true process is exponentially ergodic, then the s-period ahead forecast of agents who use pseudo-true 1-state models is given by*

$$E_t^{1*}[y_{t+s}] = a^{*s} q^* p^{*\prime} y_t, \tag{5}$$

where a^* is an eigenvalue of C_1 largest in magnitude, u denotes the corresponding eigenvector normalized so that $u'u = 1$, and $p^* = \Gamma_0^{\frac{-1}{2}} u$ and $q^* = \Gamma_0^{\frac{1}{2}} u$.

Agents forecast *as if* there is a single state with persistence a^* driving all the elements of the vector of observables. Suppose there is a change in the value of the observable. Agents incorporate this information by first projecting the change in the observable on vector p^* to form an updated estimate $E_t^{1*}[z_t]$ of the subjective state. Agents thus dismiss as irrelevant any change in the vector of observables orthogonal to the relative attention vector p^*. They then forecast the change in the s-period ahead value of the subjective state $E_t^{1*}[z_{t+s}]$ under the assumption that the state has persistence a^*. Finally, they multiply their estimate of the subjective state by the relative sensitivity vector q^* to form their forecast of the observable in period $t + s$.

The following example illustrates the result in the context of a commonly-used specification for the true process:

Example 1. Suppose the true process $\mathbb{P}$ has the following representation:

$$\begin{aligned} f_t &= F f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma), \\ y_t &= H' f_t, \end{aligned}$$

where

$$F = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix},$$

$H \in \mathbb{R}^{n \times n}$ is an invertible square matrix, and $1 > |\alpha_1| > |\alpha_2| > \cdots > |\alpha_n| > 0$. It is easy to verify that $\rho(C_l) = |\alpha_1|^l = \rho(C_1)^l$. That is, the true process is exponentially ergodic, and I can use Theorem 5 to characterize the agents' pseudo-true 1-state model and their forecasts. The perceived persistence, perceived noise, relative attention, and relative sensitivity are, respectively, given by

$a^* = \alpha_1$, $\eta^* = 0$, $p^* = (H'VH)^{\frac{1}{2}}H^{-1}V^{-1}e_1$, and $q^* = (H'VH)^{\frac{-1}{2}}H'Ve_1$, where $V \equiv (I - F^2)^{-1}\Sigma$ is the variance-covariance matrix of f_t and e_1 denotes the first coordinate vector.[20]

Agents' forecasts take a particularly simple form if H is the identity matrix, and so, $y_{it} = f_{it}$ for $i = 1, \ldots, n$. Then p^* and q^* are both multiples of the first coordinate vector e_1, and agents' forecasts simplify to

$$\begin{aligned} E_t^{1*}[y_{1,t+s}] &= \alpha_1^s y_{1t} = \mathbb{E}_t[y_{1,t+s}], \\ E_t^{1*}[y_{1,t+s}] &= 0 \qquad \forall i \neq 1. \end{aligned}$$

That is, the agents' forecast of the most persistent element of the vector of observables coincides with its rational-expectations counterpart. But agents forecast every other element of the vector of observables as if it were i.i.d.

The example illustrates that agents exhibit a form of *persistence bias*. They forecast the most persistent component of the vector of observables as accurately as under rational expectations but do so at the expense of missing the dynamics of other observables. The intuition for the result is easiest to see when the most persistent true state is close to being unit root. Then doing a poor job of tracking the most persistent state would lead to persistent mistakes in agents' forecasts. The persistence of those mistakes would make them costly from the point of view of KLDR minimization. Therefore, any pseudo-true model tracks the state close to unit root as best possible, even if doing so results in errors in forecasting the other states.

The absence of perceived noise is the other important feature of pseudo-true 1-state models when the true process is exponentially ergodic. It leads agents' forecasts to adjust rapidly in light of new information. It also leads to path-independent forecasts: The time-t forecasts depend on the agents' observation history only through the value of the observable at time t.

One can generalize Example 1 by relaxing the assumption that matrices $\mathbb{F}$ and Σ are diagonal and allowing for non-Gaussian innovations. The following theorem provides a set of sufficient conditions for the process to be exponentially ergodic:

Theorem 6. *Consider a true process $\mathbb{P}$ that can be represented as*

$$\begin{aligned} f_t &= Ff_{t-1} + \epsilon_t \\ y_t &= H'f_t, \end{aligned} \tag{6}$$

where $f_t \in \mathbb{R}^m$, $\epsilon_t \sim \mathcal{WN}(0, \Sigma)$, $F \in \mathbb{R}^{m\times m}$ is a convergent matrix, and $H \in \mathbb{R}^{m\times n}$. Suppose the variance-covariance of f_t is normalized to be the identity matrix. If H is a rank-m matrix and $\left\|\frac{F+F'}{2}\right\|_2 = \|F\|_2$, where $\|\cdot\|_2$ denotes the spectral norm, then the process is exponentially ergodic.

The assumption that the process has a representation of the form (6) is almost without loss of generality. By the Wold representation theorem, any mean zero, covariance stationary, and

[20] See the proof of Lemma A.3 for a derivation.

purely non-deterministic process can by approximated arbitrarily well by a process with a representation of the form (6). That the variance-covariance of f_t is identity is without loss of generality as well. It can always be arranged to hold by an appropriate normalization of f_t.[21]

The assumption on matrix F rules out a severe form of defectiveness by guaranteeing that the largest eigenvalue of the symmetric part of F coincides with the largest singular value of F. It is satisfied if F is diagonal or symmetric, for example. However, the assumption required for the result is much weaker than symmetry.

The most consequential assumption of the theorem is the requirement that H is a rank-m matrix. This assumption requires the vector of observables, y_t, to be rich enough to fully reveal the vector, f_t, of latent factors. The assumption can thus be seen as a *complete information* assumption: If agents observe an observable of the form (6) with a full-rank matrix H, then they have enough information to forecast the observable as well as in the full-information rational-expectations benchmark—even if they fail to do so due to their misspecified models. The following proposition shows that this assumption, in general, cannot be dispensed with:

Proposition 1. *Consider a true process $\mathbb{P}$ that can be represented as in (6) for some $f_t \in \mathbb{R}^m$, $\epsilon_t \sim \mathcal{N}(0, \Sigma)$, diagonal divergent matrix $F \in \mathbb{R}^{m\times m}$, diagonal matrix $\Sigma \in \mathbb{R}^{m\times m}$, and matrix $H \in \mathbb{R}^{m\times n}$. If the representation in (6) is minimal and $m > n = 1$, then the s-period ahead forecast of agents who use pseudo-true 1-state models is given by*

$$E_t^{1*}[y_{t+s}] = a^{*s}(1-\eta^*)\sum_{\tau=0}^{\infty} a^{*\tau}\eta^{*\tau} y_{t-\tau}$$

for some $a^* \in (-1, 1)$ and $\eta^* \in (0, 1)$.

The class of exponentially ergodic processes constitutes a subset of the class of stationary ergodic processes. However, the true processes that arise in the applications studied in this paper are all exponentially ergodic. In those applications, the forecasts of agents who are constrained to use 1-state models take the simple form given by Theorem 5.

3.4 Pseudo-True d-Dimensional Models

I next investigate whether and how the results from the $d = 1$ case generalize to the $d > 1$ case. The forecasts of agents who use d-state models take a form similar to equation (4): Given a d-state model $\theta = (A, B, Q, R)$, agents' s-period ahead forecast is given by

$$E_t^{\theta}[y_{t+s}] = B'A^{s-1}\sum_{\tau=0}^{\infty}(A - KB')^{\tau} K y_{t-\tau}, \tag{7}$$

where $K \in \mathbb{R}^{d\times n}$ is the Kalman gain matrix, which depends on (A, B, Q, R).[22] Equation (7) is valid for any d-state model θ, not just the pseudo-true ones.

[21] See Lemma A.3 of the appendix and its proof for how this can be done.
[22] See equations (A.2) and (A.3) in the proof of Theorem 2 for the definition of K.

To characterize the forecasts under pseudo-true models, one needs to find the (A, B, Q, R) matrices that minimize the KLDR from the true process and the implied Kalman gain K. This is a hard problem that involves minimizing a non-convex function over the $2nd$-dimensional non-compact manifold of d-state models Θ_d. While the problem can be simplified further, finding an analytical solution to the problem seems an unlikely possibility even for $n = 1$.

I instead solve the easier problem of minimizing the KLDR over a subset $\overline{\Theta}_d$ of Θ_d. I say model $\theta = (A, B, Q, R)$ is *Markovian in observables* (m.i.o.) if $A - KB' = \mathbf{0}$, where K is the Kalman gain matrix corresponding to model θ and $\mathbf{0} \in \mathbb{R}^{d\times d}$ is the matrix of zeros. I let

$$\overline{\Theta}_d \equiv \{\theta = (A, B, Q, R) \in \Theta_d : A - KB' = \mathbf{0}\} \subset \Theta_d.$$

denote the set of m.i.o. d-state models. The time-t forecasts of an agent using a model $\theta \in \overline{\Theta}_d$ only depend on the realized value of the vector of observables at time t (and not its past realizations)—hence, the name Markovian in observables. Any model $\theta \in \overline{\Theta}_d$ also features rapid adjustment and path independence in agents' forecasts.

Model $\theta^* \in \overline{\Theta}_d$ is a pseudo-true m.i.o. d-state model if $\text{KLDR}(\theta^*) \leq \text{KLDR}(\theta)$ for all $\theta \in \overline{\Theta}_d$. I let $\overline{\Theta}^*_d$ denote the set of pseudo-true m.i.o. d-state models. The models in $\overline{\Theta}^*_d$ have several appealing theoretical properties. They satisfy a version of the linear-invariance result of Theorem 2, and they have similar Bayesian and quasi-maximum-likelihood learning foundations as (general) pseudo-true d-state models. Perhaps most importantly, the following corollary of Theorem 5 shows that constraining agents to m.i.o. models is without loss under some conditions:

Corollary 1. *If the true process is exponentially ergodic, then any pseudo-true* 1*-state model is m.i.o.*

The result shows that—at least in the one-dimensional case—the set of pseudo-true models is a subset of the set of m.i.o. models when the true process is exponentially ergodic. Whether this result continues to hold for d-state models with $d > 1$ remains an open question. However, I can still make progress by taking the restriction to m.i.o. models as an assumption and characterizing the set of pseudo-true m.i.o. d-state models and the corresponding forecasts:

Theorem 7. *Suppose the lag-*1 *autocovariance matrix,* Γ_1*, is symmetric. Then the* s*-period ahead forecast of agents who use pseudo-true m.i.o.* d*-state models is given by*

$$E_t^{d*}[y_{t+s}] = \sum_{i=1}^{d} a_i^{*s} q_i^* p_i^{*\prime} y_t, \tag{8}$$

where $a_1^*, \ldots, a_d^*$ *are the* d *eigenvalues of* C_1 *largest in magnitude (with the possibility that some of the* a_i^**'s are equal),* u_i *denotes an eigenvector corresponding to* a_i^* *normalized such that* $u_i'u_i = 1$, $p_i^* = \Gamma_0^{\frac{-1}{2}} u_i$, $q_i^* = \Gamma_0^{\frac{1}{2}} u_i$.

The result shows that the insights from the analysis of single-state models, to a large extent, carry over to d-state models. In particular, agents who are restricted to m.i.o. d-state models exhibit a form of persistence bias. They focus on perfectly forecasting the d most persistent elements of the vector of observables at the expense of the other elements. Moreover, the perceived noise in the vector of observables is zero, as was the case in the 1-state case with an exponentially ergodic true process.

The result suggests that state-space models can be estimated consistently by principal component analysis (PCA). This conclusion is reminiscent of a central result in the theory of dynamic factor models on the consistency of the principal components estimator for the common components.[23] However, Theorem 7 is different along several dimensions. First, it concerns state-space models, not dynamic factor models. Second, the estimator suggested by the theorem uses the principal components of the lag-one autocorrelation matrix, while the PCA estimator of dynamic factor models is constructed from the principal components of the variance-covariance matrix. Last but not least, Theorem 7 suggests that the PCA estimator is consistent (at least under the theorem's assumptions) even if the number of states is misspecified.[24] I am aware of no similar result on the consistency of the PCA estimator for dynamic factor models when the number of common factors is misspecified.

3.5 Second Moments

The results of this section so far were concerned with the conditional first moments of pseudo-true d-state models. I end the section by presenting two results on the subjective second moments when agents use pseudo-true d-state models. The first result characterizes the agents' perceived variance-covariance of the vector of observables under pseudo-true 1-state models.

Theorem 8. *Given any pseudo-true* 1*-state model* θ^**, the subjective variance-covariance of the vector of observables,* $\mathrm{Var}^{1*}(y_t)$*, coincides with the true variance-covariance matrix,* Γ_0.

Agents do *not* misperceive the unconditional volatility of the vector of observables. Neither do they misperceive the unconditional means. Instead, it is the *conditional* expectations and volatilities that deviate from the corresponding values under rational expectations. The result is a direct consequence of the assumptions that (i) agents can entertain *any* stationary 1-state model and (ii) fit their models to data by minimizing the KLDR. That agents can entertain any 1-state model allows them always to match the true volatility of the observables by an appropriate choice of matrices (A, B, Q, R). The fact that agents fit their models by minimizing the KLDR (or equivalently by Bayesian updating or maximum-likelihood estimation) means that it is optimal (from the point of view of maximizing fit) to match the volatility of the observable.

I can prove a weaker version of this result for d-state models:

Theorem 9. *Suppose the first autocovariance matrix,* Γ_1*, is symmetric. Then given any pseudo-true*

[23] See, for instance, Stock and Watson (2002).

[24] An estimator for a misspecified model is consistent if the estimate converges to a pseudo-true model almost surely.

m.i.o. d-state model θ^*, the subjective variance-covariance of the vector of observables, $\mathrm{Var}^{d*}(y_t)$, *coincides with the true variance-covariance matrix,* Γ_0.

Agents who are constrained to use m.i.o. d-state models uncover the true variance-covariance matrix of the observable as long as the true process is sufficiently regular. This conclusion is a consequence of the fact that the set of m.i.o. models $\overline{\Theta}_d$ is invariant under linear transformations. Thus, for any Γ_0, the set of m.i.o. models contains a subjective model θ such that $\mathrm{Var}^{\theta}(y_t) = \Gamma_0$. KLDR minimization leads the agents to settle on such a subjective model.

4 Implications for Behavior

I next study how the bias in the agents' forecasts, which results from their use of simple models, affects their decisions in a reduced-form linear framework. I consider a population of identical agents whose time-t best responses take the following form:

$$x_t = b' y_t + E_t\left[\sum_{s=1}^{\infty} \beta^s c' y_{t+s}\right], \tag{9}$$

where $y_t \in \mathbb{R}^n$ is as before the vector of observables, $E_t[\cdot]$ denotes the agents' subjective forecasts, $b, c \in \mathbb{R}^n$ are vectors of parameters capturing agents' preferences, and $\beta \in (0,1)$ is a discount factor.

The best response in equation (9) nests various decisions such as consumption decisions in the permanent income hypothesis or price setting in the new-Keynesian model. In the next section, I further develop the implications of the general framework in the context of three canonical models in macro: the new-Keynesian model, the real business cycle model, and the Diamond–Mortensen–Pissarides model.

Throughout this section, I assume that the true process for $y_t \in \mathbb{R}^n$ has the following representation:

$$\begin{aligned} f_t &= F f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma), \\ y_t &= H' f_t, \end{aligned} \tag{10}$$

where $f_t \in \mathbb{R}^m$ is a set of latent factors, $F = \mathrm{diag}(\alpha_1, \ldots, \alpha_m)$ is a diagonal matrix with $1 > |\alpha_1| > |\alpha_2| > \cdots > |\alpha_m| > 0$ as its diagonals, $H \in \mathbb{R}^{m \times n}$ is a rank-m matrix, and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$ is a diagonal positive definite matrix. There are m "shocks," $f_1, \ldots, f_m$, which follow independent AR(1) processes, such that observables are linear functions of the shocks. The assumption that the economy is driven by independent AR(1) shocks does not play a substantial role in my analysis, but it allows me to study impulse response functions in a way that is consistent with the literature.

I can use the characterization results of Section 3 to find the agents' pseudo-true forecasts when they are constrained to use m.i.o. d-state models.[25] Agents' actions are then found by substituting for their subjective forecasts in equation (9):

[25] By Theorems 1, 5, and 6, the restriction to m.i.o. models is without loss, at the very least, when $d = 1$ and $d \geq m$.

Proposition 2. *Suppose agents' best responses are given by* (9)*, and the true process is as in* (10)*. The actions of agents who have rational expectations and those who use pseudo-true m.i.o. d-state models with $d < m$ are given, respectively, by*

$$x_t^{RE} = \left(b + \sum_{i=1}^{m} \frac{\alpha_i \beta}{1-\alpha_i \beta} H^{\dagger} e_i e_i' H c\right)' y_t,$$
$$x_t^{d*} = \left(b + \sum_{i=1}^{d} \frac{\alpha_i \beta}{1-\alpha_i \beta} H^{\dagger} e_i e_i' H c\right)' y_t,$$

where $H^{\dagger} \in \mathbb{R}^{n\times m}$ denotes the Moore–Penrose pseudo-inverse of H, and $e_i \in \mathbb{R}^m$ is the ith standard coordinate vector.

Agents' time-t optimal actions are linear functions of the time-t value of the vector of observables. This simple observation has an important consequence: By the linear invariance result, agents' pseudo-true models, forecasts, and actions are all the same whether or not they observe the actions of other agents. Moreover, each agent's expectations of other agents' actions are consistent with the agent's expectations of the observable and other agents' best-response functions.

The proposition indicates that agents' actions take similar forms whether they have rational expectations or use pseudo-true models; the only difference between the two is in the upper limit of the sum over i. Agents' actions when they forecast using a pseudo-true d-state model can be obtained from their actions under rational expectations by setting $\alpha_i = 0$ for $i > d$. This observation leads to the following equivalence result:

Corollary 2 (observational equivalence)**.** *Suppose agents' best responses can be represented as in* (9)*. When the true process is given by* (10)*, actions of an agent who uses pseudo-true m.i.o. d-state models with $d < m$ are identical to those of an agent with rational expectations in an economy where the true processes is given by* (10) *with $\alpha_{d+1} = \cdots = \alpha_m = 0$.*

Agents constrained to simple models behave as if they ignored the dynamics of all but the most persistent shocks hitting the economy. They treat the less persistent shocks as if they were i.i.d. when forming their expectations. This persistence bias is at the heart of the framework's implications for agents' behavior. In the remainder of this section, I argue that persistence bias dampens the economy's response to shocks, leads to low-dimensional dynamics, and induces additional co-movement between agents' various forward-looking decisions.

4.1 Stickiness and Dampening

Ignoring the dynamics of some shocks makes agents' forecasts and actions "sticky." Changes in the current value of an i.i.d. observable does not have any information about the future value of the observable. Therefore, agents who treat a shock as i.i.d. do not update their expectations in response to changes in the value of the shock. The stickiness of expectations translates into the stickiness of the agents' forward-looking actions.

I use the diagnostic tool of impulse-response functions (IRFs) to characterize the extent of stickiness in agents' actions. The IRF of agents' actions to an innovation in the jth shock is given by the profile $\{dx_{t+\tau}/d\epsilon_{jt}\}_\tau$. For agents who use pseudo-true m.i.o. d-state models with $d < m$ and agents who have rational expectations, the IRFs are given, respectively, by

$$\frac{d}{d\epsilon_{jt}} x_{t+\tau}^{\text{RE}} = b'H'e_j\alpha_j^\tau + \frac{\beta g'H'e_j}{1-\alpha_j\beta_j}\alpha_j^{\tau+1},$$
$$\frac{d}{d\epsilon_{jt}} x_{t+\tau}^{d*} = \underbrace{b'H'e_j\alpha_j^\tau}_{\text{direct response}} + \underbrace{\frac{\beta g'H'e_j}{1-\alpha_j\beta_j}\alpha_j^{\tau+1}\mathbb{1}\{j \le d\}}_{\text{expectational response}},$$

where $\mathbb{1}\{j \le d\}$ is one if $j \le d$ and is zero otherwise.

The IRF of actions can be decomposed into the sum of two terms: the direct response of actions and the response resulting from changes in expectations. For agents who use simple models, the direct response is identical to the direct response under rational expectations. This is a direct result of the fact that agents perfectly observe the realization of the observable at any point in time. The expectational response for constrained agents takes one of two forms. For the d most persistent shocks, the response is again the same as in rational expectations. For the remaining shocks, the expectational response is zero because agents ignore the dynamics of the shock when they form their forecasts.

When the direct and expectational responses have the same sign, the net effect of constraining agents to use simple models is a dampening of their responses to shocks:

Proposition 3. *Suppose agents' best responses are given by* (9)*, and the true process is as in* (10)*. If* $b'H'e_j$ *and* $g'H'e_j\alpha_j$ *have the same sign, then for all* t*,*

$$\left|\frac{d}{d\epsilon_{jt}} x_{t+\tau}^{d*}\right| \le \left|\frac{d}{d\epsilon_{jt}} x_{t+\tau}^{RE}\right|,$$

with the inequality strict if $d < j \le m$*, where* x^{RE} *and* x^{d*} *denote actions of agents with rational expectations and those who use pseudo-true m.i.o.* d*-state models, respectively.*

4.2 Low-Dimensional Dynamics and a Main Shock

The forward-looking actions of agents who are constrained to use simple models have low dimensional dynamics: An econometrician who analyzes those actions will conclude that the actions are driven by a small number of shocks. To make this statement precise, I consider agents who take a number of purely forward-looking actions, with the jth such action taking the following form:

$$x_{jt} = E_t\left[\sum_{s=1}^{\infty} \beta_j^s c_j' y_{t+s}\right], \tag{11}$$

where $c_j \in \mathbb{R}^n$ is a vector of parameters, and $\beta_j \in (0,1)$ is the effective discount factor for the jth decision. Note that equation (11) is simply the best response in (9) with b, the parameter capturing the contemporaneous effect of observables on the action, set to zero. The restriction to purely foreword-looking actions is a matter of convenience; the conclusions will be approximately true if the actions are sufficiently forward-looking (i.e., if $\beta_j \approx 1$ or $b_j \approx 0$ for all j).

Proposition 2 still characterizes the actions of agents who are constrained to use d-state models. But I can express the actions more conveniently in terms of the vector of shocks:

$$x_{jt}^{d*} = \sum_{i=1}^{d} \frac{\alpha_i \beta_j c_j' H' e_i}{1 - \alpha_i \beta_j} f_{it}.$$

It is immediate from this expression that agents' actions only respond to the d most persistent shocks. Consequently, an econometrician who analyzes the dynamics of agents' actions will conclude that the economy is driven by d shocks. This conclusion is independent of the specifics of agents' preferences, technology, or market structure. It holds both in partial equilibrium and in general equilibrium, as the analysis in Subsection 4.4 will make clear.

Angeletos, Collard, and Dellas (2020) find that a "main business cycle chocks" explains the bulk of movements in macroeconomic aggregates at business cycle frequencies. My analysis suggests that this finding should not come as a surprise: There is always a main shock—as long as decisions are sufficiently forward looking and agents use simple models. But in general, the main shock is an endogenous index whose composition depends on the primitives of the economy such as preferences, technology, market structure, the stochastic properties of the shocks that hit the economy, as well as the parameters of policy rules.[26]

4.3 Co-movement

The low-dimensional dynamics of agents' actions also lead to additional co-movement between their different choices. One can measure the extent of co-movement using different metrics, perhaps the most commonly used one being the correlation. The following proposition shows that constraining agents to use one-dimensional models leads to additional co-movement in their choices:

Proposition 4. *Consider actions j and k, both of the form* (11). *Suppose the true process is as in* (10) *with $m > 1$. Then*

$$\left| Corr\left(x_{jt}^{1*}, x_{kt}^{1*}\right) \right| \geq \left| Corr\left(x_{jt}^{RE}, x_{kt}^{RE}\right) \right|,$$

with the inequality strict for generic parameter values.

The result is intuitive in light of what came before. Forward-looking decisions of agents who are constrained to use 1-state models only respond to the most persistent shock in the economy. Therefore, their various actions co-move perfectly with each other, and the correlation between

[26] In the simple economies considered in this section, where there are no endogenous variables, the main shock is identical to the most persistent shock hitting the economy. In the applications considered in the next section, however, the composition of the main shock also depends on the other primitives of the economy.

the actions is generically one in absolute value. Under rational expectations, in contrast, the co-movement between different actions is imperfect (except for knife edge cases).[27]

4.4 Partial Equilibrium vs General Equilibrium

I conclude this section by arguing that the implications of the general framework are largely unchanged in a general equilibrium setting where the laws of motion for observables depend on agents' actions. I consider a stylized general equilibrium (GE) economy in which observables are linear functions of exogenous shocks and agents' actions. Specifically, I assume that, in equilibrium, the vector of observables, y_t, can be written as

$$y_t^{\mathrm{GE}} = \widetilde{H}' f_t + g x_t^{\mathrm{GE}}, \tag{12}$$

where $f_t \in \mathbb{R}^m$ is the vector of shocks, $\widetilde{H} \in \mathbb{R}^{m \times n}$ is a rank-m matrix, and $g \in \mathbb{R}^m$ is a vector that parameterizes the strength of the GE feedback from agents' actions to the observable. The agents' best-response functions are, as before, given by (9):

$$x_t^{\mathrm{GE}} = b' y_t^{\mathrm{GE}} + E_t \left[\sum_{s=1}^{\infty} \beta^s c' y_{t+s}^{\mathrm{GE}} \right]. \tag{13}$$

I continue to assume that shocks follow m independent AR(1) processes:

$$f_t = F f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \Sigma), \tag{14}$$

where $F = \mathrm{diag}(\alpha_1, \ldots, \alpha_m)$ and $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. Equations (12)–(14) together with the specification of agents' subjective expectations fully characterize the (general) equilibrium of the economy.

I contrast this economy with a partial equilibrium (PE) economy in which

$$y_t^{\mathrm{PE}} = H' f_t, \tag{15}$$

$$x_t^{\mathrm{PE}} = b' y_t^{\mathrm{PE}} + E_t \left[\sum_{s=1}^{\infty} \beta^s c' y_{t+s}^{\mathrm{PE}} \right], \tag{16}$$

and f_t follows (14). Note that the PE economy is nothing more than the model I studied so far in this section.

The term "partial equilibrium" is inspired by the following hypothetical scenario: Suppose we considered the economy described by equations (12)–(14) but ignored the fact that agents' actions affect the observable, which in turn affect agents' actions, and so on. Then the response of the GE economy to shocks would be described by equations (15)–(16).

My next result establishes an observational equivalence between the GE and PE economies:

[27] One might guess that a similar result holds when agents are constrained to use d-state models with $1 < d < m$. But this guess turns out to be incorrect: It is always possible to find actions j and k for which $\mathrm{Corr}\left(x_{jt}^{d*}, x_{kt}^{d*}\right) = 0$ and $\mathrm{Corr}\left(x_{jt}^{\mathrm{RE}}, x_{kt}^{\mathrm{RE}}\right) \neq 0$.

Proposition 5 (general equilibrium)**.** *Consider the general equilibrium economy* (12)–(14) *and the partial equilibrium economy* (14)–(16)*, and suppose that, in each economy, agents use pseudo-true m.i.o. d-state models to forecast the observable. If*

$$\widetilde{H} = H\left(I - \left(b + \sum_{k=1}^{d} \frac{\alpha_k \beta}{1-\alpha_k \beta} H^{\dagger} e_k e_k' H c\right) g'\right),$$

then the linear equilibria of the two economies are observationally equivalent.

Several remarks are in order. First, the result is a corollary of the linear invariance result (Theorem 2) and the fact that agents' actions are linear in the observable, established in Proposition 2. Second, the proposition covers the rational-expectations case by setting $d = m$. Third, when $\beta = 0$, the effect of going from PE to GE is to amplify the response of observables to shocks, as measured by matrix H', by the GE multiplier $(I - gb')^{-1}$. When $\beta > 0$, the multiplier has an additional term, which captures the general equilibrium effect of the updating of expectations by agents.

Last but not least, the distinctions between exogenous and endogenous variables, on one hand, and PE and GE, on the other, are largely inconsequential in this framework. Agents' expectations of endogenous variables are consistent with their expectations of exogenous variables and the structural equations of the economy, the GE economy is just the PE economy with a linearly transformed H matrix, and agents' expectations in the GE economy are just linear transformations of their expectations in the PE economy.

However, this conclusion relies on the assumption that the economy does not have any endogenous state variable.[28] This assumption (together with the linearity of the economy) turns all general equilibrium restrictions into linear intratemporal relationships between variables—exactly the kinds of relationships agents can comprehend. When the economy has endogenous variables, on the other hand, constraining agents to simple models could lead to behave differently in GE due to a subtle feedback between the agents' model of the economy and the economy's law of motion. In Sections 6 and 7, I illustrate this point in the context of the real business cycle model and the Diamond–Mortensen–Pissarides model.

5 Application to The New-Keynesian Model

As the first application of the general framework, I study the standard three-equation new-Keynesian model.[29]

[28] I say that a state variable is endogenous if its dynamics are endogenous, i.e., determined in equilibrium. For example, shocks are not endogenous state variables, but the capital stock is.

[29] Technically speaking, the economy will be a two-equation new-Keynesian economy, described by the dynamic IS curve and the Phillips curve. The Taylor rule will play no role in my analysis.

5.1 Primitives

The primitives of the economy are completely standard. Time is discrete, preferences are time separable, and discounting is exponential. There is a measure of household with separable preferences over the final good and leisure. In each period, households decide how much to consume and how much to save in a nominal bond, which is in zero net supply. Households also make labor-supply decisions taking the wage as given. The consumption good is a CES aggregate of a continuum of intermediate goods. Intermediate goods are produced by monopolistically competitive firms using a technology linear in labor. Intermediate-good producers are subject to a Calvo-style pricing friction. Markets for labor, the final good, and the nominal bond are competitive.

The economy is subject to technology shocks that move the natural rate of interest and cost-push shocks that affect the intermediate goods producers' desired markups. The nominal interest rate is set by a central bank. The exact rule followed by the central bank is irrelevant for my analysis. Rather, equilibrium outcomes will depend only on the statistical properties of the interest rate process (such as its serial correlation and its correlation with other aggregate observables).[30]

5.2 Log-linear Temporary Equilibrium

It is well known since Preston (2005) that recursive equilibrium equations that relate aggregate variables (e.g., the aggregate Euler equation) may not be valid away from rational expectations. Instead, one needs to separately characterize each agent's optimal behavior using only relationships that are respected by the agent's expectations.

My analysis of the new-Keynesian model thus proceeds in two steps. The first step is to characterize the *temporary equilibrium* relationships, which impose individual optimality and market clearing conditions but not rational expectations.[31] The second step is to supplement the temporary equilibrium with the model of expectations formation and characterize the resulting (full) equilibrium.

The first step of the analysis is standard. I therefore omit the details of the derivation and use the log-linearized temporary equilibrium relationships as my starting point.[32] These temporary-equilibrium conditions are given by

$$\hat{x}_t = -\sigma\left(\hat{i}_t - r_t^n\right) + E_t^h\left[\sum_{s=1}^{\infty}\beta^s\left(\frac{1-\beta}{\beta}\hat{x}_{t+s} - \sigma\left(\hat{i}_{t+s} - r_{t+s}^n\right) - \frac{\sigma}{\beta}\hat{\pi}_{t+s}\right)\right], \tag{17}$$

[30]The Taylor principle is not necessary for equilibrium uniqueness in my setup. The sunspot equilibria in the new-Keynesian model require agents to observe payoff-irrelevant common "sunspots." I restrict the set of variables that can appear in the vector of observables to be payoff relevant, thus ruling out sunspot equilibria. This is one advantage of explicitly specifying the list of observables on which agents can condition their forecasts.

[31]The idea of temporary equilibrium goes back to the writings of Hicks (1939) and Lindahl (1939). It has been extensively developed in the context of Arrow–Debreu economies by Grandmont (1977, 1982). See Woodford (2013) for a discussion of temporary equilibria in the context of modern monetary models and Farhi and Werning (2017) for an application in the context of a heterogeneous-agent new-Keynesian economy.

[32]Interested readers can find the details of this derivation, among other places, in Angeletos and Lian (2018) and Gáti (2020).

$$\hat{\pi}_t = \kappa \hat{x}_t + \mu_t + E_t^f \left[\sum_{s=1}^{\infty} (\beta\delta)^s \left(\kappa \hat{x}_{t+s} + \frac{1-\delta}{\delta} \hat{\pi}_{t+s} + \mu_{t+s} \right) \right], \tag{18}$$

where $\hat{x}_t$, $\hat{i}_t$, and $\hat{\pi}_t$ denote the log-deviations of output gap, (gross) nominal interest rate, and inflation rate, respectively, from their steady state values, β is the discount factor, σ is the elasticity of intertemporal substitution (EIS), δ is the Calvo parameter, κ is a composite parameter that determines the steepness of the Phillips curve, r_t^n denotes the technology shock that moves the natural rate of interest, and μ_t is the cost-push shock. E_t^h and E_t^f denote subjective expectations of households and firms, respectively.

I assume that the vector $(\hat{i}_t, r_t^n, \mu_t)'$, of nominal interest rate, technology shock, and cost-push shock, follows a mean-zero stationary and exponentially ergodic process.[33] This assumption allows me to use Theorem 5 to characterize the set of pseudo-true 1-state models.

5.3 Subjective Expectations and Equilibrium

For simplicity, I assume that households and firms face identical constraints on the models they can entertain, and so, end up with identical subjective expectations. Every agent knows the steady-state values of every variable. The agents' time-t information set is given by the history $\{y_\tau\}_{\tau \le t}$ of vector $y_\tau \equiv (\hat{x}_\tau, \hat{\pi}_\tau, \hat{i}_\tau, r_\tau^n, \mu_\tau)'$, consisting of time-$\tau$ log-deviations of output, inflation, and interest rate from their steady-state values, as well as realizations of every shock. Instead of imposing rational expectations, I assume agents are constrained to use one-dimensional state-space models of the form (1) to forecast y.

The equilibrium definition is straightforward. An equilibrium consists of a stochastic process $\mathbb{P}$ for $\{y_t\}_t$ and a model θ^* for agents such that (i) $\mathbb{P}$ is derived from market clearing conditions and optimal behavior by households and firms given subjective model θ^*, and (ii) θ^* is a pseudo-true 1-state model given the stochastic process $\mathbb{P}$. Following my earlier work (Molavi, 2019), I refer to this equilibrium notion as *constrained rational expectations equilibrium*.

Finding an equilibrium involves solving a fixed-point equation. I can do this in the context of the new-Keynesian model using pen and paper via a guess-and-verify method. I focus on linear equilibria, in which $\hat{x}_t$ and $\hat{\pi}_t$ are linear functions of $\hat{i}_t$, r_t^n, and μ_t.[34] In such an equilibrium, the y_t vector contains two redundant elements (which are linear combinations of other elements of y_t). Therefore, agents' forecasts of y can be obtained by first finding their forecasts of some three-

[33] The new-Keynesian literature often assumes that nominal interest rate follows a Taylor rule, which sets the rate as a linear function of output gap and inflation rate plus a monetary policy shock. As long as shocks follow a stationary and exponentially ergodic process, the standard specification leads to a process for $(\hat{i}_t, r_t^n, \mu_t)'$ that is stationary and exponentially ergodic—both in the rational expectations equilibrium and in the equilibrium in which agents are constrained to use simple state-space models. My reduced-form specification of the interest rate process is thus observationally equivalent to the standard specification. But the reduced-form specification has the advantage of allowing the model to be calibrated without taking a stand on which part of changes in interest rate are systematic and which parts are due to pure monetary policy shocks. It also enables me to study the effects of forward guidance in a theoretically coherent way. These pluses come at the expense of precluding counterfactual analyses with respect to the parameters of the Taylor rule.

[34] The existence and generic uniqueness of a linear equilibrium follows from the guess-and-verify argument. My method for finding an equilibrium is silent on whether there are other, non-linear equilibria.

dimensional vector f that spans the subspace spanned by y and then using the linear invariance result to find their forecasts of y.

I take $f_t \equiv (\hat{x}_t, \hat{\pi}_t, \hat{i}_t)'$ as my basis for the subspace spanned by y_t. This choice of f_t has two advantages over the more natural choice of the vector of shocks. First, it considerably simplifies the algebra involved in finding the equilibrium. Second, it makes the estimation of the model more straightforward. By Theorem 5, agents' model of any vector f_t depends on the autocovariance matrices of f_t at lags zero and one. When f_t consists of output gap, inflation, and interest rate, those autocovariance matrices have readily available empirical counterparts. Estimating autocovariance matrices of shocks, on the other hand, is a much harder tasks that requires a strategy to identify the unobservable shocks.

The following proposition summarizes the equilibrium characterization:

Proposition 6. *Suppose the shocks in the new-Keynesian model are stationary and exponentially ergodic, agents are constrained to use* 1*-state models, and their time-*t *information set consists of the history of vector* $y_\tau \equiv (\hat{x}_\tau, \hat{\pi}_\tau, \hat{i}_\tau, r^n_\tau, \mu_\tau)'$ *for* $\tau \leq t$*. In the constrained rational expectations equilibrium,*

$$\hat{x}_t = \frac{1}{1 - p_x\gamma_x - p_\pi(\gamma_\pi + \kappa\gamma_x)}\left[\gamma_x(p_i\hat{i}_t + p_\pi\mu_t) - \sigma(1 - \gamma_\pi p_\pi)(\hat{i}_t - r^n_t)\right], \tag{19}$$

$$\hat{\pi}_t = \frac{1}{1 - p_x\gamma_x - p_\pi(\gamma_\pi + \kappa\gamma_x)}\left[(\gamma_\pi + \kappa\gamma_x)p_i\hat{i}_t + (1 - \gamma_x p_x)\mu_t - \sigma(\kappa + \gamma_\pi p_x)(\hat{i}_t - r^n_t)\right], \tag{20}$$

where

$$\gamma_x \equiv a(q_x - \sigma q_\pi), \tag{21}$$

$$\gamma_\pi \equiv a\beta q_\pi, \tag{22}$$

Γ_0 *is the variance-covariance matrix of* $(\hat{x}_t, \hat{\pi}_t, \hat{i}_t)$*,* C_1 *is the corresponding lag-one autocorrelation matrix,* a *is the eigenvalue of* C_1 *largest in magnitude,* u *is the corresponding eigenvector normalized so that* $u'u = 1$*,* $p \equiv (p_x, p_\pi, p_i)' \equiv \Gamma_0^{\frac{-1}{2}}u$*, and* $q \equiv (q_x, q_\pi, q_i)' \equiv \Gamma_0^{\frac{1}{2}}u$*.*

The proposition provides an explicit characterization of the equilibrium given autocovariance matrices of the vector $f_t = (\hat{x}_t, \hat{\pi}_t, \hat{i}_t)'$. Although f_t contains output gap and inflation rate, which are endogenous objects, the characterization is still useful. One can directly measure the autocovariance matrices of f_t in the data and use the measured values together with values for β, σ, δ, and κ to find the response of the economy to interest rate changes as well as technology and cost-push shocks. Furthermore, in equilibrium, there is a one-to-one mapping between autocovariance matrices of f_t and autocovariance matrices of the shocks. Therefore, setting the autocovariance matrices of f_t to their empirical counterparts is equivalent to choosing the shock process to target the empirical autocovariance matrices of f_t.

One can use the result to think about optimal monetary policy. The new-Keynesian model has the so-called "divine coincidence" property under rational expectations: Without cost-push shocks, the central bank faces no trade-off between its dual goals of zero output gap and stable inflation. A similar property holds for some parameter values when agents are constrained to use one-dimensional models: The economy features divine coincidence when $p_i = 0$, i.e., agents put zero weight on the nominal interest rate when forming their forecasts, and when $q_\pi = 0$, i.e., agents' inflation expectations are insensitive to their estimate of the state of the economy. However, these are both knife-edge cases. The following corollary of Proposition 6 establishes that there is generally a trade-off between output and inflation stabilization:

Corollary 3 (failure of divine coincidence)**.** *Suppose the shocks in the new-Keynesian model are stationary and exponentially ergodic, agents are constrained to use* 1*-state models, and their time-t information set consists of the history of vector $y_\tau \equiv (\hat{x}_\tau, \hat{\pi}_\tau, \hat{i}_\tau, r^n_\tau, \mu_\tau)'$ for $\tau \leq t$. For generic values of parameters β, σ, δ, and κ and autocovariance matrices Γ_0 and Γ_1, the central bank cannot simultaneously achieve zero output gap and zero inflation, even when the cost-push shock is zero.*

The central bank is thus limited in what it can achieve with conventional monetary policy, even abstracting from the effective lower bound on the nominal interest rate. I proceed by studying if and how forward guidance can help.

5.4 Forward Guidance

I consider an economy that has been operating without forward guidance for a long time and study how implementing forward guidance then affects output and inflation. This is a good description of where the U.S. economy was in 2009, in the aftermath of the Global Financial Crisis. Consistent with this story, I end my sample in the fourth quarter of 2008 when taking the model to the data.

I assume that agents continue to forecast using a 1-state model that is pseudo true in an equilibrium without forward guidance even as they see forward guidance. This assumption captures the following scenario. Agents have lived in a new-Keynesian economy without forward guidance for a long time and have had ample opportunities to learn the equilibrium relationships. However, since agents can only entertain 1-state models, instead of learning the true model, they have settled on a pseudo-true 1-state model. Agents are then confronted with forward guidance for the first time. The key assumption is that agents do not immediately abandon their model; rather, they continue to rely on the model they had before the switch to the forward-guidance regime, even though the model may not be pseudo-true under the new regime.

The fact that agents have a fully specified model for the stochastic process of y allows me to study the effects of forward guidance in an internally consistent way. I model forward guidance as a credible announcement in period t by the central bank that the nominal rate will follow path $\{\hat{i}_{t+1}, \hat{i}_{t+2}, \ldots, \hat{i}_{t+T}\}$ going forward.

The announcement augments agents' time-t information set to include $\{\hat{i}_{t+1}, \hat{i}_{t+2}, \ldots, \hat{i}_{t+T}\}$ (in addition to $\{y_\tau\}_{\tau \leq t}$). Therefore, the agents' time-t forecasts under forward guidance are the conditional expectations $E_t^{1*}[\cdot] = E_t^{1*}[\cdot|\{y_\tau\}_{\tau\le t}, \hat{i}_{t+1}, \hat{i}_{t+2}, \ldots, \hat{i}_{t+T}]$. But agents' forecasts are Markovian in observables by Theorem 5 and the assumption that the true process is exponentially ergodic. Therefore, $E_t^{1*}[\cdot] = E_t^{1*}[\cdot|y_t, \hat{i}_{t+1}, \hat{i}_{t+2}, \ldots, \hat{i}_{t+T}]$.

On the other hand, since agents use linear-Gaussian state-space models, their forecasts are linear functions of the variables in their information set. In particular, for any observable $\zeta \in \{\hat{x}, \hat{\pi}, \hat{i}, r^n, \mu\}$

$$E_t^{1*}[\zeta_{t+s}] = E^{1*}[\zeta_{t+s}|f_t, \hat{i}_{t+1}, \hat{i}_{t+2}, \ldots, \hat{i}_{t+T}] = \Sigma_{\zeta_s\omega_T}\Sigma_{\omega_T\omega_T}^{-1}\omega_T,$$

where $\omega_T \equiv (\zeta_t, \hat{i}_{t+1}, \ldots, \hat{i}_{t+T})'$, $\Sigma_{\zeta_s\omega_T} \equiv E^{1*}[\zeta_{t+s}\omega_T']$, and $\Sigma_{\omega_T\omega_T} \equiv E^{1*}[\omega_T\omega_T']$. Note that the covariance matrices that show up in the agents' forecasts of ζ are subjective covariance matrices which depend on the agents' subjective model. But the subjective model is just the pseudo-true 1-state model, which is fully characterized by Proposition 6.

The response of the economy to forward guidance takes a relatively simple form. Substituting for the agents' forecasts in (17) and (18) and simplifying the resulting expression, I obtain

$$\hat{x}_t = \alpha_{xi}^{(T)}\hat{i}_t + \alpha_{xn}^{(T)}r_t^n + \alpha_{x\mu}^{(T)}\mu_t + \sum_{s=1}^{T}\alpha_{xi_s}^{(T)}\hat{i}_{t+s},$$

$$\hat{\pi}_t = \alpha_{\pi i}^{(T)}\hat{i}_t + \alpha_{\pi n}^{(T)}r_t^n + \alpha_{\pi\mu}^{(T)}\mu_t + \sum_{s=1}^{T}\alpha_{\pi i_s}^{(T)}\hat{i}_{t+s},$$

where α's are constants that depend on the parameters (a, p, q) of the agents' pseudo-true model and constants β, σ, δ, and κ. The expressions for α's can be found in Appendix C.1.

The α's have intuitive interpretations: α_{xi} and $\alpha_{\pi i}$ are the current interest rate elasticities of output and inflation, respectively, whereas α_{xi_s} and $\alpha_{\pi i_s}$ are the elasticities of output and inflation with respect to the s-period ahead interest rate. Note these elasticities change with the duration T of central bank's guidance. That is, committing to a zero interest rate in period $t+s$ is *not* the same as the central bank not making any announcement about period $t+s$'s interest rate. The (T) superscript in the above expressions is to emphasize this point.

The expressions for α's are rather cumbersome and hard to interpret, so I instead calibrate the model and numerically study the effects of forward guidance.

5.5 Calibration and Estimation

The model has few parameters. I calibrate the model at a quarterly frequency. Following Galí (2015)'s textbook, I set $\beta = 0.99$, $\sigma = 1$, $\delta = 3/4$, and $\kappa = 0.172$. I choose the first two autocovariance matrices of the vector $(\hat{i}_t, r_t^n, \mu_t)'$ of nominal rate, technology shock, and cost-push shock to match the first two autocovariance matrices of $f_t = (\hat{x}_t, \hat{\pi}_t, \hat{i}_t)'$. Since there is a one-to-one mapping between the two sets of autocovariance matrices, I can perfectly match the autocovariance matrices of f_t.

I estimate the empirical autocovariance matrices of f_t using the post war pre Global Financial Crisis U.S. data. For $\hat{x}_t$, I use the percentage difference between real GDP and Potential Output in period t; for $\hat{\pi}_t$, I use the percentage change in GDP Deflator; and for $\hat{i}_t$, I use the Effective Fed Funds Rate. The resulting time series are stationary, so I do not filter them. The sample period is from the first quarter of 1955 to the fourth quarter of 2008.

The estimated (lag-one) autocorrelations of interest rate, technology shock, and cost-push shock are given, respectively, by $\rho_i = 0.954$, $\rho_{r^n} = 0.955$, and $\rho_\mu = 0.925$, whereas the corresponding standard-deviations are given by $\sigma_i = 3.30$, $\sigma_{r^n} = 5.67$, and $\sigma_\mu = 0.315$. However, the estimated shocks are not independent AR(1) processes. See Appendix C.2 for the full estimated autocovariance matrices at lags zero and one, in which I also show that the estimated process is exponentially ergodic.

There are no free parameters for agents' expectations. Agents' models, beliefs, and forecasts are all pinned down by structural parameters β, σ, δ, and κ and the stochastic process of the shocks. The agents' pseudo-true 1-state model, in equilibrium, is described by

$$\begin{aligned}
a^* &= 0.985, \\
p_x^* &= 0.022, \\
p_\pi^* &= -0.42, \\
p_i^* &= -0.014, \\
q_x^* &= 0.53, \\
q_\pi^* &= -2.3, \\
q_i^* &= -2.5,
\end{aligned}$$

where a^* denotes the perceived persistence, p^* is the relative attention vector, and q^* is the relative sensitivity vector.

Agents perceive the subjective state, i.e., "the state of the economy," as highly persistent. Their estimate of the state of the economy is highly sensitive to changes in inflation, but it does not respond much to output or interest rate. High output makes agents optimistic about the state of the economy, while high inflation and high interest rate make them pessimistic. Finally, agents' forecasts of inflation and interest rate move considerably with their estimate of the state of the economy, but not so much for output.

5.6 Results

Figure 1 plots the impulse response functions to an expansionary interest-rate shock. The rate is cut by 100 basis points on impact and follows an AR(1) process with persistence parameter $\rho_i = 0.954$—the estimated (lag-one) autocorrelation of interest rate in the data. The responses of output and inflation rate are both smaller by about a factor of two than their corresponding responses under rational expectations.[35]

[35] See, for instance, Figure 3.1 of Galí (2015). Note, however, that under rational expectations, the shock being considered is typically a "pure monetary policy shock," i.e., a shock to the unsystematic part of the nominal interest rate. Here, in contrast, the shock is a shock to the nominal interest rate itself. So my impulse response functions are not directly comparable to those of Galí (2015).

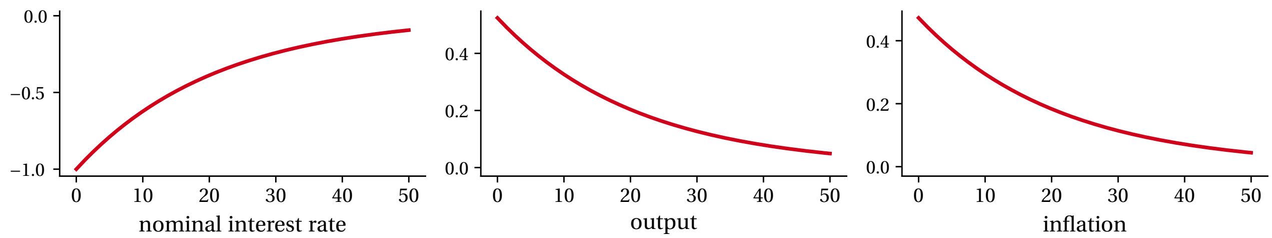

Figure 1. Impulse Response Functions to an Expansionary Interest-Rate Shock

Constraining agents to low-dimensional models makes their expectations sticky. This stickiness dampens the response of their forecasts to changes in the interest rate. Since the economy is almost purely forward looking, the stickiness of agents' forecasts translates into a dampening of the response of agents' actions to the interest rate cut. This is the mechanism behind the dampening of the response of aggregate variables to the shock.

A similar mechanism is at play in reducing the impact of forward guidance on output and inflation. Figure 2 plots the responses of output and inflation to a 100 basis point cut in the current nominal rate combined with an announcement by the central bank that the nominal rate will be kept at −1% for T quarters. The figure plots the response at the time of announcement as the duration of guidance, T, is varied. The response of output to a 100 basis point rate cut accompanied by a promise to keep the rate low for another quarter is almost 50% higher than the response to a rate cut without any guidance.

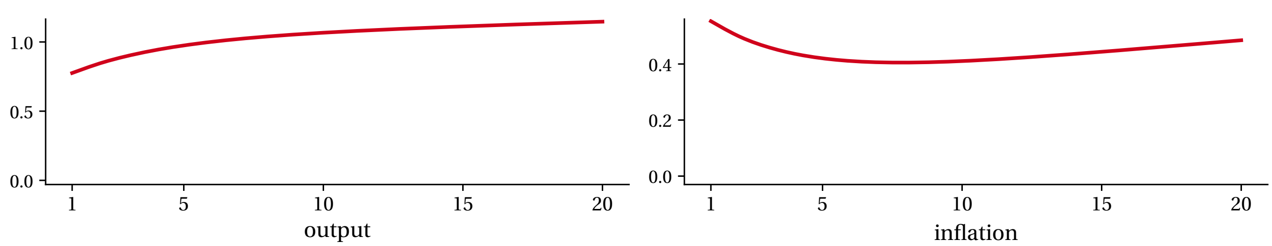

Figure 2. The Power of Forward Guidance

But the central bank quickly runs out of ammunition. Promising to keep the rate low for two quarters (instead of one) increases the response of output only by about 9%. Promising to keep the rate low for 20 quarters is only 50% more stimulative than promising to keep it low for one quarter. Likewise, the response of inflation to forward guidance remains relatively flat as T increases.

The key to understanding Figure 2 is understanding the solution to the filtering problem agents solve to incorporate forward guidance in their forecasts. Agents first use the announced path for the interest rate to estimate the path of the subjective state and then use this estimate to forecast future values of payoff-relevant variables. Therefore, the extent to which forward guidance influences agents' forecasts is a function of the sensitivity of their estimate of the subjective state to changes in the interest rate—p_i^* is exactly this parameter. The fact that p_i^* is small in equilibrium implies that agents' estimate of the subjective state is not too sensitive to changes in expected interest rate. Furthermore, since the subjective state is mean reverting with a perceived persistent $a^* < 1$, information about far future is discounted by agents when they form their expectations. The upshot is that the filtering of information through a one-dimensional model reduces the sensitivity of agents' forecasts to forward guidance.

6 Application to the Real Business Cycle Model

For my second application, I consider the textbook real business cycle (RBC) model.

6.1 Primitives

Preferences, technology, and market structure are standard. Households value consumption and labor according to the per-period utility function

$$u(c,n) = \frac{c^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \psi \frac{n^{1+\varphi}}{1+\varphi},$$

where c denotes consumption, n denotes labor, σ is the elasticity of intertemporal substitution (EIS), φ is the inverse Frisch elasticity of labor supply, and ψ is a constant that determines the steady-state working hours. The consumption good is produced by a measure of competitive firms by combining labor and capital according to the Cobb–Douglas production function

$$o_t = a_t n_t^{\alpha} k_t^{1-\alpha},$$

where o_t denotes output, a_t is total-factor productivity (TFP), and k_t denotes the capital stock at the beginning of period t. TFP follows a first-order autoregressive process in logs: $\hat{a}_t \equiv \log a_t$, and

$$\hat{a}_t = \rho \hat{a}_{t-1} + \epsilon_t. \tag{23}$$

In every period, households choose consumption, labor supply, and the next period's capital stock subject to the following flow budget constraint:

$$k_{t+1} = (1-\delta+r_t)k_t + w_t n_t - c_t,$$

where δ denotes the depreciation rate of capital, r_t is the rental rate of capital, and w_t is the wage rate. Finally, market clearing determines investment:

$$i_t = o_t - c_t.$$

6.2 Log-linear Temporary Equilibrium

As is common in the literature, I log-linearize the model around a steady state in which $\hat{a}_t = 0$, $o_t = o$, $w_t = w$, $r_t = r$, $n_t = n$, $i_t = i$, $k_t = k$, and $c_t = c$. The usual aggregate Euler equation may not hold away from rational expectations. I instead start by characterizing the temporary equilibrium relations, which impose individual optimality and market clearing conditions but not rational expectations. The log-linearized temporary-equilibrium conditions are given by

$$\hat{o}_t = \hat{a}_t + \alpha\hat{k}_t + (1-\alpha)\hat{n}_t, \tag{24}$$

$$\hat{w}_t = \hat{a}_t + \alpha(\hat{k}_t - \hat{n}_t), \tag{25}$$

$$\hat{r}_t = r\hat{a}_t + (1-\alpha)r(\hat{n}_t - \hat{k}_t), \tag{26}$$

$$\hat{n}_t = \frac{1}{\varphi}\hat{w}_t - \frac{1}{\sigma\varphi}\hat{c}_t, \tag{27}$$

$$\hat{i}_t = \frac{o}{i}\hat{o}_t - \frac{c}{i}\hat{c}_t, \tag{28}$$

$$\hat{k}_t = (1-\delta)\hat{k}_{t-1} + \delta\hat{i}_{t-1}, \tag{29}$$

$$\hat{c}_t = \frac{\chi}{\beta}\hat{k}_t + \chi\hat{r}_t + \chi\zeta\hat{w}_t + (\chi - \beta\sigma)\sum_{s=1}^{\infty}\beta^s E_t[\hat{r}_{t+s}] + \chi\zeta\sum_{s=1}^{\infty}\beta^s E_t[\hat{w}_{t+s}], \tag{30}$$

where

$$\chi \equiv (1-\beta)\left(\frac{(1-\alpha)r}{\alpha\sigma\varphi} + \frac{c}{k}\right)^{-1},$$
$$\zeta \equiv \frac{(1-\alpha)(1+\varphi)r}{\alpha\varphi},$$

and $E_t[\cdot]$ denotes the subjective forecast of households. The details of this derivation can be found in Online Appendix D. Equations (23)–(30) fully characterize the equilibrium of the economy once the subjective expectations are specified.

6.3 Subjective Expectations and Equilibrium

I find the equilibrium once imposing rational expectations (RE) and once assuming that households are constrained to d-state models. In both cases, I assume that households know the steady-state values of every variable and perfectly observe the vector $y_t \equiv (\hat{a}_t, \hat{o}_t, \hat{w}_t, \hat{r}_t, \hat{n}_t, \hat{i}_t, \hat{k}_t, \hat{c}_t)$ of log-deviations from the steady state.

Where the two cases differ is in how households model log-deviations from the steady state. Under RE, the households' model of y_t coincides with the observable's true, equilibrium stochastic process. When households are constrained to d-state models, on the other hand, they believe that y_t follows a d-dimensional model of the form (1) and use a pseudo-true d-state model to forecast the future values of y.

The equilibrium definition when households use pseudo-true d-state models is straightforward. An equilibrium consists of a stochastic process $\mathbb{P}$ for y_t and a model θ^* for households such that (i) $\mathbb{P}$ is derived from market clearing conditions and households' optimal consumption, labor supply, and investment behavior given their subjective model θ^*, and (ii) θ^* is a pseudo-true d-state model given the stochastic process $\mathbb{P}$.

The rational-expectations equilibrium has a 2-state representation. Therefore, agents constrained to d-state models with $d > 1$ recover the true process, and the equilibrium given d-state models with $d > 1$ coincides with the rational-expectations equilibrium. This observation highlights the fact that constraining agents to d-state models represents the only deviation from the full-information rational-expectations benchmark. Furthermore, the constraint is slack as long as $d > 1$.

Finding an equilibrium involves solving a fixed-point equation. The rational-expectations equilibrium can be found using existing techniques. In Online Appendix D, I discuss how one can find the equilibrium in the case where households use pseudo-true d-state models with $d = 1$. In the same appendix, I also provide a more formal definition of equilibrium.

6.4 Calibration

The exogenous parameters of the model are calibrated as follows. A period represents a quarter. The quarterly discount rate is set to $\beta = 0.99$. The EIS and the Frisch elasticity of labor supply are both set to one. The depreciation rate is set to $\delta = 0.012$ and the capital share of output to $\alpha = 0.3$. TFP has a persistence parameter of $\rho = 0.95$. I set the standard deviation of TFP innovations to one.

Note that d is the only extra free parameter relative to the full-information rational-expectations version of the model. Once one chooses a value for d, the expectations are fully pinned down by the primitives of the economy (as in the benchmark). Moreover, the d-state equilibrium nests the RE equilibrium by setting $d > 1$.

6.5 Results

Figure 3 plots the impulse response functions to a one percent increase in TFP. The responses for the case where households have rational expectations are in dashed green and for the case where agents use pseudo-true 1-state models are in solid red. Every variable except for the rental rate of capital is measured in log changes from its steady state value; the rental rate of capital is measured in percentage point changes from its steady state value. The variable labeled the "state of the economy" is defined as the households' nowcast, $\hat{z}_t$, of the subjective state, z_t, in their subjective model of the economy. Since the scale of z_t is not identifiable either to the agents or the econometrician, the scale of $\hat{z}_t$ is intrinsically meaningless.

The state of the economy at time t can be expressed as a linear combination of the time-t values of the capital stock and TFP, with the weights determined endogenously in equilibrium:[36]

$$\hat{z}_t = 0.947\hat{k}_t + 0.053\hat{a}_t.$$

The state of the economy is much more sensitive to changes in the capital stock than to changes

[36] As previously mentioned, the magnitude of $\hat{z}_t$ is irrelevant. I normalize $\hat{z}_t$ to have $\hat{z}_t = p_k\hat{k}_t + p_a\hat{a}_t$ with $|p_k| + |p_a| = 1$.

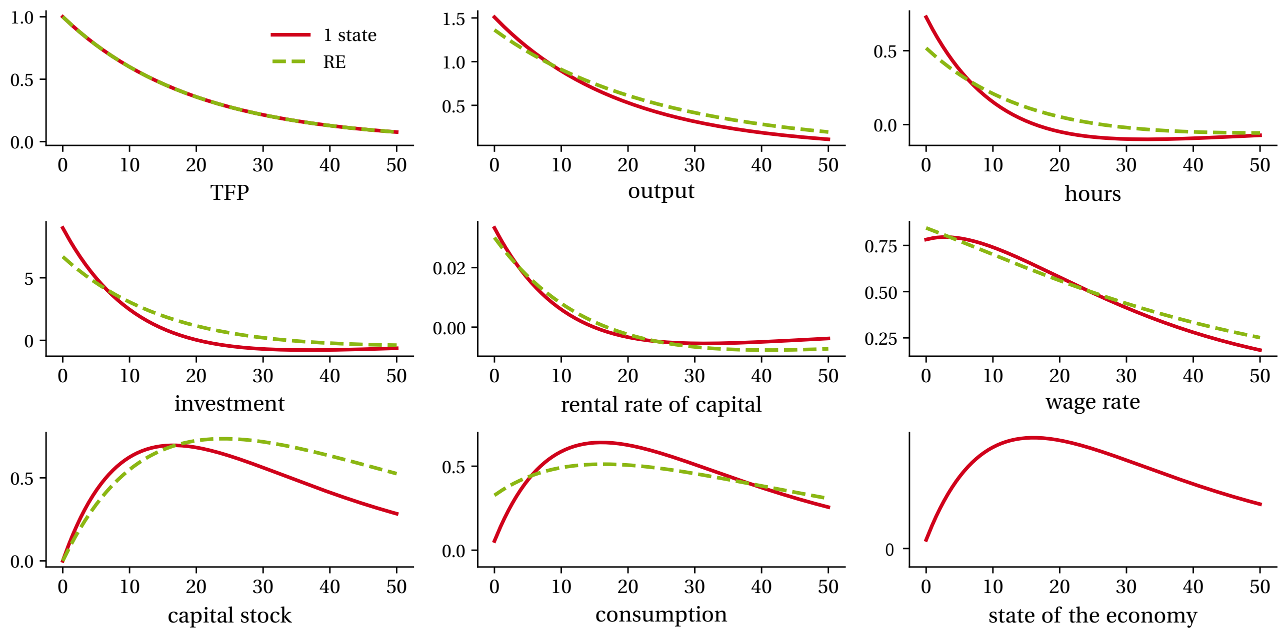

Figure 3. Impulse Response Functions to a TFP Shock

in TFP. This can also be seen in the impulse response functions: The state of the economy co-moves almost perfectly with capital.

The fact that the state of the economy inherits the dynamics of the capital stock is a manifestation of the persistence bias. In equilibrium, the capital stock is more persistent than TFP, as can be seen from the impulse response functions. Therefore, the households' nowcast of the subjective state moves almost one-for-one with changes in the capital stock.

Consumption, in turn, inherits the dynamics of the state of the economy. Since β is close to one in my calibration, consumption is almost purely forward looking. Therefore, it moves in tandem with changes in the households' forecasts, which in turn, move one-for-one with their nowcast, $\hat{z}_t$, of the subjective state. In equilibrium,

$$\hat{c}_t = 0.089\hat{k}_t + 0.088\hat{r}_t + 0.0091\hat{w}_t + 0.841\hat{z}_t.$$

That is, consumption is much more sensitive to changes in the state of the economy than to the current prices and quantities.

The upshot is that consumption co-moves with the capital stock. The (unconditional) correlation of consumption and the capital stock is 0.999 when households use pseudo-true 1-state models; in contrast, it is 0.956 when households have rational expectations. Even though consumption is an almost purely forward-looking variable, it is anchored to the most persistent backward-looking variable in the economy: capital.

The fact that consumption is anchored to capital dampens the initial response of consumption to TFP shocks. The response of consumption on impact when households use a pseudo-true 1-state model is 83% smaller than the corresponding response under RE. The consumption response in the 1-state case continues to be smaller than the RE response for sixth quarters after impact. But as the 1-state economy builds up its capital stock, the households' view of the state of the economy improves and their consumption increases. At some point, consumption in the 1-state economy overshoots its RE counterpart. The model thus provides a parsimonious account of the hump-shaped response of consumption to TFP in empirical studies.[37] Moreover, the initial underreaction and the subsequent overshooting of consumption increases its unconditional volatility relative to the RE benchmark and raises the cost of business cycles.

7 Application to the Diamond–Mortensen–Pissarides Model

I next study how the predictions of the standard labor search and matching model change when agents are constrained to use simple models. I do so in the context of the stochastic version of the Diamond–Mortensen–Pissarides (DMP) model in discrete time. I start by describing the primitives of the economy.

7.1 Primitives

There is a continuum of workers and firms in the economy. The mass of workers is normalized to one, whereas the mass of firms is determined by free entry. Workers and firms are both risk neutral and discount future at rate β. A worker matched with a firm generates a_t units of output in each period, whereas an unemployed worker produces $b < 1$ units. I assume that $a_t - b = (1 - b)\exp(\hat{a}_t)$, where $\hat{a}_t$ is a shock to labor productivity net of home production. This specification of labor productivity guarantees that $a_t > b$ for all t, so it is always efficient for workers to be employed at firms.

Unemployed workers and firms randomly match in a frictional labor market. A matching function determines the rate at which unemployed workers meet firms. Each unemployed worker finds a job in period t with probability $p_t = \mu\theta_t^{1-\alpha}$, and each vacancy is filled with probability $q_t = \mu\theta_t^{-\alpha}$, where $\theta_t \equiv v_t/u_t$ denotes market tightness, i.e., the ratio of the number of vacancies to the unemployment rate, and μ and α are parameters of the matching function. Each job is destroyed in each period with probability $s_t = s\exp(\hat{s}_t)$, where $\hat{s}_t$ is a separation shock. Firms incur a cost k per period (measured in the units of output) for maintaining a vacancy.

Wages are determined through Nash bargaining between a worker and a firm, with the threat point of the worker the value of unemployment, the threat point of the firm the value of an unfilled vacancy (which will be zero in equilibrium), and the worker's bargaining power equal to δ.

I assume that net labor productivity and separation rate shocks follow the autoregressive pro-

[37] For a meta-analysis of the response of aggregate variables to technology shocks, see Ramey (2016, pp. 135–151).

cess

$$\begin{pmatrix} \hat{a}_t \\ \hat{s}_t \end{pmatrix} = \begin{pmatrix} \rho_a & 0 \\ 0 & \rho_s \end{pmatrix} \begin{pmatrix} \hat{a}_{t-1} \\ \hat{s}_{t-1} \end{pmatrix} + \epsilon_t, \tag{31}$$

where $\epsilon_t \sim \mathcal{N}(0, \Sigma)$. This specification allows for labor productivity and separation rate to be correlated, as is the case in the data.

7.2 Temporary Equilibrium

The recursive equations that characterize the solution to the DMP model may not hold away from rational expectations. I instead start by characterizing the temporary-equilibrium relations, which hold under arbitrary expectations. I assume that firms and workers use models with the same number of states, and so, end up with the same subjective expectations in equilibrium. Market tightness and wage then satisfy the following equations:[38]

$$\theta_t^\alpha = \frac{\mu}{k} E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \prod_{k=1}^{\tau-1} (1 - s_{t+k})(a_{t+\tau} - w_{t+\tau}) \right], \tag{32}$$

$$\begin{aligned} w_t = \delta a_t + (1-\delta) b + \delta E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \prod_{k=0}^{\tau-1} (1 - s_{t+k})(a_{t+\tau} - w_{t+\tau}) \right] \\ - (1-\delta) E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \prod_{k=0}^{\tau-1} (1 - s_{t+k} - p_{t+k})(w_{t+\tau} - b) \right]. \end{aligned} \tag{33}$$

The unemployment rate follows the first-order difference equation

$$u_t = u_{t-1} + s_{t-1}(1 - u_{t-1}) - \mu \theta_{t-1}^{1-\alpha} u_{t-1}. \tag{34}$$

Equations (31)–(34) together with the specification of the subjective expectations fully characterize the equilibrium. The derivation of these equations and other omitted calculations from this section can be found in Online Appendix E. To simplify the numerical computations, I log-linearize the temporary equilibrium of the economy around a steady state in which $a_t = 1 > b$ and $s_t = s$.

7.3 Subjective Expectations and Equilibrium

I solve the model once under rational expectations and once assuming that agents are constrained to 1-state models. In both cases, I assume that every agent knows the steady-state value of every variable and perfectly observes the vector $y_t \equiv (\hat{a}_t, \hat{s}_t, \hat{\theta}_t, \hat{v}_t, \hat{u}_t, \hat{p}_t, \hat{q}_t, \hat{w}_t)$ of log-deviations from the steady state. Agents constrained to 1-state models believe that y_t follows a 1-state model of the form (1) for some $\theta = (A, B, Q, R)$ and use a pseudo-true 1-state model to forecast the future values of y.

[38] Nash bargaining only determines the total value delivered to workers and firms each and not the timing of the payoffs or the wage rate. To determine the wage, I assume that workers and firms both take the future expected wages as given and adjust the current wage to split the surplus according to the Nash bargaining solution.

Equilibrium is defined as in the previous applications. It consists of a stochastic process $\mathbb{P}$ for y_t and a model θ^* for the agents such that (i) $\mathbb{P}$ is derived from the agents' optimal behavior given their subjective model θ^*, and (ii) θ^* is a pseudo-true 1-state model given the stochastic process $\mathbb{P}$. A more formal definition can be found in Online Appendix E.

7.4 Calibration

The model is calibrated as follows. Each period corresponds to a month. The discount factor is set to $\beta = 0.99$. I set the mean of the separation rate to $s = 0.035$, so jobs last for about 2.5 years on average. The steady-state job-finding probability is set to $p = 0.4$ per month. The elasticity parameter of the matching function is set to $\alpha = 0.72$. The workers' bargaining power is set to the same value: $\delta = 0.72$. Setting $\delta = \alpha$ ensures that the Hosios condition is satisfied. I set the persistence parameter of the shock to $\rho_a = 0.96$ for labor productivity and $\rho_s = 0.90$ for separation rate. I normalize the steady-state output per worker to $a = 1$. The flow payoff to workers from unemployment is set to $b = 0.4$.[39]

The impulse response functions are independent of the volatility of the shocks and their correlation when agents have rational expectations—but not when they are constrained to one-dimensional models. I set the correlation of labor productivity and separation rate shocks to -0.4 and the ratio of the standard deviation of labor productivity to that of separation rate to ten. These choices ensure that the (pairwise) correlation coefficients between labor productivity, separation rate, and the unemployment rate are broadly consistent with the data in Shimer (2005). Finally, I normalize the standard deviation of labor productivity to one. The results that follow do not depend on this normalization.

7.5 Results

Figures 4 and 5 plot the impulse response functions to a one percent increase in labor productivity and separation rate, respectively. The responses for the case where households have rational expectations are in dashed green and for the case where agents use pseudo-true 1-state models are in solid red. Variables are all measured in log changes from their steady state values. As in the previous application, the variable labeled the "state of the economy" is defined as the agents' nowcast, $\hat{z}_t$, of the subjective state, z_t. Since the scale of z_t is not identifiable either to the agents or the econometrician, the scale of $\hat{z}_t$ is intrinsically meaningless. However, the two panels plotting the response of $\hat{z}_t$ to labor productivity and separation rate shocks use the same scale, so the responses of the state of the economy are comparable across the two shocks.

The state of the economy at time t can be expressed as a linear combination of the time-t values of the unemployment rate, labor productivity, and separation rate with the weights deter-

[39] These parameter values are all consistent with the calibration in Shimer (2005). Others, such as Hagedorn and Manovskii (2008), rely on values of b closer to one to amplify the response of unemployment to labor productivity shocks.

mined endogenously in equilibrium:[40]

$$\hat{z}_t = -0.812\hat{u}_t + 0.010\hat{a}_t - 0.177\hat{s}_t. \tag{35}$$

The state of the economy is almost five times more sensitive to changes in the unemployment rate than to changes in separation rate, and it barely responds to changes in labor productivity.

Since the shocks are significantly correlated with each other, the mapping from the persistence of the shocks to their weights in the expression for $\hat{z}_t$ is not as simple as in the RBC application. Rather, the relative attention vector p, which determines the weights different variables get in the determination of the state of economy, depends on the joint dynamics of the unemployment rate, labor productivity, and the separation rate in the way that is fleshed out in Theorem 5. Those dynamics, in turn, are determined in equilibrium as a function, among other things, of the attention vector, p.

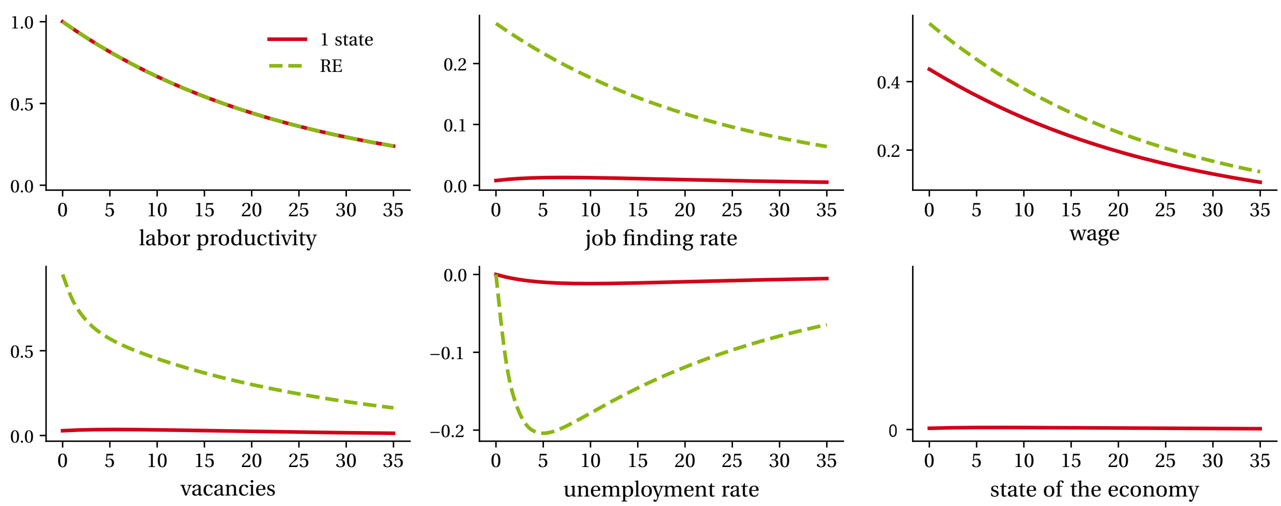

Figure 4. Impulse Response Functions to a Labor Productivity Shock

The key to understanding the economy's response to labor productivity shocks is the observation that current productivity has no direct effect on the agents' decisions. The only interesting decision in the DMP model is the firms' vacancy-creation decision, which together with the current unemployment rate, fully determines market tightness through equation (32). Market tightness and the exogenous separation rate, in turn, determine next period's unemployment rate through equation (34). Current productivity appears nowhere in these equations; it is only the firms' forecasts of future productivity that enters the dynamics of market tightness, job-finding rate, and unemployment. In fact, in equilibrium,

$$\hat{\theta}_t = 2.76\hat{z}_t,$$
$$\hat{p}_t = 0.774\hat{z}_t.$$

[40] I normalize $\hat{z}_t$ to have $\hat{z}_t = p_u\hat{u}_t + p_a\hat{a}_t + p_s\hat{s}_t$ with $|p_u| + |p_a| + |p_s| = 1$.

That is, market tightness and job-finding rate are perfectly correlated with the state of the economy.

This property of the DMP model leads to a form of complementarity in the agents' relative attention to different observables, which, in equilibrium, dampens the response of the economy to labor productivity shocks. Suppose firms reduce the weight they assign to labor productivity when forming their estimate of the current state of the economy. This makes their forecasts less sensitive to current labor productivity and dampens the effect of labor productivity on unemployment fluctuations. This, in turn, reduces labor productivity's weight in the agents' estimate of the state of the economy under a pseudo-true model. In equilibrium, labor productivity receives little weight in the agents' forecasts and has a small impact on the dynamics of vacancies, job-finding rate, and unemployment.[41]

The economy's response to a separation rate shock is perhaps even more subtle. Under rational expectations, an increase in separation rate foreshadows an increase in the unemployment rate. The increase in the unemployment rate is beneficial to would-be employers: A higher unemployment rate means a slacker labor market and a higher job-filling rate. This makes it more likely that a firm will recoup the cost of creating a vacancy, thus leading to an increase in the number of vacancies through the free-entry condition. This dynamic is behind the counterfactual positive correlation between the number of vacancies and the unemployment rate in a DMP model with only separation rate shocks.

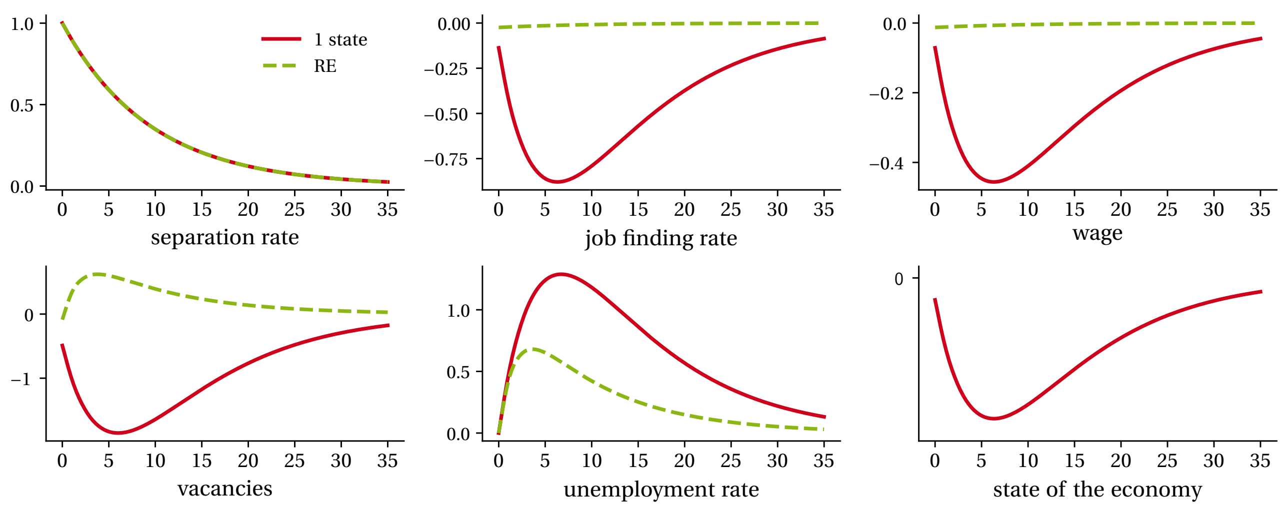

Figure 5. Impulse Response Functions to a Separation Rate Shock

Constraining agents to low-dimensional models turns this dynamic on its head. By equation (35), an increase in separations lowers the agents' nowcast of the state of the economy both directly and indirectly, through the resulting increase in unemployment. The deterioration in the firms' nowcast lowers their expectation of returns to posting a vacancy. This decrease is large

[41] Note that the wage is highly sensitive to labor productivity even when agents rely on pseudo-true 1-state models. This is due to the fact that labor productivity has a direct effect on the current wage, as can be seen in equation (33).

enough (at least, in the current calibration) to overturn the effect of the increase in the job-filling rate. As a result, firms post fewer vacancies, causing an even bigger increase in the unemployment rate.

The recession that follows an increase in separations in the 1-state version of the model has a Keynesian flavor. The increase in separations and the resulting increase in unemployment make firms pessimistic. They respond by slowing their recruiting activities, which, in turn, exacerbates the unemployment problem, darkening the outlook further, and so on. The result is an inefficiently deep and long recession.

The inability of the standard DMP model in generating realistic unemployment fluctuations in response to realistic productivity and separation shocks is known as the Shimer puzzle after Shimer (2005). The Shimer puzzle has lead to a large literature, which aims to resolve the puzzle by modifying the DMP model or Shimer's calibration of it. The exercise in this section suggests a novel path forward. It shows that constraining agents to use simple models allows even the most basic DMP model to exhibit significant amplification, propagation, and co-movement in response to separation shocks, bringing its behavior more in line with what is in the data.

A Proofs

Proof of Theorem 2

As a preliminary step, I fix an arbitrary d-state model $\theta = (A, B, Q, R)$ for the agents and compute their forecasts and the KLDR of their model from the true process. If the support of P^θ does not coincide with $\mathcal{W}$, the support of the true process, then $\text{KLDR}(\theta) = +\infty$. In what follows, I assume that P^θ is supported on $\mathcal{W}$.

Note that minimizing the KLDR over the set Θ_d of d-state models is equivalent to minimizing the KLDR over the set $\Theta_0^m \cup \Theta_1^m \cup \cdots \cup \Theta_d^m$, where Θ_k^m denotes the set of models whose *minimal* realization requires k state variables. Therefore, in the proofs, I assume without loss of generality that the d-state model θ is minimal, i.e., that there exists no d'-state model with $d' < d$ that is observationally equivalent to θ.

The Kullback–Leibler Divergence Rate. Since the entropy rate of the true process is finite, the KLDR of θ from the true process is given by

$$\text{KLDR}(\theta) = \lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[-\log f^\theta(y_1, \dots, y_t)\right] + \text{constant}.$$

On the other hand, by stationarity,

$$\lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[-\log f^\theta(y_1, \dots, y_t)\right] = \lim_{t\to\infty} \mathbb{E}\left[-\log f^\theta(y_t|y_{t-1}, \dots, y_1)\right] = \mathbb{E}\left[-\log f^\theta(y_{t+1}|y_t, \dots,)\right].$$

Therefore, to compute the KLDR, I only need to compute the subjective distribution of y_{t+1} under model θ conditional on the history of observations $\{y_t, y_{t-1}, \dots\}$.

Let $E_t^\theta[\cdot]$ denote the agents' subjective expectation given model θ and conditional on the history $\{y_\tau\}_{\tau=-\infty}^t$, and let $\text{Var}_t^\theta(\cdot)$ denote the corresponding variance-covariance matrix. Let $\hat{z}_t \equiv E_t^\theta[z_{t+1}]$ denote the agents' subjective conditional expectation of the subjective state. I can express $\hat{z}_t$ recursively using the Kalman filter:

$$\hat{z}_t = (A - KB')\hat{z}_{t-1} + Ky_t, \tag{A.1}$$

where $K \in \mathbb{R}^{d\times n}$ is the Kalman gain defined as

$$K \equiv A\hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger, \tag{A.2}$$

the dagger denotes the Moore–Penrose pseudo-inverse, $\hat{\Sigma}_z \equiv \text{Var}_t^\theta(z_{t+1})$ is the subjective conditional variance of z_{t+1}, which solves the following (generalized) algebraic Riccati equation

$$\hat{\Sigma}_z = A\left(\hat{\Sigma}_z - \hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger B'\hat{\Sigma}_z\right)A' + Q, \tag{A.3}$$

and A' denotes the transpose of matrix A.[42] Solving equation (A.1) backward, I get

$$\hat{z}_t = \sum_{\tau=0}^{\infty}(A - KB')^\tau K y_{t-\tau}.$$

[42]See, for instance, Chapter 4 of Anderson and Moore (2005). Note that I allow for the possibility that P^θ is supported on some proper subspace $\mathcal{W}$ of $\mathbb{R}^n$, in which case $B'\hat{\Sigma}_z B + R$ might not be invertible. The Moore–Penrose pseudo-inverse is then the appropriate generalization of matrix inverse in the expression for the Kalman gain.

The agents' subjective conditional expectation of y_{t+1} can be written in terms of their conditional expectation of z_{t+1}:

$$E_t^{\theta}[y_{t+1}] = B' E_t^{\theta}[z_{t+1}] = B' \sum_{\tau=0}^{\infty} (A - KB')^{\tau} K y_{t-\tau}.$$

Likewise, the conditional variance of y_{t+1} can be expressed in terms of the conditional variance of z_{t+1}:

$$\hat{\Sigma}_y \equiv \mathrm{Var}_t^{\theta}(y_{t+1}) = B' \hat{\Sigma}_z B + R. \tag{A.4}$$

More generally, the agents' s-period ahead forecast of the vector of observables is given by

$$E_t^{\theta}[y_{t+s}] = B' A^{s-1} E_t^{\theta}[z_{t+1}] = B' A^{s-1} \sum_{\tau=0}^{\infty} (A - KB')^{\tau} K y_{t-\tau}. \tag{A.5}$$

The Kullback–Leibler divergence rate is thus equal to

$$\begin{aligned} \mathrm{KLDR}(\theta) = & -\frac{1}{2} \log \det{}^{*} \left(\hat{\Sigma}_y^{\dagger} \right) + \frac{n}{2} \log (2\pi) + \frac{1}{2} \operatorname{tr} \left(\hat{\Sigma}_y^{\dagger} \Gamma_0 \right) \\ & - \frac{1}{2} \sum_{\tau=1}^{\infty} \operatorname{tr} \left(\hat{\Sigma}_y^{\dagger} \Phi_{\tau} \Gamma_{\tau}' \right) - \frac{1}{2} \sum_{\tau=1}^{\infty} \operatorname{tr} \left(\hat{\Sigma}_y^{\dagger} \Gamma_{\tau} \Phi_{\tau}' \right) \\ & + \frac{1}{2} \sum_{s=1}^{\infty} \sum_{\tau=1}^{\infty} \operatorname{tr} \left(\hat{\Sigma}_y^{\dagger} \Phi_s \Gamma_{\tau-s} \Phi_{\tau}' \right) + \text{constant}, \end{aligned} \tag{A.6}$$

where $\Gamma_l \equiv \mathbb{E}[y_t y_{t-l}']$ denotes the lag-l autocovariance matrix for the vector of observables under the true process, $\Phi_{\tau} \equiv B'(A - KB')^{\tau-1} K$, and the constant contains terms that do not depend on θ. The matrix $\hat{\Sigma}_y^{\dagger}$ denotes the Moore–Penrose pseudo-inverse of $\hat{\Sigma}_y$ and $\det^{*}(\hat{\Sigma}_y^{\dagger})$ denotes its pseudo-determinant. These objects are the appropriate counterparts of the matrix inverse and the determinant for the case when $\mathcal{W}$ does not equal $\mathbb{R}^n$, and so, the subjective model θ is degenerate.

Proof of Theorem 2. Let $\tilde{n}$ denote the dimension of vector $\tilde{y}_t = T y_t$, let $\widetilde{\mathcal{W}}$ denote the linear subspace of $\mathbb{R}^{\tilde{n}}$ defined as $\widetilde{\mathcal{W}} \equiv \{\tilde{y} \in \mathbb{R}^{\tilde{n}} : \tilde{y} = T y \text{ for some } y \in \mathcal{W}\}$, let $\widetilde{\Theta}_d$ denote the set of d-state models when the vector of observable is $\tilde{y}_t \in \mathbb{R}^{\tilde{n}}$, let $\widetilde{\mathrm{KLDR}}(\tilde{\theta})$ denote the KLDR of model $\tilde{\theta} \in \widetilde{\Theta}_d$ from the true process $\widetilde{\mathbb{P}} \equiv \mathbb{P} \circ T^{-1}$, and let $\widetilde{\Theta}_d^{*}$ denote the set of models $\tilde{\theta}^{*} \in \widetilde{\Theta}_d$ such that $\widetilde{\mathrm{KLDR}}(\tilde{\theta}^{*}) \leq \widetilde{\mathrm{KLDR}}(\tilde{\theta})$ for all $\tilde{\theta} \in \widetilde{\Theta}_d$.

I first show that $\widetilde{\mathcal{W}}$ is the support of any distribution in the set $\mathcal{P}_d^{*} \circ T^{-1}$ of distributions over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$ induced by $\mathcal{P}_d^{*}$ and T as well as the support of any distribution in the set $\widetilde{\mathcal{P}}_d^{*}$. Note that there always exists a d-state model θ for which $\mathrm{KLDR}(\theta) < \infty$—one such model is the one according to which y_t is i.i.d. over time and has a variance-covariance matrix that coincides with the true variance-covariance matrix, Γ_0. Therefore, for any pseudo-true d-state model, the KLDR is finite. Thus, any process $P \in \mathcal{P}_d^{*}$ is supported on $\mathcal{W}$, and so, any process $P \in \mathcal{P}_d^{*} \circ T^{-1}$ is supported on $\widetilde{\mathcal{W}}$. On the other hand, since the true distribution $\mathbb{P}$ is supported on $\mathcal{W}$, the induced distribution $\widetilde{\mathbb{P}} \equiv \mathbb{P} \circ T^{-1}$ is supported on $\widetilde{\mathcal{W}}$. Consequently, by the above argument, any distribution $P \in \widetilde{\mathcal{P}}_d^{*}$

is also supported on $\widetilde{\mathcal{W}}$. Therefore, I can restrict my attention to models $\theta \in \Theta_d$ such that P^θ is supported on $\mathcal{W}$ and models $\tilde{\theta} \in \widetilde{\Theta}_d$ such that $P^{\tilde{\theta}}$ is supported on $\widetilde{\mathcal{W}}$.

For any model $\theta = (A, B, Q, R) \in \Theta_d$, define the model $T(\theta) \in \widetilde{\Theta}_d$ as $T(\theta) \equiv (A, BT', Q, TRT')$. I next show that $\widetilde{\text{KLDR}}(T(\theta)) = \text{KLDR}(\theta)$, up to an additive constant that does not depend on θ. Fix some model $\theta \in \Theta_d$. Let $\hat{\Sigma}_z \equiv \text{Var}_t^\theta(z_{t+1})$ denote the subjective conditional variance of the subjective state under model θ, and let $\widetilde{\hat{\Sigma}}_z \equiv \text{Var}_t^{T(\theta)}(z_{t+1})$ denote the corresponding conditional variance under model $T(\theta)$. Matrices $\hat{\Sigma}_z$ and $\widetilde{\hat{\Sigma}}_z$ solve the following Riccati equations:

$$\hat{\Sigma}_z = A\left(\hat{\Sigma}_z - \hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger B'\hat{\Sigma}_z\right)A' + Q, \tag{A.7}$$

$$\widetilde{\hat{\Sigma}}_z = A\left(\widetilde{\hat{\Sigma}}_z - \widetilde{\hat{\Sigma}}_z BT'\left(TB'\widetilde{\hat{\Sigma}}_z BT' + TRT'\right)^\dagger TB'\widetilde{\hat{\Sigma}}_z\right)A' + Q. \tag{A.8}$$

Since matrix T has full rank, $T^\dagger = (T'T)^{-1}T$ and $T^\dagger T = I$. Therefore, $\widetilde{\hat{\Sigma}}_z = \hat{\Sigma}_z$. Next, let K denote the Kalman gain given model θ, and let denote $\tilde{K}$ denote the Kalman gain given model $T(\theta)$. Note that

$$\tilde{K} = A\widetilde{\hat{\Sigma}}_z BT'\left(TB'\widetilde{\hat{\Sigma}}_z BT' + TRT'\right)^\dagger = KT^\dagger.$$

Let $\Phi_\tau \equiv B'(A-KB')^{\tau-1}K$ given model θ, and let $\widetilde{\Phi}_\tau$ denote the corresponding matrix given model $T(\theta)$. Note that

$$\widetilde{\Phi}_\tau \equiv TB'(A - KT^\dagger TB')^{\tau-1}KT^\dagger = T\Phi_\tau T^\dagger.$$

Finally, let $\hat{\Sigma}_y \equiv \text{Var}_t^\theta(y_{t+1})$ denote the subjective conditional variance of y_{t+1} given model θ, and let $\widetilde{\hat{\Sigma}}_y \equiv \text{Var}_t^{T(\theta)}(\tilde{y}_{t+1})$ denote the corresponding conditional variance given model $T(\theta)$. Note that

$$\widetilde{\hat{\Sigma}}_y = TB'\widetilde{\hat{\Sigma}}_z BT' + TRT' = T\hat{\Sigma}_y T'.$$

One the other hand, $\tilde{\Gamma}_l \equiv \tilde{\mathbb{E}}[\tilde{y}_t\tilde{y}_{t-l}'] = T\mathbb{E}[y_t y_{t-l}]T' = T\Gamma_l T'$. Therefore, by equation (A.6),

$$\begin{aligned}
\widetilde{\text{KLDR}}(T(\theta)) = & -\frac{1}{2}\log\det{}^*\left(T^{\dagger\prime}\hat{\Sigma}_y^\dagger T^\dagger\right) + \frac{n}{2}\log(2\pi) + \frac{1}{2}\operatorname{tr}\left(T^{\dagger\prime}\hat{\Sigma}_y^\dagger T^\dagger T\Gamma_0 T'\right) \\
& -\frac{1}{2}\sum_{\tau=1}^\infty \operatorname{tr}\left(T^{\dagger\prime}\hat{\Sigma}_y^\dagger T^\dagger T\Phi_\tau T^\dagger T\Gamma_\tau' T'\right) - \frac{1}{2}\sum_{\tau=1}^\infty \operatorname{tr}\left(T^{\dagger\prime}\hat{\Sigma}_y^\dagger T^\dagger T\Gamma_\tau T' T^{\dagger\prime}\Phi_\tau' T'\right) \\
& +\frac{1}{2}\sum_{s=1}^\infty\sum_{\tau=1}^\infty \operatorname{tr}\left(T^{\dagger\prime}\hat{\Sigma}_y^\dagger T^\dagger T\Phi_s T^\dagger T\Gamma_{\tau-s} T' T^{\dagger\prime}\Phi_\tau' T'\right) + \text{constant}.
\end{aligned}$$

The fact that $T^\dagger T = I$ implies that the above expression is equal to $\text{KLDR}(\theta)$, up to an additive constant that does not depend on θ.

Likewise, for any model $\tilde{\theta} = (\tilde{A}, \tilde{B}, \tilde{Q}, \tilde{R}) \in \widetilde{\Theta}_d$, define $T^{-1}(\tilde{\theta}) \equiv (\tilde{A}, \tilde{B}T^{\dagger\prime}, \tilde{Q}, T^\dagger\tilde{R}T^{\dagger\prime}) \in \Theta_d$. By an argument similar to the one in the previous paragraph, $\text{KLDR}(T^{-1}(\tilde{\theta})) = \widetilde{\text{KLDR}}(\tilde{\theta})$, up to an additive constant that does not depend on $\tilde{\theta}$.

Therefore, the mapping T defines an isomorphism between the set of models Θ_d and the set of models $\tilde{\Theta}_d$: Any model $\theta \in \Theta_d$ can be identified with a model $T(\theta) \in \widetilde{\Theta}_d$ such that the KLDR of P^θ from the process $\mathbb{P}$ is equal to the KLDR of $P^{T(\theta)}$ from the process $\mathbb{P} \circ T^{-1}$, and any model $\tilde{\theta} \in \widetilde{\Theta}_d$

can be identified with a model $T^{-1}(\tilde{\theta}) \in \Theta_d$ such that the KLDR of $P^{T^{-1}(\tilde{\theta})}$ from the process $\mathbb{P}$ is equal to the KLDR of $P^{\tilde{\theta}}$ from the process $\mathbb{P} \circ T^{-1}$. This conclusion immediately implies that the set of pseudo-true d-state models given the true process $\mathbb{P}$ is identified with the set of pseudo-true d-state models given the true process $\mathbb{P} \circ T^{-1}$. That is, $\widetilde{\Theta}^*_d = \{T(\theta) : \theta \in \Theta^*_d\}$.

It only remains to show that $P^{T(\theta)} = P^{\theta} \circ T^{-1}$ for any model $\theta \in \Theta_d$. Since $P^{T(\theta)}$ and $P^{\theta} \circ T^{-1}$ are both zero mean, stationary, and normal distributions over $\{\tilde{y}_t\}_{t=-\infty}^{\infty}$, it is sufficient to show that the autocovariance matrices of $\tilde{y}_t$ are identical at all lags under the two distributions. But this follows the definitions of distributions $P^{T(\theta)}$ and $P^{\theta} \circ T^{-1}$. ◻

Proof of Theorem 3

Before establishing the theorem, I state and prove a useful lemma:

Lemma A.1. *Model $\theta = (A, B, Q, R)$ is a pseudo-true d-state model given true autocovariance matrices $\{\Gamma_l\}_l$ if and only if $A = M$, $B = D'N^{-1}$, $Q = I - M(I - D'D)M'$, and $R = N^{-1\prime}(I - DD')N^{-1}$, where M, D, and N are, respectively, a $d \times d$ convergent matrix, a $n \times d$ diagonal matrix with elements in the $[0, 1]$ interval, and an $n \times n$ invertible matrix that maximize*

$$
\begin{aligned}
&-\frac{1}{2}\log\det(NN') + \frac{1}{2}\operatorname{tr}(N'\Gamma_0 N) - \sum_{\tau=1}^{\infty}\operatorname{tr}\left((M(I - D'D))^{\tau-1} MD'N'\Gamma_\tau' ND\right) \\
&\quad + \frac{1}{2}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(D(M(I - D'D))^{s-1} MD'N'\Gamma_{\tau-s} NDM'((I - D'D)M')^{\tau-1} D'\right). \qquad \text{(A.9)}
\end{aligned}
$$

Proof. I can assume without loss of generality that $\hat{\Sigma}_z$ is invertible.[43] I start by expressing $\hat{\Sigma}_z^{\frac{1}{2}} B \hat{\Sigma}_y^{\frac{-1}{2}}$ as its singular value decomposition:

$$
\hat{\Sigma}_z^{\frac{1}{2}} B \hat{\Sigma}_y^{\frac{-1}{2}} = UDV',
$$

where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{d\times d}$ are orthogonal matrices, and $D \in \mathbb{R}^{n\times d}$ is a rectangular diagonal matrix with singular values of $\hat{\Sigma}_z^{\frac{1}{2}} B \hat{\Sigma}_y^{\frac{-1}{2}}$ on the diagonal. Note that

$$
VD'DV' = \hat{\Sigma}_z^{\frac{1}{2}} B \left(B'\hat{\Sigma}_z B + R\right)^{-1} B'\hat{\Sigma}_z^{\frac{1}{2}}.
$$

Since R is a symmetric positive semidefinite matrix and V is orthogonal, diagonal elements of D are weakly smaller than 1 (strictly so if R is not singular). Next define

$$
M \equiv V^{-1}\hat{\Sigma}_z^{\frac{-1}{2}} A \hat{\Sigma}_z^{\frac{1}{2}} V.
$$

Then,

$$
A = \hat{\Sigma}_z^{\frac{1}{2}} VMV^{-1}\hat{\Sigma}_z^{\frac{-1}{2}}, \qquad \text{(A.10)}
$$

$$
B = \hat{\Sigma}_z^{\frac{-1}{2}} VD'U'\hat{\Sigma}_y^{\frac{1}{2}}, \qquad \text{(A.11)}
$$

[43] This is due to the fact that any d-state model with a singular $\hat{\Sigma}_z$ is observationally equivalent to a d'-state model with $d' < d$.

$$K = \hat{\Sigma}_z^{\frac{1}{2}} VMD'U'\hat{\Sigma}_y^{\frac{-1}{2}}, \tag{A.12}$$

and so

$$KB' = \hat{\Sigma}_z^{\frac{1}{2}} VMD'DV'\hat{\Sigma}_z^{\frac{-1}{2}},$$
$$\Phi_\tau = \hat{\Sigma}_y^{\frac{1}{2}} UD\left(M\left(I-D'D\right)\right)^{\tau-1} MD'U'\hat{\Sigma}_y^{\frac{-1}{2}}.$$

I can further reduce the number of parameters in the agent's model by transforming $\hat{\Sigma}_y^{\frac{-1}{2}}$ using the orthogonal matrix U. Define

$$N \equiv \hat{\Sigma}_y^{\frac{-1}{2}} U.$$

Note that since $\hat{\Sigma}_y^{\frac{-1}{2}}$ is symmetric,

$$UN' = NU' = \hat{\Sigma}_y^{\frac{-1}{2}},$$

so

$$\hat{\Sigma}_y^{-1} = NU'UN' = NN',$$

and

$$\operatorname{tr}\left(\hat{\Sigma}_y^{-1}\Gamma_0\right) = \operatorname{tr}\left(\hat{\Sigma}_y^{\frac{-1}{2}}\Gamma_0\hat{\Sigma}_y^{\frac{-1}{2}}\right) = \operatorname{tr}\left(UN'\Gamma_0 NU'\right) = \operatorname{tr}\left(N'\Gamma_0 N\right).$$

On the other hand,

$$\begin{aligned}\operatorname{tr}\left(\hat{\Sigma}_y^{-1}\Phi_\tau\Gamma_\tau'\right) &= \operatorname{tr}\left(\hat{\Sigma}_y^{\frac{-1}{2}} UD\left(M\left(I-D'D\right)\right)^{\tau-1} MD'U'\hat{\Sigma}_y^{\frac{-1}{2}}\Gamma_\tau'\right)\\ &= \operatorname{tr}\left(\left(M\left(I-D'D\right)\right)^{\tau-1} MD'N'\Gamma_\tau' ND\right),\end{aligned}$$

and

$$\operatorname{tr}\left(\hat{\Sigma}_y^{-1}\Phi_s\Gamma_{\tau-s}\Phi_\tau'\right) = \operatorname{tr}\left(D\left(M\left(I-D'D\right)\right)^{s-1} MD'N'\Gamma_{\tau-s}NDM'\left(\left(I-D'D\right)M'\right)^{\tau-1}D'\right).$$

Therefore, the KLDR can be expressed in terms of matrices M, D, and N as

$$\text{KLDR}(\theta) = \text{KLDR}(M,D,N) + \text{constant},$$

where, with some abuse of notation, I let

$$\begin{aligned}\text{KLDR}(M,D,N) \equiv\ & -\frac{1}{2}\log\det\left(NN'\right) + \frac{1}{2}\operatorname{tr}\left(N'\Gamma_0 N\right) - \sum_{\tau=1}^{\infty}\operatorname{tr}\left(\left(M\left(I-D'D\right)\right)^{\tau-1} MD'N'\Gamma_\tau' ND\right)\\ &+\frac{1}{2}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(D\left(M\left(I-D'D\right)\right)^{s-1} MD'N'\Gamma_{\tau-s}NDM'\left(\left(I-D'D\right)M'\right)^{\tau-1}D'\right).\end{aligned} \tag{A.13}$$

Any non-singular matrix N has a unique decomposition as $N = \hat{\Sigma}_y^{\frac{-1}{2}} U$, where $\hat{\Sigma}_y^{\frac{-1}{2}}$ is a symmetric positive definite matrix and U is a orthogonal matrix.[44] On the other hand, for any positive definite matrix $\hat{\Sigma}_y$, there exists a positive definite matrix R that satisfies the Riccati equation (A.3).

[44] This is known as the polar decomposition of matrix N.

Therefore, minimizing KLDR(θ) with respect to θ is equivalent to minimizing KLDR(M, D, N) with respect to M, D, and N subject to the constraints that M is a convergent matrix, D is a rectangular diagonal matrix with diagonal elements in the interval $[0, 1]$, and N is a non-singular matrix.

Next I show how given a tuple (M, D, N) that minimizes KLDR(M, D, N) one can find the corresponding parameters (A, B, Q, R) of the representation in (1). Let $T \equiv \hat{\Sigma}_z^{\frac{1}{2}} V$. Then, by Equation (A.10),

$$A = TMT^{-1}. \tag{A.14}$$

Substituting for $\hat{\Sigma}_y$ in equations (A.11) and (A.4), I get

$$B = (T^{-1})' D' N^{-1}, \tag{A.15}$$

$$R = N^{-1\prime} \left(I - DD'\right) N^{-1}. \tag{A.16}$$

Substituting in (A.3) for A from equation (A.10) and for B from the above equation, I get

$$\begin{aligned}
Q &= \hat{\Sigma}_z - A\left(\hat{\Sigma}_z - \hat{\Sigma}_z B \left(B'\hat{\Sigma}_z B + R\right)^{-1} B'\hat{\Sigma}_z\right) A' \\
&= \hat{\Sigma}_z - \hat{\Sigma}_z^{\frac{1}{2}} V M V^{-1} \hat{\Sigma}_z^{\frac{-1}{2}} \left(\hat{\Sigma}_z - \hat{\Sigma}_z^{\frac{1}{2}} V D' D V \hat{\Sigma}_z^{\frac{1}{2}}\right) \hat{\Sigma}_z^{\frac{-1}{2}} V M' V^{-1} \hat{\Sigma}_z^{\frac{1}{2}} \\
&= \hat{\Sigma}_z - \hat{\Sigma}_z^{\frac{1}{2}} V M \left(I - D'D\right) M' V^{-1} \hat{\Sigma}_z^{\frac{1}{2}}.
\end{aligned}$$

Therefore,

$$Q = T\left(I - M\left(I - D'D\right) M'\right) T'. \tag{A.17}$$

While (M, D, N) are pinned down by the minimization of the Kullback–Leibler divergence, T is not identified. However, for any non-singular T and $\widetilde{T} \neq T$, the models described by (M, D, N, T) and $(M, D, N, \widetilde{T})$ are observationally equivalent.[45] Therefore, without loss of generality, I can set $T = I$ to get to the following representation:

$$\begin{aligned}
A &= M, \\
B &= D' N^{-1}, \\
Q &= I - M\left(I - D'D\right) M', \\
R &= N^{-1\prime}\left(I - DD'\right) N^{-1}.
\end{aligned}$$

This completes the proof of the lemma.

For future reference, I also compute several other objects under the above representation. The matrix of Kalman gain is given by

$$K = MD'N'. \tag{A.18}$$

The subjective forecasts can then be found by substituting for A, B, and K in (A.5):

$$E_t^{\theta}\left[y_{t+s}\right] = N'^{-1} D M^{s-1} \sum_{\tau=0}^{\infty} \left(M\left(I - D'D\right)\right)^{\tau} MD'N' y_{t-\tau}. \tag{A.19}$$

[45] See, for instance, Gevers and Wertz (1984).

The subjective variance of y_{t+1} conditional on the information available to the agents at time t is given by

$$\hat{\Sigma}_y = (NN')^{-1}.$$

The unconditional subjective variance of y is given by

$$\text{Var}^\theta(y) = B'\text{Var}^\theta(z)B + R,$$

where $\text{Var}(z)$ solves the discrete Lyapunov equation

$$\text{Var}^\theta(z) = A\text{Var}^\theta(z)A' + Q.$$

Solving the above equation forward, I get

$$\text{Var}^\theta(z) = I + \sum_{\tau=1}^{\infty} M^\tau D'DM'^\tau.$$

Therefore,

$$\text{Var}^\theta(y) = B'\sum_{\tau=0}^{\infty} A^\tau QA'^\tau B + R = N^{-1\prime}\left(I + \sum_{\tau=1}^{\infty} DM^\tau D'DM'^\tau D'\right)N^{-1}. \tag{A.20}$$

□

I can now establish Theorem 3.

Proof of Theorem 3. Let M, D, and N be as in Lemma A.1. When $d = 1$, then

$$M = a$$

for some $a \in [-1, 1]$ and

$$D = \begin{pmatrix} d_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = d_1 e_1$$

for some $d_1 \in [0, 1]$, where e_1 denotes the first coordinate vector. Define

$$\eta \equiv 1 - d_1^2,$$
$$S \equiv \Gamma_0^{\frac{1}{2}} N.$$

Then KLDR, defined in (A.9), can be written as a function of a, η, and S, with slight abuse of notation:

$$\text{KLDR}(a, \eta, T) = -\frac{1}{2}\log\det(SS') + \frac{1}{2}\text{tr}(S'S) - \frac{1}{2}e_1'S'\Omega(a, \eta)Se_1,$$

where

$$\Omega(a, \eta) \equiv a(1-\eta)\sum_{\tau=1}^{\infty}(a\eta)^{\tau-1}\Gamma_0^{\frac{-1}{2}}(\Gamma_\tau + \Gamma_\tau')\Gamma_0^{\frac{-1}{2}} - a^2(1-\eta)^2\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}(a\eta)^{s+\tau-2}\Gamma_0^{\frac{-1}{2}}\Gamma_{\tau-s}\Gamma_0^{\frac{-1}{2}}.$$

I can simplify the second term of $\Omega(a,\eta)$ further:

$$\begin{aligned}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}(a\eta)^{s+\tau-2}\Gamma_0^{\frac{-1}{2}}\Gamma_{\tau-s}\Gamma_0^{\frac{-1}{2}} &= \sum_{s=1}^{\infty}\sum_{\tau=s+1}^{\infty}(a\eta)^{s+\tau-2}\Gamma_0^{\frac{-1}{2}}\left(\Gamma_{\tau-s}+\Gamma'_{\tau-s}\right)\Gamma_0^{\frac{-1}{2}}+\sum_{s=1}^{\infty}(a\eta)^{2(s-1)}I\\ &= \sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}(a\eta)^{2(s-1)+\tau}\Gamma_0^{\frac{-1}{2}}\left(\Gamma_{\tau}+\Gamma'_{\tau}\right)\Gamma_0^{\frac{-1}{2}}+\sum_{s=1}^{\infty}(a\eta)^{2(s-1)}I\\ &= \left(\sum_{s=1}^{\infty}(a\eta)^{2(s-1)}\right)\left(I+\sum_{\tau=1}^{\infty}(a\eta)^{\tau}\Gamma_0^{\frac{-1}{2}}(\Gamma_\tau+\Gamma'_\tau)\Gamma_0^{\frac{-1}{2}}\right)\\ &= \frac{1}{1-a^2\eta^2}\left(I+a\eta\sum_{\tau=1}^{\infty}(a\eta)^{\tau-1}\Gamma_0^{\frac{-1}{2}}(\Gamma_\tau+\Gamma'_\tau)\Gamma_0^{\frac{-1}{2}}\right).\end{aligned}$$

Therefore,

$$\Omega(a,\eta)=-\frac{a^2(1-\eta)^2}{1-a^2\eta^2}I+\frac{(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\sum_{\tau=1}^{\infty}a^{\tau}\eta^{\tau-1}\Gamma_0^{\frac{-1}{2}}(\Gamma_\tau+\Gamma'_\tau)\Gamma_0^{\frac{-1}{2}}. \tag{A.21}$$

By Lemma A.1, minimizing the KLDR with respect to A, B, Q, and R is equivalent to minimizing KLDR(M,D,N) with respect to M, D, and N. But for any a, η, and S, one can construct a corresponding M, D, and N. Therefore, I can instead minimize KLDR(a,η,S) with respect to a, η, and S.

I first minimize KLDR(a,η,S) with respect to S taking a and η as given. The first-order optimality condition with respect to S is given by

$$S^{-1}=S'-e_1e_1'S'\Omega(a,\eta),$$

which implies that

$$S'S-e_1e_1'T'\Omega(a,\eta)S=I. \tag{A.22}$$

Therefore, for any solution to the problem of minimizing KLDR(a,η,S),

$$n=\mathrm{tr}(I)=\mathrm{tr}\left(S'S\right)-\mathrm{tr}\left(e_1e_1'S'\Omega(a,\eta)S\right)=\mathrm{tr}\left(S'S\right)-e_1'S'\Omega(a,\eta)Se_1.$$

So, minimizing KLDR(a,η,S) respect to a, η, and S is equivalent to solving the following program:

$$\max_{a,\eta}\ \det\left(S(a,\eta)S'(a,\eta)\right),$$

where

$$S(a,\eta)\in\arg\min_{S}\ -\frac{1}{2}\log\det\left(SS'\right)+\frac{1}{2}\mathrm{tr}\left(S'S\right)-\frac{1}{2}e_1'S'\Omega(a,\eta)Se_1. \tag{A.23}$$

I proceed by first characterizing $S(a,\eta)$. Note that the necessary first-order optimality conditions for problem (A.23) are given by the matrix equation (A.22).

Claim A.1. *For any matrix S that solves equation* (A.22)*, the necessary first-order optimality condition for problem* (A.23)*,* *(i)* $Se_1 = \frac{1}{\sqrt{1-\lambda}} u$,

(ii) $S'^{-1} e_1 = \sqrt{1-\lambda} u$,

(iii) $SS' = I + \frac{\lambda}{1-\lambda} uu'$,

where λ is an eigenvalue of the real symmetric matrix $\Omega(a, \eta)$ and u is a corresponding eigenvector normalized such that $u'u = 1$.

I return to proving the claim toward the end of the proof. Equation (A.22) in general has multiple solutions, with each solution corresponding to a local extremum of problem (A.23). The global optimum of problem (A.23) is given by the solution to equation (A.22) that results in the largest value for $\det(SS')$. But by part (iii) of Claim A.1, $\det(SS') = (1-\lambda)^{-1}$. Thus, for any pseudo-true 1-state model, a and η maximize $\lambda_{\max}(\Omega(a, \eta))$ and S satisfies parts (i)–(iii) of Claim A.1, with $\lambda = \lambda_{\max}(\Omega)$ and $u = u_{\max}(\Omega)$ the corresponding eigenvector.

I next find the parameters A, B, Q, and R representing the a, η, and S that minimize KLDR(a, η, S). First, note that

$$\begin{aligned} M &= a, \\ D &= \sqrt{1-\eta} e_1, \\ N &= \Gamma_0^{\frac{-1}{2}} S. \end{aligned}$$

The representation in Lemma A.1 is thus given by

$$\begin{aligned} A &= a, \\ B &= \sqrt{1-\eta} e_1' S^{-1} \Gamma_0^{\frac{1}{2}}, \\ Q &= 1 - a^2 \eta, \\ R &= \Gamma_0^{\frac{1}{2}} S^{-1'} \left(I - (1-\eta) e_1 e_1' \right) S^{-1} \Gamma_0^{\frac{1}{2}}. \end{aligned}$$

By Claim A.1 and the argument above,

$$\begin{aligned} e_1' S^{-1} &= \sqrt{1 - \lambda_{\max}(\Omega)} u_{\max}'(\Omega), \\ S^{-1'} S^{-1} &= (SS')^{-1} = I - \lambda_{\max}(\Omega) u_{\max}(\Omega) u_{\max}'(\Omega). \end{aligned}$$

Thus,

$$B = \sqrt{(1-\eta)\left(1 - \lambda_{\max}(\Omega)\right)} u_{\max}'(\Omega) \Gamma_0^{\frac{1}{2}},$$

and

$$\begin{aligned} R &= \Gamma_0^{\frac{1}{2}} \left(I - \lambda_{\max}(\Omega) u_{\max}(\Omega) u_{\max}'(\Omega) \right) \Gamma_0^{\frac{1}{2}} - (1-\eta)\left(1 - \lambda_{\max}(\Omega)\right) \Gamma_0^{\frac{1}{2}} u_{\max}(\Omega) u_{\max}'(\Omega) \Gamma_0^{\frac{1}{2}} \\ &= \Gamma_0^{\frac{1}{2}} \left[I - \left(1 - \eta + \eta \lambda_{\max}(\Omega)\right) u_{\max}(\Omega) u_{\max}'(\Omega) \right] \Gamma_0^{\frac{1}{2}}. \end{aligned}$$

Finally, note that $M = a$, $D = \sqrt{1-\eta}e_1$, and $N = \Gamma_0^{\frac{-1}{2}} S$. Therefore, by equation (A.19), the subjective forecasts are given by

$$E_t^{\theta}[y_{t+s}] = a^s(1-\eta)\Gamma_0^{\frac{1}{2}} S'^{-1} e_1 e_1' S' \Gamma_0^{\frac{-1}{2}} \sum_{\tau=0}^{\infty} a^{\tau}\eta^{\tau} y_{t-\tau}. \tag{A.24}$$

Using Claim A.1 to substitute for the optimal S, I get

$$E_t^{\theta}[y_{t+s}] = a^s(1-\eta)\Gamma_0^{\frac{1}{2}} u_{\max}(\Omega) u_{\max}'(\Omega) \Gamma_0^{\frac{-1}{2}} \sum_{\tau=0}^{\infty} a^{\tau}\eta^{\tau} y_{t-\tau},$$

where $u_{\max}(\Omega)$ is a unit-norm eigenvector of Ω with eigenvalue $\lambda_{\max}(\Omega)$. The theorem then follows by the definition of p and q. □

Proof of Claim A.1. The first-order optimality condition with respect to S is given by

$$S'S - e_1 e_1' S' \Omega T = I. \tag{A.25}$$

Multiplying the transpose of the above equation from right by e_1 and from left by S'^{-1}, I get

$$Se_1 - \Omega S e_1 = S'^{-1} e_1. \tag{A.26}$$

On the other hand, multiplying equation (A.25) from left by S and from right by S^{-1}, I get

$$SS' = I + Se_1 e_1' S' \Omega. \tag{A.27}$$

By the Sherman–Morrison formula,

$$S'^{-1} S^{-1} = I - \frac{Se_1 e_1' S' \Omega}{1 + e_1' S' \Omega S e_1}$$

Multiplying the above equation from right by Se_1, I get

$$S'^{-1} e_1 = \frac{1}{1 + e_1' S' \Omega S e_1} S e_1. \tag{A.28}$$

Substituting for $S'^{-1}e_1$ from the above equation in (A.26) and rearranging the terms, I get

$$\Omega S e_1 = \frac{e_1' S' \Omega S e_1}{1 + e_1' S' \Omega S e_1} S e_1. \tag{A.29}$$

That is, Se_1 is an eigenvector of Ω. Let λ denote the corresponding eigenvalue and let $u = Se_1/\sqrt{e_1' S' S e_1}$. Then equation (A.29) implies

$$\lambda = \frac{\lambda e_1' S' S e_1}{1 + \lambda e_1' S' S e_1}.$$

I separately consider the cases $\lambda \neq 0$ and $\lambda = 0$. If $\lambda \neq 0$, then

$$e_1' S' S e_1 = (1-\lambda)^{-1},$$

and so

$$Se_1 = \frac{1}{\sqrt{1-\lambda}}u.$$

Equation (A.28) then implies that

$$S'^{-1}e_1 = \sqrt{1-\lambda}u,$$

and equation (A.27) implies that

$$SS' = I + \frac{\lambda}{1-\lambda}uu'.$$

If $\lambda = 0$, then equation (A.26) implies that $Se_1 = S'^{-1}e_1$, and so, Se_1 and $S'^{-1}e_1$ are both multiples of u. Furthermore, $e_1'S^{-1}Se_1 = e_1'e_1 = 1$. Therefore, $Se_1 = S'^{-1}e_1 = u$. On the other hand, equation (A.27) implies that $SS' = I$. This completes the proof of the claim. □

Proof of Theorem 4

I first prove a useful lemma about the spectral radius of autocorrelation matrices when the true process is stationary ergodic:

Lemma A.2. *For a stationary ergodic process with autocorrelation matrices $\{C_l\}_l$, the spectral radius of autocorrelation matrices satisfies $\rho(C_l) \leq 1$ for any l with the inequality strict for $l = 1$.*

Proof. Let λ_l denote an eigenvalue of C_l largest in magnitude and let u_l denote the corresponding eigenvector normalized such that $u_l'u_l = 1$. Define the process $\omega_t^{(l)} \equiv u_l'\Gamma_0^{\frac{-1}{2}}y_t \in \mathbb{R}$. Since y_t is stationary ergodic, so is $\omega_t^{(l)}$ for any l. Furthermore, since Γ_0 is non-singular, the process $\omega_t^{(l)}$ is non-degenerate for any l. I first show that λ_l is the autocorrelation of the process $\omega_t^{(l)}$ at lag l. Note that

$$\mathbb{E}[\omega_t^{(l)}\omega_{t-l}^{(l)}] = u_l'\Gamma_0^{\frac{-1}{2}}\mathbb{E}[y_ty_{t-l}']\Gamma_0^{\frac{-1}{2}}u_l = u_l'\Gamma_0^{\frac{-1}{2}}\Gamma_l\Gamma_0^{\frac{-1}{2}}u_l = u_l'\Gamma_0^{\frac{-1}{2}}\left(\frac{\Gamma_l+\Gamma_l'}{2}\right)\Gamma_0^{\frac{-1}{2}}u_l = u_l'C_lu_l = \lambda_l.$$

Furthermore,

$$\mathbb{E}[\omega_t^{(l)}\omega_t^{(l)}] = u_l'\Gamma_0^{\frac{-1}{2}}\mathbb{E}[y_ty_t']\Gamma_0^{\frac{-1}{2}}u_l = u_l'\Gamma_0^{\frac{-1}{2}}\Gamma_0\Gamma_0^{\frac{-1}{2}}u_l = u_l'u_l = 1.$$

Therefore, since $\omega_t^{(l)}$ is stationary,

$$\rho(C_l) = |\lambda_l| = \frac{\mathbb{E}[\omega_t^{(l)}\omega_{t-l}^{(l)}]}{\mathbb{E}[\omega_t^{(l)}\omega_t^{(l)}]} \leq 1.$$

Next, toward a contradiction suppose that $\rho(C_1) = 1$. Then $\omega_t^{(1)}$ is perfectly correlated with $\omega_{t-1}^{(1)}$, and so, with $\omega_{t-l}^{(1)}$ for every l, contracting the fact that $\omega_t^{(1)}$ is stationary ergodic and non-degenerate. □

I can now prove the theorem.

Proof of Theorem 4. Define

$$C(a,\eta) \equiv \sum_{\tau=1}^{\infty} a^{\tau}\eta^{\tau-1}C_{\tau}. \tag{A.30}$$

Then

$$\lambda_{\max}(\Omega(a,\eta)) = -\frac{a^2(1-\eta)^2}{1-a^2\eta^2} + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\lambda_{\max}(C(a,\eta)), \tag{A.31}$$

where $\lambda_{\max}(C(a,\eta))$ denotes the largest eigenvalue of $C(a,\eta)$. To simplify the exposition, I prove the result under the assumption that the largest eigenvalue of $C(a,\eta)$ is simple at the point (a^*,η^*) that maximizes $\lambda_{\max}(C(a,\eta))$.[46] The partial derivatives of $\lambda_{\max}(\Omega(a,\eta))$ with respect to a and η are given by

$$\begin{aligned}
\frac{\partial\lambda_{\max}(\Omega(a,\eta))}{\partial a} &= \frac{-2a(1-\eta)^2}{\left(1-a^2\eta^2\right)^2} - \frac{4a\eta(1-\eta)^2}{\left(1-a^2\eta^2\right)^2}\lambda_{\max}(C) \\
&\quad + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}u'_{\max}(C)\frac{\partial C}{\partial a}u_{\max}(C), &\text{(A.32)}\\
\frac{\partial\lambda_{\max}(\Omega(a,\eta))}{\partial \eta} &= \frac{2a^2(1-\eta)(1-a^2\eta)}{\left(1-a^2\eta^2\right)^2} - \frac{2\left(1+a^4\eta^2+a^2(1-4\eta+\eta^2)\right)}{\left(1-a^2\eta^2\right)^2}\lambda_{\max}(C) \\
&\quad + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}u'_{\max}(C)\frac{\partial C}{\partial \eta}u_{\max}(C), &\text{(A.33)}
\end{aligned}$$

where $u_{\max}(C)$ denotes the eigenvector of C with eigenvalue $\lambda_{\max}(C)$, normalized such that $u'_{\max}(C)u_{\max}(C) = 1$, and

$$\begin{aligned}
\frac{\partial C}{\partial a} &= \sum_{\tau=1}^{\infty} \tau a^{\tau-1}\eta^{\tau-1}C_{\tau}, \\
\frac{\partial C}{\partial \eta} &= \sum_{\tau=1}^{\infty} (\tau-1)a^{\tau}\eta^{\tau-2}C_{\tau}.
\end{aligned}$$

Note that

$$\eta u'_{\max}(C)\frac{\partial C}{\partial \eta}u_{\max}(C) + \lambda_{\max}(C) = a u'_{\max}(C)\frac{\partial C}{\partial a}u_{\max}(C). \tag{A.34}$$

for any a and η.

Let a^* and η^* be scalars in the $[-1,1]$ and $[0,1]$ intervals, respectively, that maximize $\lambda_{\max}(\Omega(a,\eta))$. I separately consider the cases $\eta^* = 1$ and $\eta^* < 1$. If $\eta^* = 1$, then $B = 0$ in the representation in the proof of Theorem 3 and so the pseudo-true 1-state model is i.i.d.

In the rest of the proof, I assume that $\eta^* < 1$ and show that this implies $a^* \neq 1$—by a similar argument $a^* \neq -1$. Toward a contradiction, suppose $a^* = 1$. Setting $a = 1$ in the partial derivatives of $\lambda_{\max}(\Omega(a,\eta))$, I get

$$\left.\frac{\partial\lambda_{\max}(\Omega(a,\eta))}{\partial a}\right|_{a=1} = \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left[-1-2\eta\lambda_{\max}(C) + (1-\eta^2)u'_{\max}(C)\frac{\partial C}{\partial a}u_{\max}(C)\right],$$

[46] The argument can easily be adapted to the case where the largest eigenvalue of $C(a^*,\eta^*)$ is not necessarily simple by replacing the gradient of $\lambda_{\max}(C(a,\eta))$ with its subdifferential and replacing the usual first-order optimality condition with the condition that the zero vector belongs to the subdifferential.

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial \eta}\right|_{a=1} = \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left[1 - 2\lambda_{\max}(C) + (1-\eta^2)u'_{\max}(C)\frac{\partial C}{\partial \eta}u_{\max}(C)\right],$$

where $C = C(1,\eta)$ and its partial derivatives are computed at $a = 1$. Multiplying the second equation above by η and subtracting from it the first equation, I get

$$\begin{aligned}
&\eta\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial \eta}\right|_{a=1} - \left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial a}\right|_{a=1} \\
&= \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left[1+\eta+(1-\eta^2)\left(\eta u'_{\max}(C)\frac{\partial C}{\partial \eta}u_{\max}(C) - u'_{\max}(C)\frac{\partial C}{\partial a}u_{\max}(C)\right)\right] \\
&= \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left[1+\eta-(1-\eta^2)\lambda_{\max}(C)\right],
\end{aligned}$$

where in the second equality I am using identity (A.34). Therefore,

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial a}\right|_{a=1} = \eta\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial \eta}\right|_{a=1} - \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left(1+\eta-(1-\eta^2)\lambda_{\max}(C(1,\eta))\right).$$

Note that

$$\lambda_{\max}(C(1,\eta)) \leq \sum_{\tau=1}^{\infty}\eta^{\tau-1}\lambda_{\max}(C_\tau) < \sum_{\tau=1}^{\infty}\eta^{\tau-1} = \frac{1}{1-\eta},$$

where the inequality is by Lemma A.2. Therefore,

$$-\frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}\left(1+\eta-(1-\eta^2)\lambda_{\max}(C(1,\eta))\right) < \frac{2(1-\eta)^2}{\left(1-\eta^2\right)^2}(1+\eta-1-\eta) = 0.$$

On the other hand, by the optimality of $a^* = 1$ and $\eta^* < 1$,

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial \eta}\right|_{a^*=1,\eta=\eta^*} \leq 0.$$

Thus,

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial a}\right|_{a^*=1,\eta=\eta^*} < 0,$$

a contradiction to the assumption of optimality of $a^* = 1$ and $\eta^* < 1$. This proves that $a^* < 1$ and establishes the stationarity of the 1-state model with $a = a^*$ and $\eta = \eta^*$. □

Proof of Theorem 5

Let λ denote the eigenvalue of C_1 largest in magnitude.[47] If $\rho(C_1) = 0$, then $\rho(C_\tau) = 0$ and so $\rho(C_\tau) = 0$ for all $\tau \geq 1$. Since C_τ are symmetric matrices, this implies that $C_\tau = 0$ for all $\tau \geq 1$. Therefore,

$$\lambda_{\max}(\Omega(a,\eta)) = -\frac{a^2(1-\eta)^2}{1-a^2\eta^2}.$$

[47] The proof does not assume that λ is unique. I allow for the possibility that λ and $\lambda' = -\lambda$ are both eigenvalues of C_1 and $|\lambda| = |\lambda'| = \rho(C_1)$.

The above expression is maximized by setting $(1-\eta)a = 0$. Therefore, by Theorem 3, for any pseudo-true 1-state model, $E_t^\theta[y_{t+s}] = a^s(1-\eta)qp'\sum_{\tau=0}^{\infty} a^\tau \eta^\tau y_{t-\tau} = 0$. On the other hand, if $\rho(C_1) = 0$, then $\lambda = 0$. Therefore, the theorem holds in the case $\rho(C_1) = 0$.

For the rest of the proof, I assume $\rho(C_1) > 0$. Define

$$\begin{aligned}\overline{f}(a,\eta) &\equiv -\frac{a^2(1-\eta)^2}{1-a^2\eta^2} + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\sum_{\tau=1}^{\infty}|a|^\tau\eta^{\tau-1}\rho(C_1)^\tau \\ &= -\frac{a^2(1-\eta)^2}{1-a^2\eta^2} + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\frac{|a|\rho(C_1)}{1-\eta|a|\rho(C_1)},\end{aligned}$$

where in the second equality I am using the fact that $\rho(C_\tau) < 1$. The function $\overline{f}(a,\eta)$ has two maximizers given by $(\overline{a}^*,\overline{\eta}^*) = (-\rho(C_1),0)$ and $(\overline{a}^*,\overline{\eta}^*) = (\rho(C_1),0)$ with the maximum given by $\overline{f}^* = \rho(C_1)^2$. I establish the theorem by showing that $\lambda_{\max}(\Omega(a,\eta)) \le \overline{f}(a,\eta)$ for all a and η, $\lambda_{\max}(\Omega(\lambda,0)) = \overline{f}(\lambda,0) = \overline{f}^*$, and $\lambda_{\max}(\Omega(-\lambda,0)) \le \overline{f}(-\lambda,0) = \overline{f}^*$ with the inequality strict if $-\lambda$ is not an eigenvalue of C_1. This establishes that $(a^*,\eta^*) = (\lambda,0)$ is the unique maximizer of $\lambda_{\max}(\Omega(a,\eta))$ if $-\lambda$ is not eigenvalue of C_1 and that $(a^*,\eta^*) = (\lambda,0)$ and $(a^*,\eta^*) = (-\lambda,0)$ are the only maximizers of $\lambda_{\max}(\Omega(a,\eta))$ if λ and $-\lambda$ are both eigenvalues of C_1.

As the first step in doing so, I show that for all a and t,

$$\lambda_{\max}\left(a^\tau C_\tau\right) \le |a|^\tau\rho(C_1)^\tau,$$

by considering four disjoint cases: If $a \le 0$ and $\lambda_{\min}(C_\tau) \le 0$, then

$$\lambda_{\max}\left(a^\tau C_\tau\right) = a^\tau\lambda_{\min}(C_\tau) = |a|^\tau\left|\lambda_{\min}(C_\tau)\right| \le |a|^\tau\rho(C_1)^\tau.$$

If $a \le 0$ and $\lambda_{\min}(C_\tau) > 0$, then

$$\lambda_{\max}\left(a^\tau C_\tau\right) = a^\tau\lambda_{\min}(C_\tau) \le 0 \le |a|^\tau\rho(C_1)^\tau.$$

If $a > 0$ and $\lambda_{\max}(C_\tau) \le 0$, then

$$\lambda_{\max}\left(a^\tau C_\tau\right) = a^\tau\lambda_{\max}(C_\tau) \le 0 \le |a|^\tau\rho(C_1)^\tau.$$

Finally, if $a > 0$ and $\lambda_{\max}(C_\tau) > 0$, then

$$\lambda_{\max}\left(a^\tau C_\tau\right) = a^\tau\lambda_{\max}(C_\tau) = |a|^\tau\left|\lambda_{\max}(C_\tau)\right| \le |a|^\tau\rho(C_1)^\tau.$$

So $\lambda_{\max}\left(a^\tau C_\tau\right) \le |a|^\tau\rho(C_1)^\tau$ regardless of the values of a and the eigenvalues of C_1. Therefore,

$$\lambda_{\max}\left(\sum_{\tau=1}^{\infty}a^\tau\eta^{\tau-1}C_\tau\right) \le \sum_{\tau=1}^{\infty}\eta^{\tau-1}\lambda_{\max}\left(a^\tau C_\tau\right) \le \sum_{\tau=1}^{\infty}\eta^{\tau-1}|a|^\tau\rho(C_1)^\tau = \frac{|a|\rho(C_1)}{1-\eta|a|\rho(C_1)},$$

where the first inequality is using the fact that $\eta^{\tau-1} \ge 0$ for all $\tau \ge 1$ and Weyl's inequality. Consequently,

$$\lambda_{\max}(\Omega(a,\eta)) \le \overline{f}(a,\eta) < \rho(C_1)^2$$

for any a, η such that $(|a|, \eta) \neq (\rho(C_1), 0)$.

I finish the proof by arguing that $\lambda_{\max}(\Omega(\lambda, 0)) = \rho(C_1)^2$ and $\lambda_{\max}(\Omega(-\lambda, 0)) \leq \overline{f}(-\lambda, 0) = \rho(C_1)^2$ with the inequality strict if $-\lambda$ is not an eigenvalue of C_1. To see this first note that

$$\lambda_{\max}(\Omega(a,0)) = -a^2 + 2\lambda_{\max}(aC_1) = \begin{cases} -a^2 + 2a\lambda_{\min}(C_1) & \text{if} \quad a < 0, \\ -a^2 + 2a\lambda_{\max}(C_1) & \text{if} \quad a \geq 0. \end{cases}$$

Thus,

$$\max_{a\in[-1,1]} \lambda_{\max}(\Omega(a,0)) = \begin{cases} \lambda_{\min}(C_1)^2 & \text{if} \quad |\lambda_{\min}(C_1)| > \lambda_{\max}(C_1), \\ \lambda_{\max}(C_1)^2 & \text{if} \quad |\lambda_{\min}(C_1)| \leq \lambda_{\max}(C_1), \end{cases}$$

and

$$\arg\max_{a\in[-1,1]} \lambda_{\max}(\Omega(a,0)) = \begin{cases} \{\lambda_{\min}(C_1)\} & \text{if} \quad |\lambda_{\min}(C_1)| > \lambda_{\max}(C_1), \\ \{\lambda_{\min}(C_1), \lambda_{\max}(C_1)\} & \text{if} \quad |\lambda_{\min}(C_1)| = \lambda_{\max}(C_1), \\ \{\lambda_{\max}(C_1)\} & \text{if} \quad |\lambda_{\min}(C_1)| < \lambda_{\max}(C_1). \end{cases}$$

Since C_1 is a symmetric matrix, the eigenvalues of C_1 are all real and so

$$\rho(C_1) = \begin{cases} -\lambda_{\min}(C_1) & \text{if} \quad |\lambda_{\min}(C_1)| > \lambda_{\max}(C_1), \\ \lambda_{\max}(C_1) & \text{if} \quad |\lambda_{\min}(C_1)| \leq \lambda_{\max}(C_1). \end{cases}$$

This establishes that, in any 1-state constrained rational model, $\eta = 0$, $a = \lambda$, and

$$\Omega(a,\eta) = -\lambda^2 I + 2\lambda C_1.$$

By Theorem 3, u is an eigenvector of $\Omega(a,\eta)$ with eigenvalue $\lambda_{\max}(\Omega(a,\eta)) = \lambda^2$ and $u'u = 1$. Therefore, u is also an eigenvector of C_1, but with eigenvalue λ. This completes the proof of the theorem. □

Proof of Theorem 6

I first prove a useful lemma, which offers a canonical representation of matrices C_l:[48]

Lemma A.3. *Suppose $\{C_l\}_l$ are the autocorrelation matrices of an n-dimensional stationary ergodic process that can be represented as in (6) with $f_t \in \mathbb{R}^m$. There exists a convergent $m \times m$ matrix $\mathbb{F}$ with $\|\mathbb{F}\|_2 \leq 1$, and a semi-orthogonal $m \times n$ matrix $\mathbb{H}$ such that*

$$C_l = \mathbb{H}' \left(\frac{\mathbb{F}^l + \mathbb{F}'^l}{2} \right) \mathbb{H}. \tag{A.35}$$

Conversely, for any positive integers $m \geq n$, $m \times m$ convergent matrix $\mathbb{F}$ with $\|\mathbb{F}\|_2 \leq 1$, and semi-orthogonal $m \times n$ matrix $\mathbb{H}$, there exists an n-dimensional stationary ergodic process with autocorrelation matrices $\{C_l\}_l$ of the form (A.35), which can be represented as in (6).[49]

[48] Versions of this result have previously appeared in the control and time-series literatures. For early examples, see Ho and Kálmán (1966) and Akaike (1975).

[49] Matrix $\mathbb{H} \in \mathbb{R}^{m\times n}$ is semi-orthogonal if $\mathbb{H}'\mathbb{H} = I$, where I denotes the $n \times n$ identity matrix.

Proof. Recall that I have assumed (without loss of generality) that the true process is non-degenerate, i.e., $\mathbb{E}[y_t y_t']$ is invertible. Invertibility of $\mathbb{E}[y_t y_t']$ requires $m \geq n$, an assumption I maintain throughout the first part of the proof. Given representation (6), the autocovariance matrices are given by

$$\Gamma_l = \mathbb{E}\left[y_t y_{t-l}'\right] = H'F^l \mathbb{E}\left[f_{t-l} f_{t-l}'\right] H = H'\mathbb{F}^l V H,$$

where $V \equiv \mathbb{E}\left[f_t f_t'\right]$ is the unique solution to the following discrete-time Lyapunov equation:

$$V = FVF' + \Sigma. \tag{A.36}$$

Therefore,

$$C_l = (H'VH)^{\frac{-1}{2}} \left(\frac{H'F^l VH + H'VF'^l H}{2}\right) (H'VH)^{\frac{-1}{2}}.$$

Matrix V is positive semidefinite; it is positive definite if the representation in (6) is minimal.[50] Without loss of generality, I assume that that is the case. Define

$$\mathbb{H}' \equiv (H'VH)^{\frac{-1}{2}} H'V^{\frac{1}{2}},$$
$$\mathbb{F} \equiv V^{\frac{-1}{2}} F V^{\frac{1}{2}}.$$

Then

$$C_l = \mathbb{H}' \left(\frac{\mathbb{F}^l + \mathbb{F}'^l}{2}\right) \mathbb{H}. \tag{A.37}$$

Note that since F is a convergent matrix, so is $\mathbb{F}$. Substituting $\mathbb{F} = V^{\frac{-1}{2}} F V^{\frac{1}{2}}$ in equation (A.36), I get

$$1 - \mathbb{F}\mathbb{F}' = V^{\frac{-1}{2}} \Sigma V^{\frac{-1}{2}}.$$

Therefore, since Σ is positive semidefinite, the spectral radius of $\mathbb{F}\mathbb{F}'$ is weakly smaller than one. This implies that $\|\mathbb{F}\|_2 \leq 1$. On the other hand,

$$\mathbb{H}'\mathbb{H} = (H'VH)^{\frac{-1}{2}} H'VH \, (H'VH)^{\frac{-1}{2}} = I.$$

That is, $\mathbb{H}$ is a (full-rank) semi-orthogonal matrix. This proves the first part of the theorem.

I next argue that given a convergent matrix $\hat{\mathbb{F}} \in \mathbb{R}^{m\times m}$ with $\|\hat{\mathbb{F}}\|_2 \leq 1$ and a semi-orthogonal matrix $\hat{\mathbb{H}} \in \mathbb{R}^{m\times n}$ with $m \geq n$, there exists a stationary ergodic process such that the corresponding autocorrelation matrices are given by (A.37) with $\mathbb{F} = \hat{\mathbb{F}}$ and $\mathbb{H} = \hat{\mathbb{H}}$. Given any such $\hat{\mathbb{F}}$ and $\hat{\mathbb{H}}$, let $F = \hat{\mathbb{F}}$, $H = \hat{\mathbb{H}}$, and $\Sigma = I - \hat{\mathbb{F}}\hat{\mathbb{F}}'$. The solution to the Lyapunov equation (A.36) is then given by $V = I$. Therefore, $\mathbb{F} = F = \hat{\mathbb{F}}$ and $\mathbb{H} = \hat{\mathbb{H}}(\hat{\mathbb{H}}'\hat{\mathbb{H}})^{\frac{-1}{2}} = \hat{\mathbb{H}}$, where in the last equality I am using the assumption of semi-orthogonality of $\hat{\mathbb{H}}$. By construction, then the autocorrelation matrices of the process (6) with matrices F, H, and Σ as above are given by (A.37) with $\mathbb{F} = \hat{\mathbb{F}}$ and $\mathbb{H} = \hat{\mathbb{H}}$. □

[50] See, for instance, Akaike (1975).

Proof of Theorem 6. I assume without loss of generality that the representation in (6) in minimal. By Lemma A.3 then,

$$C_l = \mathbb{H}'\left(\frac{\mathbb{F}^l + \mathbb{F}'^l}{2}\right)\mathbb{H},$$

where $\mathbb{H}' \equiv (H'VH)^{\frac{-1}{2}} H'V^{\frac{1}{2}}$, $\mathbb{F} \equiv V^{\frac{-1}{2}} F V^{\frac{1}{2}}$, and $V \equiv \mathbb{E}\left[f_t f_t'\right]$ is the variance-covariance of f_t. Note that since the variance-covariance of f_t is normalized to be the identity matrix, $V = I$, $\mathbb{F} = F$, and $\mathbb{H} = H$. Recall that vector y_t does not contain any redundant observables (which are linear combinations of other observables). This assumption, together with the assumption that H is a rank-m matrix, ensures that H is an invertible $m \times m$ matrix. Therefore, by Lemma A.3, $\mathbb{H} = H$ is an orthogonal matrix. Thus,

$$\rho(C_l) = \rho\left(\mathbb{H}'\left(\frac{\mathbb{F}^l + \mathbb{F}'^l}{2}\right)\mathbb{H}\right) = \rho\left(\frac{\mathbb{F}^l + \mathbb{F}'^l}{2}\right) = \rho\left(\frac{F^l + F'^l}{2}\right) \tag{A.38}$$

for all l. But since the spectral radius of a symmetric matrix equals its spectral norm,

$$\rho\left(\frac{F^l + F'^l}{2}\right) = \left\|\frac{F^l + F'^l}{2}\right\|_2 \leq \frac{1}{2}\left\|F^l\right\|_2 + \frac{1}{2}\left\|F'^l\right\|_2 = \left\|F^l\right\|_2 \leq \|F\|_2^l. \tag{A.39}$$

Therefore,

$$\rho(C_l) \leq \|F\|_2^l.$$

On the other hand, by equations (A.38) and (A.39),

$$\rho(C_1) = \left\|\frac{F + F'}{2}\right\|_2 = \|F\|_2,$$

where the second equality is by assumption. Thus,

$$\rho(C_l) \leq \|F\|_2^l = \rho(C_1)^l,$$

and the process is exponentially ergodic. □

Proof of Proposition 1

I first state and prove a useful lemma:

Lemma A.4. *Suppose C_1 has a unique and simple eigenvalue λ with $|\lambda| = \rho(C_1) > 0$ and let u denote the corresponding eigenvector normalized to have $u'u = 1$.*[51] *If $u'C_2u > \rho(C_1)^2$, then the agents' forecasts in any pseudo-true 1-state model are given by (4) with a tuple (a^*, η^*, p^*, q^*) such that $\eta^* > 0$.*

Proof. Define $C(a, \eta)$ as in the proof of Theorem 4. As in the proof of Theorem 4, I present the argument under the assumption that the largest eigenvalue of $C(a, \eta)$ is simple at the point (a^*, η^*)

[51] The assumption that λ is unique and simple is not necessary for the result. The result generalizes to arbitrary matrices C_1 with $\rho(C_1) \neq 0$ by replacing $u'C_2u$ with the maximum of $u'C_2u$ over all unit-norm eigenvectors u of C_1 with eigenvalues λ such that $|\lambda| = \rho(C_1)$.

that maximizes $\lambda_{\max}(C(a,\eta))$.[52] I start by proposing a candidate solution to problem of maximizing $\lambda_{\max}(\Omega(a,\eta))$ at which $\eta = 0$ and argue that the candidate does not satisfy the necessary first-order optimality conditions. Setting $\eta = 0$ in equations (A.32) and (A.33), I get

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial a}\right|_{\eta=0} = -2a + 2u'_{\max}(aC_1)C_1 u_{\max}(aC_1),$$
$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta)))}{\partial \eta}\right|_{\eta=0} = 2a^2 - 2(1+a^2)\lambda_{\max}(aC_1) + 2a^2 u'_{\max}(aC_1)C_2 u_{\max}(aC_1),$$

where I am using the fact that $C = aC_1$ when $\eta = 0$. Any solution to $\partial\lambda_{\max}(\Omega(a,\eta))/\partial a|_{\eta=0} = 0$ satisfies $a = \lambda$, where $\lambda = \lambda_{\min}(C_1)$ if $\lambda_{\max}(C_1) \le 0$, $\lambda = \lambda_{\max}(C_1)$ if $\lambda_{\min}(C_1) \ge 0$, and $\lambda \in \{\lambda_{\max}(C_1), \lambda_{\min}(C_1)\}$ otherwise. Evaluating $\lambda_{\max}(\Omega(a,\eta))$ at $a = \lambda$ and $\eta = 0$, I get $\lambda_{\max}(\Omega(\lambda,0)) = \lambda^2$. Therefore, for the solution $(a,\eta) = (\lambda,0)$ to the first-order condition $\partial\lambda_{\max}(\Omega(a,\eta))/\partial a = 0$ to be a maximizer of $\lambda_{\max}(\Omega(a,\eta))$, it must be the case that λ is the eigenvalue of C_1 largest in magnitude and $u = u_{\max}(aC_1)$ is a corresponding eigenvector normalized such that $u'u = 1$. Substituting in the expression for $\partial\lambda_{\max}(\Omega(a,\eta))/\partial\eta|_{\eta=0}$, I get

$$\left.\frac{\partial \lambda_{\max}(\Omega(a,\eta))}{\partial \eta}\right|_{a=\lambda,\eta=0} = 2\rho(C_1)^2\left(u'C_2u - \rho(C_1)^2\right) > 0,$$

where the inequality follows the assumption that $u'C_2u > \rho(C_1)^2$. This implies that the pair $\eta = 0$ and $a = \lambda$ does not constitute a local maximizer of $\lambda_{\max}(\Omega(a,\eta))$. Since this pair is the only candidate with $\eta = 0$ that satisfies the first-order conditions, in any pseudo-true 1-state model, $\eta > 0$. This establishes the lemma. $\square$

Proof of Proposition 1. Let σ^2 denote the variance of y_t. By the argument in the proof of Lemma A.3, the lag-l autocorrelation of y_t is given by

$$C_l = H'\left(\frac{F^l + F'^l}{2}\right)H,$$

where $F = V^{\frac{-1}{2}}\mathbb{F}V^{\frac{1}{2}}$, $H' = (\mathbb{H}'V\mathbb{H})^{\frac{-1}{2}}\mathbb{H}'V^{\frac{1}{2}}$, and V is the solution to the discrete-time Lyapunov equation (A.36). Since $\mathbb{F}$ and Σ are diagonal matrices, so is V. Therefore, $F = \mathbb{F}$. On the other hand, by Lemma A.3, H is a semi-orthogonal matrix. Therefore, $H'H = 1$, and so,

$$C_l = \sum_{i=1}^{m} w_i \alpha_i^l,$$

where $w_i = H_i^2 \ge 0$, $\sum_{i=1}^m w_i = 1$, and α_i is the ith diagonal element of $\mathbb{F}$. That is, $C_l^{\frac{1}{l}}$ is equal to the weighted l-norm of the vector $(\alpha_1,\ldots,\alpha_m)$ with weights $w = (w_1,\ldots,w_m)$.

Since the representation in (6) is minimal, $w_i > 0$ for all i, and all α_i are distinct. If that were not the case, there would exist some $\tilde{m} < m$ such that $C_l = \sum_{i=1}^{\tilde{m}} \tilde{w}_i\tilde{\alpha}_i^l$ for some non-negative weights

[52] See footnote 46 for how the argument can be generalized.

$\tilde{w}_i$ that sum up to one and some $\tilde{\alpha}_i \in (-1,1)$. Consider the process $\tilde{\mathbb{P}}$ represented as in (6) with $\mathbb{F} = \text{diag}(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_{\tilde{m}})$, $\epsilon_t \sim \mathcal{N}(0, \Sigma)$, $\Sigma = I - \mathbb{F}\mathbb{F}'$, and $\mathbb{H} = \sigma\, \text{diag}(\sqrt{\tilde{w}_1}, \ldots, \sqrt{\tilde{w}_{\tilde{m}}})$. By the argument in the proof of Lemma A.3, $\tilde{\mathbb{P}}$ has the same autocorrelation matrices as $\mathbb{P}$. Moreover, both $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are mean-zero and normal and both have variance σ^2. Therefore, $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are observationally equivalent, a contradiction to the assumption that the representation I started with was minimal.

Next, note that, by the generalized mean inequality, $C_l^{\frac{1}{l}} > C_1$ for all $l \geq 2$, where the strictness of the inequality follows the facts that $w_i > 0$ for all i and all α_i are distinct. In particular, $u'C_2u = C_2 > C_1^2 = \rho(C_1)^2$, where I am using the fact that y_t is a scalar. Thus, by Lemma A.4, $\eta^* > 0$. To see why $\eta^* < 1$, recall that by Theorem 3, the (a^*, η^*) pair maximizes

$$\Omega(a,\eta) = -\frac{a^2(1-\eta)^2}{1-a^2\eta^2} + \frac{2(1-\eta)(1-a^2\eta)}{1-a^2\eta^2}\sum_{\tau=1}^{\infty} a^\tau \eta^{\tau-1} C_\tau.$$

But $\Omega(a,1) = 0 < C_1^2 = \Omega(C_1, 0)$. Therefore, $\eta^* = 1$ cannot be part of the description of a pseudo-true 1-state model. Finally, $a^* \in (1,1)$ by Theorem 4. The proposition then follows Theorem 3 by noting that $q^*p^{*\prime} = 1$ whenever y_t is a scalar. □

Proof of Theorem 7

I first find a transformation $\widetilde{y}_t = Ty_t$ of the vector of observables such that T is invertible and $\widetilde{\Gamma}_0^{\frac{-1}{2}}\widetilde{\Gamma}_1\widetilde{\Gamma}_0^{\frac{-1}{2}}$ is diagonal. Since matrices Γ_0 and Γ_1 are both symmetric and Γ_0 is non-singular, Γ_0 and Γ_1 can be diagonalized simultaneously by a real congruence transformation. Note that since Γ_1 is symmetric,

$$\Gamma_0^{\frac{-1}{2}}\Gamma_1\Gamma_0^{\frac{-1}{2}} = \frac{1}{2}\Gamma_0^{\frac{-1}{2}}\left(\Gamma_1 + \Gamma_1'\right)\Gamma_0^{\frac{-1}{2}} = C_1$$

is symmetric. Therefore, there exists a diagonal matrix Λ, with the eigenvalues of C_1 as its diagonal elements, and an orthogonal matrix U such that $C_1 = U\Lambda U'$. Define

$$T \equiv U'\Gamma_0^{\frac{-1}{2}}.$$

It is easy to verify that $T\Gamma_0T' = I$ and $T\Gamma_1T' = \Lambda$. The autocovariance matrices of $\widetilde{y}_t \equiv Ty_t$ are given by $\widetilde{\Gamma}_l \equiv \mathbb{E}[\widetilde{y}_t\widetilde{y}_{t-\tau}'] = T\Gamma_lT'$. In particular, $\widetilde{\Gamma}_0 = I$, $\widetilde{\Gamma}_1 = \Lambda$, and so $\widetilde{\Gamma}_0^{\frac{-1}{2}}\widetilde{\Gamma}_1\widetilde{\Gamma}_0^{\frac{-1}{2}} = \Lambda$.

I next find the pseudo-true m.i.o. d-state models given the vector of observables $\widetilde{y}_t$ and then transform the models back using the linear invariance result to find the pseudo-true m.i.o. d-state models given observables y_t. By Lemma A.1, the KLDR and the agents' forecasts can be represented in terms of matrices $\widetilde{M}$, $\widetilde{N}$, and $\widetilde{D}$ of the transformed model as in (A.13) and (A.19). Let $S \equiv \widetilde{\Gamma}_0^{\frac{1}{2}}\widetilde{N} = \widetilde{N}$ and use the restriction to the set of m.i.o. models to set $\widetilde{D} = D = (I\ \ \mathbf{0})'$. The expression for the KLDR in (A.13) then simplifies to

$$\text{KLDR}(\theta) = -\frac{1}{2}\log\det(SS') + \frac{1}{2}\text{tr}(S'S) - \text{tr}\left(\widetilde{M}D'S'\Lambda SD\right) + \frac{1}{2}\text{tr}\left(\widetilde{M}D'S'SD\widetilde{M}'\right) + \text{constant}.$$

I begin by ignoring the constraint that $\widetilde{M}$ is a convergent matrix and showing that the solution to the relaxed problem is also a solution to the original problem. For the relaxed problem, the necessary first-order optimality conditions with respect to S and $\widetilde{M}$ are given by

$$I = S'S - S'\Lambda'SD\widetilde{M}D' - S'\Lambda SD\widetilde{M}'D' + S'SD\widetilde{M}'\widetilde{M}D', \tag{A.40}$$

$$D'S'\Lambda SD = \widetilde{M}D'S'SD. \tag{A.41}$$

I proceed by first characterizing the set of all solutions for S and $\widetilde{M}$ to these optimality conditions. Given any such solution, the KLDR is given by $-\log\det(SS')/2+\text{constant}$. Therefore, if there are multiple solutions to equations (A.40) and (A.41), the optimal solution is the one with the largest value of $\log\det(SS')$.

In the next step, I write $S = (S_1\ \ S_2)$ where $S_1 \in \mathbb{R}^{n\times d}$ and $S_2 \in \mathbb{R}^{n\times(n-d)}$. Equation (A.40) then can be written as

$$\begin{aligned}\begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & I\end{pmatrix} &= \begin{pmatrix} S_1'S_1 & S_1'S_2 \\ S_2'S_1 & S_2'S_2\end{pmatrix} - \begin{pmatrix} S_1'\Lambda'S_1 & S_1'\Lambda'S_2 \\ S_2'\Lambda'S_1 & S_2'\Lambda'S_2\end{pmatrix}\begin{pmatrix} \widetilde{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \begin{pmatrix} S_1'\Lambda S_1 & S_1'\Lambda S_2 \\ S_2'\Lambda S_1 & S_2'\Lambda S_2\end{pmatrix}\begin{pmatrix} \widetilde{M}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix} \\ &\quad + \begin{pmatrix} S_1'S_1 & S_1'S_2 \\ S_2'S_1 & S_2'S_2\end{pmatrix}\begin{pmatrix} \widetilde{M}'\widetilde{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{pmatrix} \\ &= \begin{pmatrix} S_1'S_1 & S_1'S_2 \\ S_2'S_1 & S_2'S_2\end{pmatrix} - \begin{pmatrix} S_1'\Lambda'S_1\widetilde{M} & \mathbf{0} \\ S_2'\Lambda'S_1\widetilde{M} & \mathbf{0}\end{pmatrix} - \begin{pmatrix} S_1'\Lambda S_1\widetilde{M}' & \mathbf{0} \\ S_2'\Lambda S_1\widetilde{M}' & \mathbf{0}\end{pmatrix} + \begin{pmatrix} S_1'S_1\widetilde{M}'\widetilde{M} & \mathbf{0} \\ S_2'S_1\widetilde{M}'\widetilde{M} & \mathbf{0}\end{pmatrix}.\end{aligned}$$

Therefore,

$$S_1'S_1 - S_1'\Lambda'S_1\widetilde{M} - S_1'\Lambda S_1\widetilde{M}' + S_1'S_1\widetilde{M}'\widetilde{M} = I, \tag{A.42}$$

$$S_2'S_1 = \mathbf{0}, \tag{A.43}$$

$$S_2'\Lambda'S_1\widetilde{M} + S_2'\Lambda S_1\widetilde{M}' = \mathbf{0}, \tag{A.44}$$

$$S_2'S_2 = I. \tag{A.45}$$

Likewise, equation (A.41) can be written as

$$S_1'\Lambda S_1 = \widetilde{M}S_1'S_1. \tag{A.46}$$

Equation (A.45) implies that S_2 is a full-rank matrix. On the other hand, since $S'S = \begin{pmatrix} S_1'S_1 & \mathbf{0} \\ \mathbf{0} & I\end{pmatrix}$ is invertible, S_1 is also a full-rank matrix. Equation (A.43) then implies that the column space (or range) of S_1 is the same as the null space (or kernel) of S_2', and equation (A.44) implies that the column space of $\Lambda'S_1\widetilde{M} + \Lambda S_1\widetilde{M}'$ is a subspace of the null space of S_2'. Therefore,

$$\Lambda'S_1\widetilde{M} + \Lambda S_1\widetilde{M}' = S_1 Y \tag{A.47}$$

for some matrix Y.

Define $X \equiv S_1'S_1$ and $Z \equiv S_1'\Lambda S_1$. Since S_1 is full rank, X is invertible. Left-multiplying equation (A.47) by S_1', I get

$$Z'\widetilde{M} + Z\widetilde{M}' = XY. \tag{A.48}$$

Meanwhile, equations (A.42) and (A.46) can be written as

$$X - Z'\widetilde{M} - Z\widetilde{M}' + X\widetilde{M}'\widetilde{M} = I, \tag{A.49}$$

$$Z = \widetilde{M}X. \tag{A.50}$$

Solving for $\widetilde{M}$ from (A.50) and substituting in (A.49), I get

$$X = I + ZX^{-1}Z'. \tag{A.51}$$

Therefore, X is a symmetric positive-definite matrix, with eigenvalues weakly larger than one. Combining (A.48) and (A.49), I get

$$X - XY + X\widetilde{M}'\widetilde{M} = I,$$

and so

$$Y = I + X^{-1}Z'ZX^{-1} - X^{-1}.$$

Because the eigenvalues of X are weakly larger than one, Y is a symmetric positive semi-definite matrix.

Let $\mathcal{K} \subseteq \mathbb{R}^d$ denote the (possibly empty) null space of $Z' = Z$. If $p \in \mathcal{K}$, then

$$Xp = p + ZX^{-1}Z'p = p,$$

$$X^{-1}p = p,$$

and

$$Yp = p + X^{-1}Z'ZX^{-1}p - X^{-1}p = X^{-1}Z'Zp = 0,$$

where in the last equality I am using the symmetry of Z. Therefore, $\mathcal{K}$ is an eigenspace of both X and X^{-1} with eigenvalue one and a subspace of the null space of Y. It is easy to show that, if p is an eigenvector of X with eigenvalue one, then it is also in the null spaces of $Z = Z'$ and Y, and that if p is in the null space of Y, it is also in the null space of $Z = Z'$ and an eigenvector of X with eigenvalue one. Therefore, $\mathcal{K}$ is identical to the null space of Y and the eigenspace of X with eigenvalue one. By a similar argument, equations (A.48) and (A.50) imply that $\mathcal{K}$ is also the null space of $\widetilde{M} + \widetilde{M}'$.

I next show that the column space of S_1 is spanned by d eigenvectors of Λ. Decompose $\mathbb{R}^d$ as a direct sum: $\mathbb{R}^d = \mathcal{K} \oplus \mathcal{R}$, where $\mathcal{K}$ is the kernel of $Z' = Z$ and $\mathcal{R}$ is the range of Z. Let d_K denote the dimension of $\mathcal{K}$ and d_R denote the dimension of $\mathcal{R}$. By the rank-nullity theorem, $d_K + d_R = d$. Any vector $p \in \mathbb{R}^d$ can be written as $p = p_K + p_R$, where $p_K \in \mathcal{K}$ and $p_R \in \mathcal{R}$. First, consider the set of $S_1 p_K$ for $p_K \in \mathcal{K}$. By construction, $p_K S_1' \Lambda S_1 p_K = 0$ for any $p_K \in \mathcal{K}$. Therefore, since S_1 is a full rank matrix, $S_1 p_K$ is in the null space of Λ. But Λ is a diagonal matrix with its null space the same as its (possibly empty) eigenspace with eigenvalue zero. Therefore, the set $\{S_1 p_K : p_K \in \mathcal{K}\}$ is a d_K-dimensional invariant subspace of Λ spanned by eigenvectors of Λ with eigenvalue zero. Next consider the set of $S_1 p_R$ for $p_R \in \mathcal{R}$. By the argument in the previous paragraph, matrices Y and $\widetilde{M} + \widetilde{M}'$ are invertible when restricted to the subspace $\mathcal{R}$ with the inverse over the subspace given by the Moore–Penrose pseudo-inverse. Therefore, equation (A.47) can be written as

$$\Lambda S_1 = S_1 Y \left(\widetilde{M} + \widetilde{M}'\right)^{\dagger}$$

over this subspace. But since Y and $(\widetilde{M} + \widetilde{M}')^\dagger$ are invertible when restricted to $\mathcal{R}$, the above equation implies that

$$\{S_1 p_R : p_R \in \mathcal{R}\} = \left\{S_1 Y \left(\widetilde{M} + \widetilde{M}'\right)^\dagger p_R : p_R \in \mathcal{R}\right\} = \{\Lambda S_1 p_R : p_R \in \mathcal{R}\}.$$

That is, the set $\{S_1 p_R : p_R \in \mathcal{R}\}$ is a d_R-dimensional invariant subspace of Λ. But the invariant subspaces of Λ are spanned by its eigenvectors.

The eigenvectors of Λ are the n coordinate vectors. Without loss of generality, assume that the elements of $\widetilde{y}_t$ are ordered such that the column space of S_1 is spanned by the first d coordinate vectors $e_1, \ldots, e_d$. In such coordinates, S_1 can be written as

$$S_1 = \begin{pmatrix} S_{11} \\ \mathbf{0} \end{pmatrix},$$

where $S_{11} \in \mathbb{R}^{d\times d}$ is an invertible matrix. Therefore,

$$X = S_1' S_1 = S_{11}' S_{11},$$
$$Z = S_{11}' \Lambda_1 S_{11},$$

where $\Lambda = \begin{pmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \Lambda_2 \end{pmatrix}$, and Λ_1 is a $d \times d$ diagonal matrix. Substituting in (A.51), I get

$$S_{11}' S_{11} = I + S_{11}' \Lambda_1 S_{11} \left(S_{11}' S_{11}\right)^{-1} S_{11}' \Lambda_1 S_{11} = I + S_{11}' \Lambda_1^2 S_{11},$$

and so,

$$S_{11}' \left(I - \Lambda_1^2\right) S_{11} = I.$$

Multiplying the above equation from left by S_{11} and from right by $S_{11}^{-1}(I - \Lambda_1^2)^{-1}$, I get

$$S_{11} S_{11}' = \left(I - \Lambda_1^2\right)^{-1}. \tag{A.52}$$

Therefore,

$$\log\det(SS') = \log\det(S'S) = \log\det\left(S_1' S_1\right) = \log\det\left(S_{11}' S_{11}\right) = \log\det\left(S_{11} S_{11}'\right) = -\log\det\left(I - \Lambda_1^2\right).$$

Thus, minimizing the KLDR requires the column space of S_1 to be spanned by eigenvectors $p_1, \ldots, p_d$ of Λ corresponding to the d eigenvalues of Λ largest in magnitude.

It only remains to compute matrices $\widetilde{A}, \widetilde{B}, \widetilde{Q}, \widetilde{R}$. Note that matrix S can be written as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ \mathbf{0} & S_{22} \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} = S'S = \begin{pmatrix} S_{11}' S_{11} & S_{11}' S_{12} \\ S_{12}' S_{11} & S_{12}' S_{12} + S_{22}' S_{22} \end{pmatrix}.$$

Since S_{11} is an invertible matrix, the above equation implies that $S_{12} = \mathbf{0}$ and $S_{22}' S_{22} = I$. Therefore, since S_{22} is a symmetric invertible matrix, $S_{22}' = S_{22}^{-1}$. I can now find $\widetilde{A}, \widetilde{B}, \widetilde{Q}, \widetilde{R}$:

$$\widetilde{A} = \widetilde{M} = S_{11}' \Lambda_1 {S_{11}'}^{-1},$$

$$\widetilde{B} = D'S^{-1}\widetilde{\Gamma}_0^{\frac{1}{2}} = \begin{pmatrix} S_{11}^{-1} & \mathbf{0} \end{pmatrix},$$
$$\widetilde{Q} = I - \widetilde{M}\,(I - D'D)\,\widetilde{M}' = I,$$
$$\widetilde{R} = S'^{-1}\,(I - DD')\,S^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}.$$

I can use the above expressions to compute the forecasts of $\widetilde{y}_t$. Equation (A.19) implies

$$E_t^\theta[\widetilde{y}_{t+s}] = \begin{pmatrix} S_{11}'^{-1} \\ \mathbf{0} \end{pmatrix} S_{11}'\Lambda_1^s S_{11}'^{-1}\,(S_{11} \quad \mathbf{0})\,\widetilde{y}_t = \begin{pmatrix} \Lambda_1^s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\widetilde{y}_t.$$

Using $\widetilde{y}_t = Ty_t$ for all t in the above equation, I get

$$E_t^\theta[y_{t+s}] = T^{-1}\begin{pmatrix} \Lambda_1^s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Ty_t = \Gamma_0^{\frac{1}{2}}U\begin{pmatrix} \Lambda_1^s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} U'\Gamma_0^{\frac{-1}{2}}y_t = \Gamma_0^{\frac{1}{2}}UD'U'C_1^sUDU'\Gamma_0^{\frac{-1}{2}}y_t.$$

Using the definition of U, I can simplify the above expression to

$$E_t^\theta[y_{t+s}] = \sum_{i=1}^{d}\Gamma_0^{\frac{1}{2}}u_i\lambda_i^s u_i'\Gamma_0^{\frac{-1}{2}}y_t,$$

where λ_i is the ith largest eigenvalue of C_1, and u_i is the corresponding eigenvector normalized such that $u_i'u_i = 1$. The theorem then follows the definitions of a_i^*, p_i^*, and q_i^*. □

Proof of Theorem 8

Setting $M = a$, $D = \sqrt{1-\eta}e_1$, and $N = \Gamma_0^{\frac{-1}{2}}S$ in equation (A.20), I get

$$\mathrm{Var}^\theta(y) = \Gamma_0^{\frac{1}{2}}\left[I + \frac{1}{1-a^2}\left[a^2(1-\eta)^2 - \left(1 - 2a^2\eta + a^2\eta^2\right)\lambda\right]uu'\right]\Gamma_0^{\frac{1}{2}},$$

where a, η, $\lambda = \lambda_{\max}(\Omega(a,\eta))$, and u are as in Theorem 3. Substituting for $\lambda_{\max}(\Omega(a,\eta))$ from equation (A.31) in the above equation, I get

$$\mathrm{Var}^*(y_t) = \Gamma_0^{\frac{1}{2}}\left[I + \frac{2(1-\eta)(1-a^2\eta)}{(1-a^2)(1-a^2\eta^2)}\left(a^2(1-\eta) - (1-2a^2\eta + a^2\eta^2)\lambda_{\max}(C)\right)uu'\right]\Gamma_0^{\frac{1}{2}}. \qquad \text{(A.53)}$$

Let a^* and η^* be scalars in the $[-1,1]$ and $[0,1]$ intervals, respectively, that maximize $\lambda_{\max}(\Omega(a,\eta))$. I separately consider the cases $\eta^* = 1$ and $\eta^* < 1$. If $\eta^* = 1$, then the right-hand side of equation (A.53) is equal to Γ_0.

Next suppose $\eta^* < 1$. By the argument in the proof of Theorem 4, the first-order optimality condition with respect to a must hold with equality at $a = a^*$ and $\eta = \eta^* < 1$. Setting $\partial\lambda_{\max}(\Omega(a,\eta))/\partial a = 0$ and multiplying both sides of the equation by a^*, I get

$$\begin{aligned}&\frac{2a^{*2}(1-\eta^*)^2}{\left(1-a^{*2}\eta^{*2}\right)^2} + \frac{4a^{*2}\eta^*(1-\eta^*)^2}{\left(1-a^{*2}\eta^{*2}\right)^2}\lambda_{\max}(C) \\ &\quad = \frac{2(1-\eta^*)(1-a^{*2}\eta^*)}{1-a^{*2}\eta^{*2}}\lambda_{\max}(C) + \frac{2(1-\eta^*)(1-a^{*2}\eta^*)}{1-a^{*2}\eta^{*2}}\eta^* u_{\max}'(C)\frac{\partial C}{\partial\eta}u_{\max}(C). \qquad \text{(A.54)}\end{aligned}$$

Setting $\eta^* = 0$ in the above equation, I get $a^{*2} = \lambda_{\max}(C)$. Setting $a^{*2} = \lambda_{\max}(C)$ in equation (A.53) then establishes that $\mathrm{Var}^*(y_t) = \Gamma_0$ in the case where $\eta^* = 0$.

Finally, I consider the case where $\eta^* \in (0,1)$. Then additionally the first-order optimality condition with respect to η must hold with equality. Setting $\partial\lambda_{\max}(\Omega(a,\eta))/\partial\eta = 0$, multiplying the resulting equation by η^*, solving for $\eta^* u'_{\max}(C)\frac{\partial C}{\partial \eta}u_{\max}(C)$, and substituting in equation (A.54), I get

$$\begin{aligned}
&\frac{2a^{*2}(1-\eta^*)^2}{\left(1-a^{*2}\eta^{*2}\right)^2} + \frac{4a^{*2}\eta^*(1-\eta^*)^2}{\left(1-a^{*2}\eta^{*2}\right)^2}\lambda_{\max}(C) \\
&\quad = \frac{2(1-\eta^*)(1-a^{*2}\eta^*)}{1-a^{*2}\eta^{*2}}\lambda_{\max}(C) - \frac{2a^{*2}\eta^*(1-\eta^*)(1-a^{*2}\eta^*)}{\left(1-a^{*2}\eta^{*2}\right)^2} \\
&\quad\quad + \frac{2\eta^*\left(1+a^{*4}\eta^{*2}+a^{*2}(1-4\eta^*+\eta^{*2})\right)}{\left(1-a^{*2}\eta^{*2}\right)^2}\lambda_{\max}(C).
\end{aligned}$$

Simplifying the above expression leads to

$$a^{*2}(1-\eta^*) = \left(1-2a^{*2}\eta^* + a^{*2}\eta^{*2}\right)\lambda_{\max}(C).$$

Combining the above identity with equation (A.53) implies that $\mathrm{Var}^*(y_t) = \Gamma_0$ and finishes the proof of the theorem. □

Proof of Theorem 9

Define T as in the proof of Theorem 7 and let $\widetilde{y}_t = Ty_t$. Then by the argument in the proof of Theorem 7, $\widetilde{\Gamma}_0 = I$ and $\widetilde{\Gamma}_1 = \Lambda$, where Λ is a diagonal matrix with the eigenvalues of C_1 as its diagonals. Furthermore, any pseudo-true m.i.o. d-state model $\widetilde{\theta} = (\widetilde{A}, \widetilde{B}, \widetilde{Q}, \widetilde{R})$ given observable $\widetilde{y}_t$ satisfies

$$\begin{aligned}
\widetilde{A} &= S'_{11}\Lambda_1 {S'_{11}}^{-1}, \\
\widetilde{B} &= \begin{pmatrix} S_{11}^{-1} & \mathbf{0} \end{pmatrix}, \\
\widetilde{Q} &= I, \\
\widetilde{R} &= \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix},
\end{aligned}$$

where S_{11} is $d \times d$ matrix that satisfies $S_{11}S'_{11} = \left(I - \Lambda_1^2\right)^{-1}$, and Λ_1 is a $d \times d$ diagonal matrix containing the d largest eigenvalues of C_1. On the other hand, equation (A.20) implies

$$\begin{aligned}
\mathrm{Var}^\theta(\widetilde{y}) &= \widetilde{B}' \sum_{\tau=0}^{\infty} \widetilde{A}^\tau \widetilde{Q}\left(\widetilde{A}'\right)^\tau \widetilde{B} + \widetilde{R} = \begin{pmatrix} {S'_{11}}^{-1} \\ \mathbf{0} \end{pmatrix} \sum_{\tau=0}^{\infty} S'_{11}\Lambda_1^\tau \left(S_{11}S'_{11}\right)^{-1}\Lambda_1^\tau S_{11}\begin{pmatrix} S_{11}^{-1} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \\
&= \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} = I.
\end{aligned}$$

Therefore, $\mathrm{Var}^\theta(y) = T^{-1}\mathrm{Var}^\theta(\widetilde{y})T'^{-1} = (T'T)^{-1} = \Gamma_0$. □

Proof of Proposition 6

I first show that in a linear equilibrium r_t^n and μ_t can be written as linear functions of $\hat{x}_t$, $\hat{\pi}_t$, and $\hat{i}_t$. Suppose r_t^n and μ_t can be written as linear functions of $\hat{x}_t$, $\hat{\pi}_t$, and $\hat{i}_t$. Then by the linear invariance result, agents' forecasts are the same whether they observe vector $f_t \equiv (\hat{x}_t, \hat{\pi}_t, \hat{i}_t)'$ or vector y_t, consisting of all the observables. Furthermore, since shocks follow an exponentially ergodic process and f_t is an invertible linear transformation of the vector of shocks, f_t follows an exponentially ergodic process as well. Therefore, by the linear invariance result and Theorem 5,

$$E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s\left(\frac{1-\beta}{\beta}\hat{x}_{t+s}-\sigma\left(\hat{i}_{t+s}-r_{t+s}^n\right)-\frac{\sigma}{\beta}\hat{\pi}_{t+s}\right)\right]=\gamma_x\hat{z}_t, \tag{A.55}$$

$$E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s\left(\kappa\hat{x}_{t+s}+\frac{1-\delta}{\delta}\hat{\pi}_{t+s}+\mu_{t+s}\right)\right]=\gamma_\pi\hat{z}_t, \tag{A.56}$$

where γ_x and γ_π are constants that are to be determined in equilibrium, $\hat{z}_t = p'f_t$ is the agents' time-t estimate of the subjective state, and $p \equiv (p_x, p_\pi, p_i)'$ is the relative attention vector. Substituting in (17) and (18) and collecting terms, I get

$$\sigma r_t^n = \hat{x}_t + \sigma\hat{i}_t - \gamma_x\left(p_x\hat{x}_t + p_\pi\hat{\pi}_t + p_i\hat{i}_t\right), \tag{A.57}$$

$$\mu_t = \hat{\pi}_t - \kappa\hat{x}_t - \gamma_\pi\left(p_x\hat{x}_t + p_\pi\hat{\pi}_t + p_i\hat{i}_t\right). \tag{A.58}$$

These expressions verify my guess that r_t^n and μ_t can be written as linear functions of $\hat{x}_t$, $\hat{\pi}_t$, and $\hat{i}_t$.

I next find constants γ_x and γ_π. Using the linear invariance result to substitute for σr_{t+s}^n and μ_{t+s} from the above equations in (A.55) and (A.56) and using Theorem 5 to characterize the resulting subjective expectations, I get

$$E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s\left(\frac{1-\beta}{\beta}\hat{x}_{t+s}-\sigma\left(\hat{i}_{t+s}-r_{t+s}^n\right)-\frac{\sigma}{\beta}\hat{\pi}_{t+s}\right)\right]=\frac{a\left((1-\beta\gamma_x p_x)q_x-(\sigma+\beta\gamma_x p_\pi)q_\pi-\beta\gamma_x p_i q_i\right)}{1-a\beta}\hat{z}_t,$$

$$E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s\left(\kappa\hat{x}_{t+s}+\frac{1-\delta}{\delta}\hat{\pi}_{t+s}+\mu_{t+s}\right)\right]=\frac{a\beta\left(-\delta\gamma_\pi p_x q_x+(1-\delta\gamma_\pi p_\pi)q_\pi-\delta\gamma_\pi p_i q_i\right)}{1-a\beta\delta}\hat{z}_t,$$

where a is the perceived persistence, and $q \equiv (q_x, q_\pi, q_i)'$ is the relative sensitivity vector. The above equations give two linear equations for the two unknowns γ_x and γ_π. The solution is given by

$$\gamma_x = a(q_x - \sigma q_\pi),$$

$$\gamma_\pi = a\beta q_\pi,$$

where I am using the fact that $p'q = 1$. Finally, solving equations (A.57) and (A.58) for $\hat{x}_t$ and $\hat{\pi}_t$ results in equations (19) and (20). □

B Weighted Mean-Squared Forecast Error

The agents' time-t one-step-ahead forecast error given model θ is defined as

$$e_t(\theta) \equiv y_{t+1} - E_t^\theta[y_{t+1}],$$

where E_t^θ denotes the agents' subjective expectation conditional on their information at time t and given model θ. The weighted average of mean-squared forecast errors given the symmetric weight matrix $W \in \mathbb{R}^{n\times n}$ is defined as

$$\text{MSE}_W(\theta) = \mathbb{E}\left[e_t'(\theta) W e_t(\theta)\right].$$

Instead of assuming that agents use a model that minimizes the KLDR, one can assume that they make their forecasts using the model θ that minimizes $\text{MSE}_W(\theta)$ for some matrix W.

Using the mean-squared forecast error as the notion of fit has two disadvantages relative to the KLDR. First, the choice of matrix W introduces additional degrees of freedom when the observable is not a scalar. Second, the minimizer of weighted mean-squared errors is in general not invariant to linear transformations of the vector of observable (unless if the weight vector W is transformed accordingly).

Let θ^* denote a pseudo-true d-state model, and let $\hat{\Sigma}_y^*$ denote the implied subjective variance of y_{t+1} conditional on the agents' information at time t.

Proposition B.1. *If W is set to be the inverse of $\hat{\Sigma}_y^*$, then $\theta^* \in \arg\min_{\theta\in\Theta_d} MSE_W(\theta)$.*

The proposition establishes that mean-squared forecast error minimization coincides with KLDR minimization under the appropriate choice of the weighting matrix W.

C Details of the NK Application (For Online Publication)

C.1 Forward Guidance

By the linear invariance result, agents' expectations respect any intratemporal linear relationships that hold in the equilibrium without forward guidance. In particular, by equations (A.57) and (A.58),

$$\sigma E^{1*}[r_{t+s}^n] = E^{1*}[\hat{x}_{t+s}] + \sigma E^{1*}[\hat{i}_{t+s}] - \gamma_x\left(p_x E^{1*}[\hat{x}_{t+s}] + p_\pi E^{1*}[\hat{\pi}_{t+s}] + p_i E^{1*}[\hat{i}_{t+s}]\right),$$

$$E^{1*}[\mu_{t+s}] = E^{1*}[\hat{\pi}_{t+s}] - \kappa E^{1*}[\hat{x}_{t+s}] - \gamma_\pi\left(p_x E^{1*}[\hat{x}_{t+s}] + p_\pi E^{1*}[\hat{\pi}_{t+s}] + p_i E^{1*}[\hat{i}_{t+s}]\right).$$

Substituting in (17) and (18), I get

$$E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s\left(\frac{1-\beta}{\beta}\hat{x}_{t+s} - \sigma\left(\hat{i}_{t+s} - r_{t+s}^n\right) - \frac{\sigma}{\beta}\hat{\pi}_{t+s}\right)\right] = E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s v_x' f_{t+s}\right], \tag{C.1}$$

$$E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s\left(\kappa\hat{x}_{t+s} + \frac{1-\delta}{\delta}\hat{\pi}_{t+s} + \mu_{t+s}\right)\right] = E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s v_\pi' f_{t+s}\right], \tag{C.2}$$

where $v_x, v_\pi \in \mathbb{R}^3$ are vectors that satisfy

$$v_x' f_t = \frac{1}{\beta}\left[(1-\beta\gamma_x p_x)\hat{x}_t - (\sigma+\beta\gamma_x p_\pi)\hat{\pi}_t - \beta\gamma_x p_i \hat{i}_t\right],$$
$$v_\pi' f_t = \frac{1}{\delta}\left[-\delta\gamma_\pi p_x \hat{x}_t + (1-\delta\gamma_\pi p_\pi)\hat{\pi}_t - \delta\gamma_\pi p_i \hat{i}_t\right].$$

On the other hand,

$$E_t^{1*}[f_{t+s}] = \Sigma_{f_s\omega_T}\Sigma_{\omega_T\omega_T}^{-1}\omega_T, \tag{C.3}$$

where $\omega_T \equiv (f_t', \hat{i}_{t+1}, \ldots, \hat{i}_{t+T})' \in \mathbb{R}^{3+T}$, $\Sigma_{f_s\omega_T} \equiv E^{1*}[f_{t+s}\omega_T']$, and $\Sigma_{\omega_T\omega_T} \equiv E^{1*}[\omega_T\omega_T']$. Therefore,

$$E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s v_x' f_{t+s}\right] = \psi_{xT}'\omega_T,$$
$$E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s v_\pi' f_{t+s}\right] = \psi_{\pi T}'\omega_T.$$

where $\psi_{xT}, \psi_{\pi T} \in \mathbb{R}^{3+T}$ are vectors defined as

$$\psi_{xT}' \equiv (\psi_{xf}', \psi_{xi_1}, \ldots, \psi_{xi_T})' \equiv v_x'\left(\sum_{s=1}^{\infty}\beta^s\Sigma_{f_s\omega_T}\right)\Sigma_{\omega_T\omega_T}^{-1}, \tag{C.4}$$
$$\psi_{\pi T}' \equiv (\psi_{\pi f}', \psi_{\pi i_1}, \ldots, \psi_{\pi i_T})' \equiv v_\pi'\left(\sum_{s=1}^{\infty}(\beta\delta)^s\Sigma_{f_s\omega_T}\right)\Sigma_{\omega_T\omega_T}^{-1}, \tag{C.5}$$

and $\psi_{xf} \equiv (\psi_{xx}, \psi_{x\pi}, \psi_{xi})'$ and $\psi_{\pi f} \equiv (\psi_{\pi x}, \psi_{\pi\pi}, \psi_{\pi i})'$ are vectors in $\mathbb{R}^3$. Therefore,

$$E_t^{1*}\left[\sum_{s=1}^{\infty}\beta^s\left(\frac{1-\beta}{\beta}\hat{x}_{t+s} - \sigma\left(\hat{i}_{t+s} - r_{t+s}^n\right) - \frac{\sigma}{\beta}\hat{\pi}_{t+s}\right)\right] = \psi_{xf}' f_t + \sum_{s=1}^{T}\psi_{xi_s}\hat{i}_{t+s},$$
$$E_t^{1*}\left[\sum_{s=1}^{\infty}(\beta\delta)^s\left(\kappa\hat{x}_{t+s} + \frac{1-\delta}{\delta}\hat{\pi}_{t+s} + \mu_{t+s}\right)\right] = \psi_{\pi f}' f_t + \sum_{s=1}^{T}\psi_{\pi i_s}\hat{i}_{t+s}.$$

Substituting in equations (17) and (18), I get

$$\hat{x}_t = -\sigma\left(\hat{i}_t - r_t^n\right) + \psi_{xx}\hat{x}_t + \psi_{x\pi}\hat{\pi}_t + \psi_{xi}\hat{i}_t + \sum_{s=1}^{T}\psi_{xi_s}\hat{i}_{t+s},$$
$$\hat{\pi}_t = \kappa\hat{x}_t + \mu_t + \psi_{\pi x}\hat{x}_t + \psi_{\pi\pi}\hat{\pi}_t + \psi_{\pi i}\hat{i}_t + \sum_{s=1}^{T}\psi_{\pi i_s}\hat{i}_{t+s}.$$

The above equations can be solved for $\hat{x}_t$ and $\hat{i}_t$ to get,

$$\hat{x}_t = \alpha_{xi}\hat{i}_t + \alpha_{xn}r_t^n + \alpha_{x\mu}\mu_t + \sum_{s=1}^{T}\alpha_{xi_s}\hat{i}_{t+s},$$
$$\hat{\pi}_t = \alpha_{\pi i}\hat{i}_t + \alpha_{\pi n}r_t^n + \alpha_{\pi\mu}\mu_t + \sum_{s=1}^{T}\alpha_{\pi i_s}\hat{i}_{t+s}$$

for some constants that depend on the ψ's.

It only remains to compute ψ_{xT} and $\psi_{\pi T}$. I first compute the elements of $\Sigma_{f_s \omega_T}$. By the law of iterated expectations and Theorem 5,

$$E^{1*}[f_{t+s}f_t'] = E^{1*}\left[E^{1*}[f_{t+s}|f_t]f_t'\right] = E^{1*}\left[a^s q p' f_t f_t'\right] = a^s q p' \Gamma_0.$$

Next consider elements of the form $E^{1*}[f_{t+s}\hat{i}_{t+\tau}]$. If $s < \tau$, then

$$E^{1*}[f_{t+s}\hat{i}_{t+\tau}] = E^{1*}\left[f_{t+s}E^{1*}[\hat{i}_{t+\tau}|f_{t+s}]\right] = E^{1*}\left[f_{t+s}a^{\tau-s}q_i p' f_{t+s}\right] = a^{\tau-s}q_i E^{1*}\left[f_{t+s}f_{t+s}'\right] p = a^{\tau-s}q_i\Gamma_0 p.$$

Likewise, if $s > \tau$, then

$$E^{1*}[f_{t+s}\hat{i}_{t+\tau}] = E^{1*}\left[\hat{i}_{t+\tau}E^{1*}[f_{t+s}|f_{t+\tau}]\right] = E^{1*}\left[e_i' f_{t+\tau}a^{s-\tau}qp' f_{t+\tau}\right] = a^{s-\tau}qp' E^{1*}\left[f_{t+s}f_{t+s}'\right] e_i = a^{s-\tau}qp'\Gamma_0 e_i,$$

where e_i is the coordinate vector that selects element $\hat{i}_t$ of vector $f_t = (\hat{o}_t, \hat{\pi}_t, \hat{i}_t)$, i.e., $\hat{i}_t = e_i' f_t$. Finally, if $s = \tau$, then

$$E^{1*}[f_{t+s}\hat{i}_{t+\tau}] = E^{1*}[f_{t+s}f_{t+s}' e_i] = \Gamma_0 e_i.$$

I next compute the elements of $\Sigma_{\omega_T\omega_T}$. First, note that

$$E^{1*}[f_t f_t'] = \Gamma_0,$$

and

$$E^{1*}[f_t'\hat{i}_{t+\tau}] = E^{1*}\left[f_t' E^{1*}[\hat{i}_{t+\tau}|f_t]\right] = E^{1*}\left[a^\tau q_i p' f_t f_t'\right] = a^\tau q_i p'\Gamma_0.$$

Finally, if $\tau < \tau'$, then

$$E^{1*}[\hat{i}_{t+\tau}\hat{i}_{t+\tau'}] = E^{1*}\left[\hat{i}_{t+\tau}E^{1*}[\hat{i}_{t+\tau'}|f_{t+\tau}]\right] = E^{1*}\left[e_i' f_{t+\tau}a^{\tau'-\tau}q_i p' f_{t+\tau}\right] = a^{\tau'-\tau}q_i p'\Gamma_0 e_i,$$

and

$$E^{1*}[\hat{i}_{t+\tau}\hat{i}_{t+\tau}] = e_i' E^{1*}[f_{t+\tau}f_{t+\tau}']e_i = e_i'\Gamma_0 e_i.$$

Putting everything together, I get

$$\Sigma_{\omega_T\omega_T} = \begin{pmatrix} \Gamma_0 & aq_i\Gamma_0 p & a^2 q_i\Gamma_0 p & \dots & a^T q_i\Gamma_0 p \\ aq_i p'\Gamma_0 & e_i'\Gamma_0 e_i & aq_i p'\Gamma_0 e_i & \dots & a^{T-1}q_i p'\Gamma_0 e_i \\ a^2 q_i p'\Gamma_0 & aq_i p'\Gamma_0 e_i & e_i'\Gamma_0 e_i & \dots & a^{T-2}q_i p'\Gamma_0 e_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a^T q_i p'\Gamma_0 & a^{T-1}q_i p'\Gamma_0 e_i & a^{T-2}q_i p'\Gamma_0 e_i & \dots & e_i'\Gamma_0 e_i \end{pmatrix}. \tag{C.6}$$

and

$$\Sigma_{f_s\omega_T} = \begin{cases} \begin{pmatrix} a^s qp'\Gamma_0 & \Gamma_0 e_i & aq_i\Gamma_0 p & a^2 q_i\Gamma_0 p & \dots & a^{T-1}q_i\Gamma_0 p \end{pmatrix} & \text{if } s = 1, \\ \begin{pmatrix} a^s qp'\Gamma_0 & a^{s-1}qp'\Gamma_0 e_i & \dots & aqp'\Gamma_0 e_i & \Gamma_0 e_i & aq_i\Gamma_0 p & \dots & a^{T-s}q_i\Gamma_0 p \end{pmatrix} & \text{if } 1 < s < T, \\ \begin{pmatrix} a^s qp'\Gamma_0 & a^{s-1}qp'\Gamma_0 e_i & \dots & aqp'\Gamma_0 e_i & \Gamma_0 e_i \end{pmatrix} & \text{if } s = T, \\ \begin{pmatrix} a^s qp'\Gamma_0 & a^{s-1}qp'\Gamma_0 e_i & \dots & a^{s-T}qp'\Gamma_0 e_i \end{pmatrix} & \text{if } s > T. \end{cases} \tag{C.7}$$

Therefore,

$$\sum_{s=1}^{\infty} \beta^s \Sigma_{f_s \omega_T}$$
$$= \left(\sum_{s=1}^{\infty} (a\beta)^s qp'\Gamma_0 \quad \beta\Gamma_0 e_i + \sum_{s=2}^{\infty} a^{s-1}\beta^s qp'\Gamma_0 e_i \quad \ldots \quad \sum_{s=1}^{T-1} a^{T-s}\beta^s q_i \Gamma_0 p + \beta^T \Gamma_0 e_i + \sum_{s=T+1}^{\infty} a^{s-T}\beta^s qp'\Gamma_0 e_i \right)$$
$$= \left(\frac{a\beta qp'\Gamma_0}{1-a\beta} \quad \beta\Gamma_0 e_i + \frac{a\beta^2 qp'\Gamma_0 e_i}{1-a\beta} \quad a\beta q_i\Gamma_0 p + \beta^2\Gamma_0 e_i + \frac{a\beta^3 qp'\Gamma_0 e_i}{1-a\beta} \quad \ldots \quad \frac{(a^T\beta - \beta^T a) q_i \Gamma_0 p}{a-\beta} + \beta^T\Gamma_0 e_i + \frac{a\beta^{T+1} qp'\Gamma_0 e_i}{1-a\beta} \right).$$

Likewise,

$$\sum_{s=1}^{\infty} (\beta\delta)^s \Sigma_{f_s\omega_T} = \left(\frac{a\beta\delta qp'\Gamma_0}{1-a\beta\delta} \quad \beta\delta\Gamma_0 e_i + \frac{a(\beta\delta)^2 qp'\Gamma_0 e_i}{1-a\beta\delta} \quad \ldots \quad \frac{(a^T\beta\delta - (\beta\delta)^T a) q_i\Gamma_0 p}{a-\beta\delta} + (\beta\delta)^T\Gamma_0 e_i + \frac{a(\beta\delta)^{T+1} qp'\Gamma_0 e_i}{1-a\beta\delta} \right).$$

Given the expressions for $\Sigma_{\omega_T\omega_T}$, $\sum_{s=1}^{\infty} \beta^s \Sigma_{f_s\omega_T}$, and $\sum_{s=1}^{\infty} (\beta\delta)^s \Sigma_{f_s\omega_T}$, one can use (C.4) and (C.5) to find ψ_{xT} and $\psi_{\pi T}$.

C.2 Estimation

I choose the variance-covariance and lag-one autocovariance of $s_t \equiv (\hat{i}_t, r_t^n, \mu_t)'$ to match the variance-covariance and lag-one autocovariance of $f_t = (\hat{x}_t, \hat{\pi}_t, \hat{i}_t)'$. The estimated values are given by

$$E[s_t s_t'] = \begin{pmatrix} 10.9 & 16.4 & 0.200 \\ 16.4 & 32.1 & -0.0827 \\ 0.200 & -0.0827 & 0.0994 \end{pmatrix},$$

and

$$E[s_t s_{t-1}'] = \begin{pmatrix} 10.4 & 16.2 & 0.155 \\ 15.0 & 30.7 & -0.146 \\ 0.302 & 0.129 & 0.0920. \end{pmatrix}.$$

Figure C.1 plots $\rho(C_l)$ in solid red and $\rho(C_1)^l$ in dashed green, where $\rho(C_l)$ denotes the spectral radius of the lag-l autocorrelation matrix of f_t.[53] The figure verifies that the estimated process is exponentially ergodic, and so, the agents' pseudo-true model is described by Theorem 5.

D Details of the RBC Application (For Online Publication)

D.1 Temporary Equilibrium

The (log-)linearized temporary-equilibrium conditions are given by

$$\hat{y}_t = \hat{a}_t + \alpha\hat{k}_t + (1-\alpha)\hat{n}_t, \tag{D.1}$$
$$\hat{w}_t = \hat{a}_t + \alpha(\hat{k}_t - \hat{n}_t), \tag{D.2}$$
$$\hat{r}_t = r\hat{a}_t + (1-\alpha)r(\hat{n}_t - \hat{k}_t), \tag{D.3}$$

[53] The result would be identical if I instead used the autocorrelation matrices of s_t. This is a corollary of the linear invariance result.

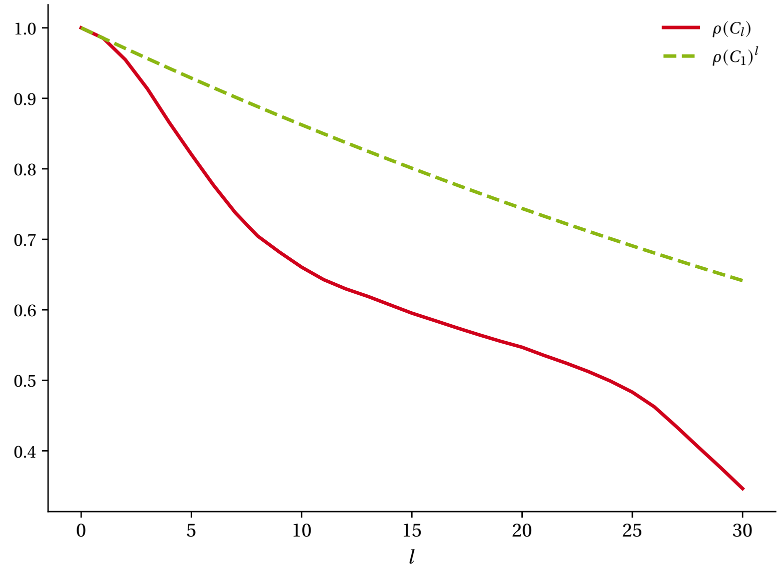

Figure C.1. Test of Exponential Ergodicity

$$\hat{n}_t = \frac{1}{\varphi}\hat{w}_t - \frac{1}{\sigma\varphi}\hat{c}_t, \tag{D.4}$$

$$\hat{i}_t = \frac{y}{i}\hat{y}_t - \frac{c}{i}\hat{c}_t, \tag{D.5}$$

$$\hat{k}_t = (1-\delta)\hat{k}_{t-1} + \delta\hat{i}_{t-1}, \tag{D.6}$$

$$\hat{a}_t = \rho\hat{a}_{t-1} + \epsilon_t, \tag{D.7}$$

$$\hat{c}_t = E_t[\hat{c}_{t+1}] - \sigma\beta E_t[\hat{r}_{t+1}], \tag{D.8}$$

where $\hat{r}_t$ denotes the first-order deviation of the rental rate of capital from its steady-state value and the remaining hatted variables are log-deviations from the corresponding steady-state values. The Euler equation (D.8) may not hold away from rational expectations if $\hat{c}_t$ denotes the aggregate consumption; it is valid under arbitrary expectations only if $\hat{c}_t$ denotes individual consumption. However, the individual consumption Euler equation can be combined with the households' intertemporal budget constraint and the transversality condition to obtain an aggregate consumption function that is valid under arbitrary expectations. The log-linearized household budget constraint is given by:

$$\hat{k}_{t+1} = (1-\delta+r)\hat{k}_t + \hat{r}_t + \frac{r(1-\alpha)}{\alpha}(\hat{w}_t + \hat{n}_t) - \frac{c}{k}\hat{c}_t.$$

Substituting for labor supply in the budget constraint, I get

$$\hat{k}_{t+1} = \frac{1}{\beta}\hat{k}_t + \hat{r}_t + \frac{(1-\alpha)(1+\varphi)r}{\alpha\varphi}\hat{w}_t - \left(\frac{(1-\alpha)r}{\alpha\sigma\varphi} + \frac{c}{k}\right)\hat{c}_t,$$

where I am using the fact that $1-\delta+r=\beta^{-1}$. Multiplying the above equation by β^t, summing over t, and taking subjective expectations of both sides, I get

$$\left(\frac{(1-\alpha)r}{\alpha\sigma\varphi}+\frac{c}{k}\right)\sum_{s=0}^{\infty}\beta^s E_t[\hat{c}_{t+s}]=\frac{1}{\beta}\hat{k}_t+\sum_{s=0}^{\infty}\beta^s\left(E_t[\hat{r}_{t+s}]+\frac{(1-\alpha)(1+\varphi)r}{\alpha\varphi}E_t[\hat{w}_{t+s}]\right).$$

Define

$$\chi\equiv(1-\beta)\left(\frac{(1-\alpha)r}{\alpha\sigma\varphi}+\frac{c}{k}\right)^{-1},$$
$$\zeta\equiv\frac{(1-\alpha)(1+\varphi)r}{\alpha\varphi}.$$

Then the above equation can be written as

$$\frac{1-\beta}{\chi}\sum_{s=0}^{\infty}\beta^s E_t[\hat{c}_{t+s}]=\frac{1}{\beta}\hat{k}_t+\sum_{s=0}^{\infty}\beta^s E_t[\hat{r}_{t+s}]+\zeta\sum_{s=0}^{\infty}\beta^s E_t[\hat{w}_{t+s}]. \tag{D.9}$$

On the other hand, the Euler equation implies

$$E_t[\hat{c}_{t+s}]=\hat{c}_t+\sigma\beta\sum_{\tau=1}^{s}E_t[\hat{r}_{t+\tau}].$$

Therefore,

$$\begin{aligned}\sum_{s=0}^{\infty}\beta^s E_t[\hat{c}_{t+s}]&=\sum_{s=0}^{\infty}\beta^s\hat{c}_t+\sigma\beta\sum_{s=1}^{\infty}\sum_{\tau=1}^{s}\beta^s E_t[\hat{r}_{t+s}]\\&=\frac{1}{1-\beta}\hat{c}_t+\sigma\beta\sum_{\tau=1}^{\infty}\sum_{s=\tau}^{\infty}\beta^s E_t[\hat{r}_{t+\tau}]\\&=\frac{1}{1-\beta}\hat{c}_t+\frac{\beta\sigma}{1-\beta}\sum_{\tau=1}^{\infty}\beta^\tau E_t[\hat{r}_{t+\tau}].\end{aligned}$$

Combining the above with equation (D.9), I get

$$\hat{c}_t=\frac{\chi}{\beta}\hat{k}_t+\chi\hat{r}_t+\chi\zeta\hat{w}_t+(\chi-\beta\sigma)\sum_{s=1}^{\infty}\beta^s E_t[\hat{r}_{t+s}]+\chi\zeta\sum_{s=1}^{\infty}\beta^s E_t[\hat{w}_{t+s}]. \tag{D.10}$$

D.2 Constrained Rational Expectations Equilibrium

Suppose households use a pseudo-true 1-state model to forecast the wage and the rental rate of capital. Define $\omega_t\equiv(\hat{o}_t,\hat{n}_t,\hat{w}_t,\hat{r}_t,\hat{c}_t,\hat{i}_t)$, $f_t\equiv(\hat{k}_t,\hat{a}_t)'$, and $\xi_t\equiv(f_t',\omega_t')'$. Let $v\in\mathbb{R}^8$ be a vector that satisfies

$$v'\xi_t=(\chi-\beta\sigma)\,\hat{r}_t+\chi\zeta\hat{w}_t.$$

Then equation (D.10) can be written as

$$\hat{c}_t=\frac{\chi}{\beta}\hat{k}_t+\chi\hat{r}_t+\chi\zeta\hat{w}_t+\sum_{s=1}^{\infty}\beta^s v'E_t[\xi_{t+s}].$$

Suppose $\xi_t = Tf_t$ for some full-rank matrix T—I later verify that this is indeed the case. Then by the linear-invariance result,

$$\hat{c}_t = \frac{\chi}{\beta}\hat{k}_t + \chi\hat{r}_t + \chi\zeta\hat{w}_t + \sum_{s=1}^{\infty}\beta^s v'TE_t[f_{t+s}].$$

The households' forecasts of f_t when they use model θ is given by (4). This can be written recursively as

$$E_t^{\theta}[f_{t+s}] = a^s(1-\eta)\hat{z}_t q, \tag{D.11}$$

$$\hat{z}_t = a\eta\hat{z}_{t-1} + p'f_t = a\eta\hat{z}_{t-1} + p_k\hat{k}_t + p_a\hat{a}_t, \tag{D.12}$$

where $\hat{z}_t$ denote the households' estimate of the subjective state at time t. Therefore,

$$\hat{c}_t = \frac{\chi}{\beta}\hat{k}_t + \chi\hat{r}_t + \chi\zeta\hat{w}_t + \frac{a\beta(1-\eta)}{1-a\beta}v'Tq\hat{z}_t. \tag{D.13}$$

I guess that $\eta = 0$ in equilibrium and later verify this guess. Solving for $\hat{z}_t$ from (D.12) and substituting in (D.13), I get

$$\hat{c}_t = \left(\frac{\chi}{\beta} + \gamma_k\right)\hat{k}_t + \chi\hat{r}_t + \chi\zeta\hat{w}_t + \gamma_a\hat{a}_t, \tag{D.14}$$

where

$$\gamma_k \equiv \frac{a\beta}{1-a\beta}v'Tqp_k, \tag{D.15}$$

$$\gamma_a \equiv \frac{a\beta}{1-a\beta}v'Tqp_a. \tag{D.16}$$

Equations (D.1)–(D.5) and (D.14) can be solved for ω_t as a function of f_t. This verifies the guess that $\xi_t = (f_t', \omega_t')' = Tf_t$ and leads to an expression for matrix T. In particular,

$$\hat{i}_t = \psi_k\hat{k}_t + \psi_a\hat{a}_t,$$

for some ψ_k and ψ_a. Substituting for $\hat{i}_{t-1}$ from above in (D.6), I get

$$\hat{k}_t = (1-\delta+\delta\psi_k)\hat{k}_{t-1} + \delta\psi_a\hat{a}_{t-1}. \tag{D.17}$$

I can now describe the constrained rational expectations equilibrium. Equations (D.7) and (D.17) can be written in vector form as

$$f_t = \mathbb{F}(\gamma_k, \gamma_a)f_{t-1} + \epsilon_t. \tag{D.18}$$

An equilibrium is given by tuples (γ_k^*, γ_a^*) and (a^*, η^*, p^*, q^*) such that (i) (a^*, η^*, p^*, q^*) is the pseudo-true 1-state model given the true process (D.18) where $\gamma_k = \gamma_k^*$ and $\gamma_a = \gamma_a^*$, (ii) γ_k^* and γ_a^* are given by equations (D.15) and (D.16) for $a = a^*$, $p = p^*$, and $q = q^*$, and (iii) $\eta^* = 0$.

Finding an equilibrium requires solving a fixed-point equation. I start with a candidate $(\gamma_k, \gamma_a, \eta)$, with $\eta = 0$. The candidate defines a true process as in (D.18). This process in turn leads to a pseudo-true 1-state model $(\tilde{a}, \tilde{\eta}, \tilde{p}, \tilde{q})$. Such a pseudo-true 1-state model, in turn, defines a $(\tilde{\gamma}_k, \tilde{\gamma}_a)$ pair through equations (D.15) and (D.16). I solve for the equilibrium by numerically minimizing the Euclidean distance between tuples $(\tilde{\gamma}_k, \tilde{\gamma}_a, \tilde{\eta})$ and $(\gamma_k, \gamma_a, \eta)$ over the set of all (γ_k, γ_a) pairs. The fixed-point turns out to satisfy $\tilde{\eta} = \eta = 0$, verifying my earlier conjecture.

E Details of the DMP Application (For Online Publication)

E.1 Non-Linear Equilibrium

I start with the workers' problem. Let U_t and V_t denote the time-t value to a worker of unemployment and employment, respectively. Those random variables solve the following Bellman equations:

$$U_t = b + \beta E_t \left[p_t V_{t+1} + (1 - p_t) U_{t+1} \right], \tag{E.1}$$

$$V_t = w_t + \beta E_t \left[s_t U_{t+1} + (1 - s_t) V_{t+1} \right], \tag{E.2}$$

where b denotes the workers' flow payoff from being unemployed, w_t denotes the wage rate, and $p_t = \mu \theta_t^{1-\alpha}$ denotes the job-finding probability, with θ_t the labor market tightness and μ and α parameters of the matching function. Subtracting U_t from V_t, I get

$$V_t - U_t = w_t - b + \beta E_t \left[(1 - s_t - p_t)(V_{t+1} - U_{t+1}) \right]. \tag{E.3}$$

Define

$$\lambda^w_{t,t+\tau} \equiv \prod_{k=0}^{\tau-1} (1 - s_{t+k} - p_{t+k}). \tag{E.4}$$

Solving (E.3) forward, I get

$$V_t - U_t = w_t - b + E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \lambda^w_{t,t+\tau} (w_{t+\tau} - b) \right]. \tag{E.5}$$

This equation is valid under arbitrary expectations.

I consider the firms next. Let J_t denote the time-t value to a firm of a job. It solves the following Bellman equation:

$$J_t = a_t - w_t + \beta E_t [(1 - s_t) J_{t+1}].$$

Solving the equation forward, I get

$$J_t = a_t - w_t + E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \lambda^f_{t,t+\tau} (a_{t+\tau} - w_{t+\tau}) \right], \tag{E.6}$$

where

$$\lambda^f_{t,t+\tau} \equiv \prod_{k=0}^{\tau-1} (1 - s_{t+k}). \tag{E.7}$$

Free entry by firms implies

$$0 = -k + \beta E_t [q_t J_{t+1}], \tag{E.8}$$

where $q_t = \mu\theta_t^{-\alpha}$ is the probability of filling a vacancy in each period. Substituting for J_t in (E.8) from (E.6), I get

$$\theta_t^\alpha = \frac{\mu}{k} E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \lambda_{t+1,t+\tau}^f (a_{t+\tau} - w_{t+\tau}) \right]. \tag{E.9}$$

Equation (E.9) determines tightness as a function of the firms' expectations of wage and labor productivity.

The wage rate is determined by Nash bargaining. Under Nash bargaining,

$$\frac{J_t}{1-\delta} = \frac{V_t - U_t}{\delta},$$

where δ denotes the workers' bargaining power. Combining the above equation with (E.5) and (E.6) and solving for w_t, I get

$$w_t = \delta a_t + (1-\delta)b + \delta E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \lambda_{t,t+\tau}^f (a_{t+\tau} - w_{t+\tau}) \right] - (1-\delta) E_t \left[\sum_{\tau=1}^{\infty} \beta^\tau \lambda_{t,t+\tau}^w (w_{t+\tau} - b) \right]. \tag{E.10}$$

The unemployment rate follows the first-order difference equation

$$u_t = u_{t-1} + s_{t-1}(1 - u_{t-1}) - p_{t-1} u_{t-1}. \tag{E.11}$$

E.2 Steady State

I first consider a steady state in which $a_t = 1 > b$, $w_t = w$, $\theta_t = \theta$, $s_t = s$, and agents have perfect foresight. Equation (E.10) implies that in the steady state,

$$\frac{(1-\delta)(w-b)}{1-\beta(1-s-p)} = \frac{\delta(1-w)}{1-\beta(1-s)}.$$

Therefore,

$$w = \frac{\delta(1-\beta(1-s-p)) + (1-\delta)(1-\beta(1-s))b}{1-\beta(1-s-\delta p)}.$$

Equation (E.6) implies that the value of a job to a firm is constant in the steady state:

$$J_t = J = \frac{1}{1-\beta(1-s)}(1-w).$$

Equation (E.8) and the definition of q_t imply

$$\frac{\mu}{k\theta^\alpha} = \frac{1}{\beta J}.$$

The steady-state unemployment rate satisfies

$$s\frac{1-u}{u} = p.$$

E.3 Log-Linear Model

I next log-linearize the model around the steady state. Log-linearizing (E.4) and (E.7), I get

$$\hat{\lambda}^{w}_{t,t+\tau} = -\frac{1}{1-s-p}\sum_{k=0}^{\tau-1}\left(p\hat{p}_{t+k} + s\hat{s}_{t+k}\right),$$

$$\hat{\lambda}^{f}_{t,t+\tau} = -\frac{1}{1-s}\sum_{k=0}^{\tau-1} s\hat{s}_{t+k}.$$

Log-linearizing $p_t = \mu\theta_t^{1-\alpha}$, I get

$$\hat{p}_t = (1-\alpha)\hat{\theta}_t.$$

Log-linearizing (E.9),

$$\hat{\theta}_t = E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s)^{\tau-1}\left(\frac{(1-b)\mu}{\alpha k\theta^{\alpha}}\hat{a}_{t+\tau} - \frac{w\mu}{\alpha k\theta^{\alpha}}\hat{w}_{t+\tau} + \frac{(1-w)\mu}{\alpha k\theta^{\alpha}}\hat{\lambda}_{t+1,t+\tau}\right)\right].$$

Substituting for $\mu/(k\theta^{\alpha})$, I get

$$\hat{\theta}_t = E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\left(\frac{1-b}{\alpha J}\hat{a}_{t+\tau} - \frac{w}{\alpha J}\hat{w}_{t+\tau} + \frac{1-w}{\alpha J}\hat{\lambda}_{t+1,t+\tau}\right)\right].$$

The term involving $\hat{\lambda}^{f}_{t,t+\tau}$ can be simplified further:

$$\begin{aligned}
\sum_{\tau=1}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\hat{\lambda}_{t+1,t+\tau} &= -\sum_{\tau=2}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\frac{s}{1-s}\sum_{k=0}^{\tau-2}\hat{s}_{t+1+k} \\
&= -s\sum_{k=0}^{\infty}\hat{s}_{t+1+k}\sum_{\tau=k+2}^{\infty}\beta^{\tau-1}(1-s)^{\tau-2} \\
&= -\frac{\beta s}{1-\beta(1-s)}\sum_{\tau=1}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\hat{s}_{t+\tau}.
\end{aligned}$$

Define

$$\zeta \equiv \frac{\beta s(1-w)}{1-\beta(1-s)}.$$

Then,

$$\hat{\theta}_t = \frac{1}{\alpha J}E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\left((1-b)\hat{a}_{t+\tau} - w\hat{w}_{t+\tau} - \zeta\hat{s}_{t+\tau}\right)\right]. \tag{E.12}$$

Log-linearizing (E.10),

$$\begin{aligned}
w\hat{w}_t = \delta(1-b)\hat{a}_t + \delta E_t&\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s)^{\tau}\left((1-b)\hat{a}_{t+\tau} - w\hat{w}_{t+\tau} + (1-w)\hat{\lambda}^{f}_{t,t+\tau}\right)\right] \\
-(1-\delta)E_t&\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s-p)^{\tau}\left((w-b)\hat{\lambda}^{w}_{t,t+\tau} + w\hat{w}_{t+\tau}\right)\right].
\end{aligned} \tag{E.13}$$

The terms involving $\hat{\lambda}^w_{t,t+\tau}$ and $\hat{\lambda}^f_{t,t+\tau}$ can be simplified further:

$$
\begin{aligned}
\sum_{\tau=1}^{\infty} \beta^{\tau}(1-s-p)^{\tau} \hat{\lambda}^w_{t,t+\tau} &= -\sum_{\tau=1}^{\infty} \beta^{\tau}(1-s-p)^{\tau} \frac{1}{1-s-p} \sum_{k=0}^{\tau-1} (p\hat{p}_{t+k} + s\hat{s}_{t+k}) \\
&= -\beta \sum_{k=0}^{\infty} (p\hat{p}_{t+k} + s\hat{s}_{t+k}) \sum_{\tau=k+1}^{\infty} (\beta(1-s-p))^{\tau-1} \\
&= -\frac{\beta}{1-\beta(1-s-p)} \sum_{k=0}^{\infty} \beta^k (1-s-p)^k (p\hat{p}_{t+k} + s\hat{s}_{t+k}) .
\end{aligned}
$$

$$
\sum_{\tau=1}^{\infty} \beta^{\tau}(1-s)^{\tau} \hat{\lambda}^f_{t,t+\tau} = -\frac{\beta s}{1-\beta(1-s)} \sum_{k=0}^{\infty} \beta^k (1-s)^k \hat{s}_{t+k}.
$$

Define

$$
\chi \equiv \frac{\beta(1-\delta)(w-b)}{1-\beta(1-s-p)}
$$

Then, (E.13) can be written as

$$
\begin{aligned}
w\hat{w}_t &= \delta(1-b)\hat{a}_t + (s\chi - \delta\zeta)\hat{s}_t + p\chi(1-\alpha)\hat{\theta}_t \\
&+ E_t \left[\sum_{\tau=1}^{\infty} \beta^{\tau}(1-s)^{\tau} (\delta(1-b)\hat{a}_{t+\tau} - \delta w\hat{w}_{t+\tau} - \delta\zeta\hat{s}_{t+\tau}) \right] \\
&- E_t \left[\sum_{\tau=1}^{\infty} \beta^{\tau}(1-s-p)^{\tau} \left((1-\delta)w\hat{w}_{t+\tau} - p\chi(1-\alpha)\hat{\theta}_{t+\tau} - s\chi\hat{s}_{t+\tau} \right) \right] .
\end{aligned}
\tag{E.14}
$$

Finally, log-linearizing (E.11),

$$
\hat{u}_t = (1-s-p)\hat{u}_{t-1} - (1-\alpha)p\hat{\theta}_{t-1} + p\hat{s}_{t-1}. \tag{E.15}
$$

E.4 Rational Expectations Equilibrium

I guess and verify that under rational expectations $\hat{\theta}_t = \gamma_{\theta a}\hat{a}_t + \gamma_{\theta s}\hat{s}_t$ and $w\hat{w}_t = \gamma_{wa}\hat{a}_t + \gamma_{ws}\hat{s}_t$. Substituting in (E.12) and (E.14), I get

$$
\hat{\theta}_t = \frac{\rho_a}{1-\beta\rho_a(1-s)} \frac{1-b-\gamma_{wa}}{\alpha J} \hat{a}_t - \frac{\rho_s}{1-\beta\rho_s(1-s)} \frac{\zeta+\gamma_{ws}}{\alpha J} \hat{s}_t,
$$

and

$$
\begin{aligned}
w\hat{w}_t = &\left[\delta(1-b) + p\chi(1-\alpha)\gamma_{\theta a} + \frac{\beta\delta\rho_a(1-s)(1-b-\gamma_{wa})}{1-\beta\rho_a(1-s)} \right] \hat{a}_t \\
&+ \left[\frac{\beta\rho_a(1-s-p)}{1-\beta\rho_a(1-s-p)} (p\chi(1-\alpha)\gamma_{\theta a} - (1-\delta)\gamma_{wa}) \right] \hat{a}_t \\
&+ \left[s\chi - \delta\zeta + p\chi(1-\alpha)\gamma_{\theta s} - \frac{\beta\delta\rho_s(1-s)(\zeta+\gamma_{ws})}{1-\beta\rho_s(1-s)} \right] \hat{s}_t \\
&+ \left[\frac{\beta\rho_s(1-s-p)}{1-\beta\rho_s(1-s-p)} (p\chi(1-\alpha)\gamma_{\theta s} + s\chi - (1-\delta)\gamma_{ws}) \right] \hat{s}_t .
\end{aligned}
$$

These equations validate the guess and yield four linear equations for the four unknowns $\gamma_{\theta a}$, $\gamma_{\theta s}$, γ_{wa}, and γ_{ws}, which can be solved given values for the exogenous parameters. The rational-expectations equilibrium is then described by (31) and (E.15) with $\hat{w}_t = \gamma^*_{wa}\hat{a}_t + \gamma^*_{ws}\hat{s}_t$, $\hat{\theta}_t = \gamma^*_{\theta a}\hat{a}_t + \gamma^*_{\theta s}\hat{s}_t$, and $(\gamma^*_{\theta a}, \gamma^*_{\theta s}, \gamma^*_{wa}, \gamma^*_{ws})$ the solution to the above linear equations.

E.5 Constrained Rational Expectations Equilibrium

I next consider the equilibrium where agents are constrained to use pseudo-true 1-state models. I guess (and later verify) that, in equilibrium,

$$\hat{\theta}_t = \psi_{\theta u}\hat{u}_t + \psi_{\theta a}\hat{a}_t + \psi_{\theta s}\hat{s}_t,$$
$$w\hat{w}_t = \psi_{wu}\hat{u}_t + \psi_{wa}\hat{a}_t + \psi_{ws}\hat{s}_t.$$

Using the linear-invariance result to substitute for $\hat{\theta}_{t+\tau}$ and $\hat{w}_{t+\tau}$ in (E.12) and (E.14), I get

$$\hat{\theta}_t = \frac{1}{\alpha J}E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau-1}(1-s)^{\tau-1}\left((1-b-\psi_{wa})\hat{a}_{t+\tau} - \psi_{wu}\hat{u}_{t+\tau} - (\zeta+\psi_{ws})\hat{s}_{t+\tau}\right)\right], \tag{E.16}$$

and

$$\begin{aligned}
w\hat{w}_t = {} & p\chi(1-\alpha)\psi_{\theta u}\hat{u}_t + (\delta(1-b) + p\chi(1-\alpha)\psi_{\theta a})\hat{a}_t + (s\chi - \delta\zeta + p\chi(1-\alpha)\psi_{\theta s})\hat{s}_t \\
& + E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s)^{\tau}(\delta(1-b-\psi_{wa})\hat{a}_{t+\tau} - \delta\psi_{wu}\hat{u}_{t+\tau} - \delta(\zeta+\psi_{ws})\hat{s}_{t+\tau})\right] \\
& + E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s-p)^{\tau}(p\chi(1-\alpha)\psi_{\theta u} - (1-\delta)\psi_{wu})\hat{u}_{t+\tau}\right] \\
& + E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s-p)^{\tau}(p\chi(1-\alpha)\psi_{\theta a} - (1-\delta)\psi_{wa})\hat{a}_{t+\tau}\right] \\
& + E_t\left[\sum_{\tau=1}^{\infty}\beta^{\tau}(1-s-p)^{\tau}(p\chi(1-\alpha)\psi_{\theta s} - (1-\delta)\psi_{ws} + s\chi)\hat{s}_{t+\tau}\right].
\end{aligned} \tag{E.17}$$

The agents' forecasts are given by equation (4). I guess that $\eta = 0$ in equilibrium and later verify this guess. Given the guess,

$$E_t[\hat{u}_{t+\tau}] = a^{\tau}q_u p_u\hat{u}_t + a^{\tau}q_u p_a\hat{a}_t + a^{\tau}q_u p_s\hat{s}_t,$$
$$E_t[\hat{a}_{t+\tau}] = a^{\tau}q_a p_u\hat{u}_t + a^{\tau}q_a p_a\hat{a}_t + a^{\tau}q_a p_s\hat{s}_t,$$
$$E_t[\hat{s}_{t+\tau}] = a^{\tau}q_s p_u\hat{u}_t + a^{\tau}q_s p_a\hat{a}_t + a^{\tau}q_s p_s\hat{s}_t.$$

Using the linear-invariance result to substitute for $E_t[\hat{u}_{t+\tau}]$, $E_t[\hat{a}_{t+\tau}]$, and $E_t[\hat{s}_{t+\tau}]$ in (E.16) and (E.17) and collecting terms verifies the guess that $\hat{\theta} = \psi_{\theta u}\hat{u}_t + \psi_{\theta a}\hat{a}_t + \psi_{\theta s}\hat{s}_t$ and $\hat{w}_t = \psi_{wu}\hat{u}_t + \psi_{wa}\hat{a}_t + \psi_{ws}\hat{s}_t$ and leads to the following linear equations for $\psi_{\theta u}$, $\psi_{\theta a}$, $\psi_{\theta s}$, ψ_{wu}, ψ_{wa}, and ψ_{ws}:

$$\psi_{\theta u} = \frac{ap_u}{1-a\beta(1-s)}\left(\frac{1-b-\psi_{wa}}{\alpha J}q_a - \frac{\psi_{wu}}{\alpha J}q_u - \frac{\zeta+\psi_{ws}}{\alpha J}q_s\right), \tag{E.18}$$

$$\psi_{\theta a} = \frac{ap_a}{1 - a\beta(1-s)} \left(\frac{1 - b - \psi_{wa}}{\alpha J} q_a - \frac{\psi_{wu}}{\alpha J} q_u - \frac{\zeta + \psi_{ws}}{\alpha J} q_s \right), \tag{E.19}$$

$$\psi_{\theta s} = \frac{ap_s}{1 - a\beta(1-s)} \left(\frac{1 - b - \psi_{wa}}{\alpha J} q_a - \frac{\psi_{wu}}{\alpha J} q_u - \frac{\zeta + \psi_{ws}}{\alpha J} q_s \right), \tag{E.20}$$

$$\begin{aligned} \psi_{wu} = {} & p\chi(1-\alpha)\psi_{\theta u} + \frac{a\beta\delta(1-s)p_u}{1 - a\beta(1-s)} \left[(1 - b - \psi_{wa})q_a - \psi_{wu}q_u - (\zeta + \psi_{ws})q_s \right] \\ & + \frac{a\beta(1-s-p)p_u}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta a} - (1-\delta)\psi_{wa})q_a + (p\chi(1-\alpha)\psi_{\theta u} - (1-\delta)\psi_{wu})q_u \right] \\ & + \frac{a\beta(1-s-p)p_u}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta s} - (1-\delta)\psi_{ws} + s\chi)q_s \right], \end{aligned} \tag{E.21}$$

$$\begin{aligned} \psi_{wa} = {} & \delta(1-b) + p\chi(1-\alpha)\psi_{\theta a} + \frac{a\beta\delta(1-s)p_a}{1 - a\beta(1-s)} \left[(1 - b - \psi_{wa})q_a - \psi_{wu}q_u - (\zeta + \psi_{ws})q_s \right] \\ & + \frac{a\beta(1-s-p)p_a}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta a} - (1-\delta)\psi_{wa})q_a + (p\chi(1-\alpha)\psi_{\theta u} - (1-\delta)\psi_{wu})q_u \right] \\ & + \frac{a\beta(1-s-p)p_a}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta s} - (1-\delta)\psi_{ws} + s\chi)q_s \right], \end{aligned} \tag{E.22}$$

$$\begin{aligned} \psi_{ws} = {} & s\chi - \delta\zeta + p\chi(1-\alpha)\psi_{\theta s} + \frac{a\beta\delta(1-s)p_s}{1 - a\beta(1-s)} \left[(1 - b - \psi_{wa})q_a - \psi_{wu}q_u - (\zeta + \psi_{ws})q_s \right] \\ & + \frac{a\beta(1-s-p)p_s}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta a} - (1-\delta)\psi_{wa})q_a + (p\chi(1-\alpha)\psi_{\theta u} - (1-\delta)\psi_{wu})q_u \right] \\ & + \frac{a\beta(1-s-p)p_s}{1 - a\beta(1-s-p)} \left[(p\chi(1-\alpha)\psi_{\theta s} - (1-\delta)\psi_{ws} + s\chi)q_s \right]. \end{aligned} \tag{E.23}$$

I can now describe the constrained rational expectations equilibrium. Given $\hat{\theta} = \psi_{\theta u}\hat{u}_t + \psi_{\theta a}\hat{a}_t + \psi_{\theta s}\hat{s}_t$, equations (31) and (E.15) can be written in vector form as

$$f_t = \mathbb{F}(\psi_{\theta u}, \psi_{\theta a}, \psi_{\theta s}) f_{t-1} + \epsilon_t. \tag{E.24}$$

An equilibrium is then given by tuples $(\psi^*_{\theta u}, \psi^*_{\theta a}, \psi^*_{\theta s}, \psi^*_{wu}, \psi^*_{wa}, \psi^*_{ws})$ and (a^*, η^*, p^*, q^*) such that (i) (a^*, η^*, p^*, q^*) is the pseudo-true 1-state model given the true process (E.24) with $\psi_{\theta u} = \psi^*_{\theta u}$, $\psi_{\theta a} = \psi^*_{\theta a}$, and $\psi_{\theta s} = \psi^*_{\theta s}$, (ii) $(\psi^*_{\theta u}, \psi^*_{\theta a}, \psi^*_{\theta s}, \psi^*_{wu}, \psi^*_{wa}, \psi^*_{ws})$ solves (E.18)–(E.23) given $a = a^*$, $p = p^*$, $q = q^*$, and (iii) $\eta^* = 0$.

Finding an equilibrium requires solving a fixed-point equation. I start with a candidate $(\psi_{\theta u}, \psi_{\theta a}, \psi_{\theta s}, \eta)$, with $\eta = 0$. The candidate defines a true process as in (E.24). The process leads to a pseudo-true 1-state model $(\tilde{a}, \tilde{\eta}, \tilde{p}, \tilde{q})$. Such a pseudo-true 1-state model, in turn, defines a $(\tilde{\psi}_{\theta u}, \tilde{\psi}_{\theta a}, \tilde{\psi}_{\theta s})$ pair through equations (E.18)–(E.23). I solve for the equilibrium by numerically minimizing the Euclidean distance between pairs $(\tilde{\psi}_{\theta u}, \tilde{\psi}_{\theta a}, \tilde{\psi}_{\theta s}, \tilde{\eta})$ and $(\psi_{\theta u}, \psi_{\theta a}, \psi_{\theta s}, \eta)$ over the set of all $(\psi_{\theta u}, \psi_{\theta a}, \psi_{\theta s})$ tuples. The fixed-point turns out to satisfy $\tilde{\eta} = \eta = 0$, verifying my earlier conjecture.

References

Adam, Klaus, Pei Kuang, and Albert Marcet (2012), "House Price Booms and the Current Account." In *NBER Macroeconomics Annual*, volume 26, 77–122.

Adam, Klaus and Albert Marcet (2011), "Internal Rationality, Imperfect Market Knowledge and Asset Prices." *Journal of Economic Theory*, 146, 1224–1252.

Adam, Klaus, Albert Marcet, and Johannes Beutel (2017), "Stock Price Booms and Expected Capital Gains." *American Economic Review*, 107, 2352–2408.

Afrouzi, Hassan, Spencer Yongwook Kwon, Augustin Landier, Yueran Ma, and David Thesmar (2021), "Overreaction in Expectations: Evidence and Theory." Working paper.

Akaike, Hirotugu (1975), "Markovian Representation of Stochastic Processes by Canonical Variables." *SIAM Journal on Control*, 13, 162–173.

Alvarez, Fernando E., Francesco Lippi, and Luigi Paciello (2015), "Monetary Shocks in Models with Inattentive Producers." *The Review of economic studies*, 83, 421–459.

Anderson, Brian D.O. and John B. Moore (2005), *Optimal Filtering*. Dover Publications, Mineola, N.Y.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2020), "Business-Cycle Anatomy." *American Economic Review*, 110, 3030–70.

Angeletos, George-Marios and Zhen Huo (2021), "Myopia and Anchoring." *American Economic Review*, 111, 1166–1200.

Angeletos, George-Marios and Jennifer La'O (2009), "Incomplete Information, Higher-Order Beliefs and Price Inertia." *Journal of Monetary Economics*, 56, S19–S37.

Angeletos, George-Marios and Chen Lian (2016), "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination." In *Handbook of Macroeconomics*, volume 2, 1065–1240, Elsevier.

Angeletos, George-Marios and Chen Lian (2018), "Forward Guidance without Common Knowledge." *American Economic Review*, 108, 2477–2512.

Angeletos, George-Marios and Karthik Sastry (2021), "Inattentive Economies." Working paper.

Berk, Robert H. (1966), "Limiting Behavior of Posterior Distributions When the Model is Incorrect." *Annals of Mathematical Statistics*, 37, 51–58.

Bhandari, Anmol, Jaroslav Borovička, and Paul Ho (2019), "Survey Data and Subjective Beliefs in Business Cycle Models." Working paper.

Bianchi, Francesco, Cosmin L. Ilut, and Hikaru Saijo (2021), "Diagnostic Business Cycles." Working paper.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018), "Diagnostic Expectations and Credit Cycles." *Journal of Finance*, 73, 199–227.

Bray, M. M. and N. E. Savin (1986), "Rational Expectations Equilibria, Learning, and Model Specification." *Econometrica*, 54, 1129–1160.

Bray, Margaret (1982), "Learning, Estimation, and the Stability of Rational Expectations." *Journal of Economic Theory*, 26, 318–339.

Bullard, James, George W. Evans, and Seppo Honkapohja (2008), "Monetary Policy, Judgment, and Near-Rational Exuberance." *American Economic Review*, 98, 1163–1177.

Bullard, James and Kaushik Mitra (2002), "Learning About Monetary Policy Rules." *Journal of Monetary Economics*, 49, 1105–1129.

Bunke, Olaf and Xavier Milhaud (1998), "Asymptotic Behavior of Bayes Estimates Under Possibly Incorrect Models." *Annals of Statistics*, 26, 617–644.

Caplin, Andrew and Mark Dean (2013), "Behavioral Implications of Rational Inattention with Shannon Entropy." Working paper.

Chahrour, Ryan, Kristoffer Nimark, and Stefan Pitschner (2021), "Sectoral Media Focus and Aggregate Fluctuations." *American Economic Review*, 111, 3872–3922.

Cho, In Koo, Noah Williams, and Thomas J. Sargent (2002), "Escaping Nash Inflation." *Review of Economic Studies*, 69, 1–40.

Da Silveira, Rava Azeredo, Yeji Sung, and Michael Woodford (2020), "Optimally Imprecise Memory and Biased Forecasts." Working paper.

Douc, Randal and Eric Moulines (2012), "Asymptotic Properties of the Maximum Likelihood Estimation in Misspecified Hidden Markov Models." *Annals of Statistics*, 40, 2697–2732.

Esponda, Ignacio and Demian Pouzo (2016), "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models." *Econometrica*, 84, 1093–1130.

Esponda, Ignacio and Demian Pouzo (2021), "Equilibrium in Misspecified Markov Decision Processes." *Theoretical Economics*, 16, 717–757.

Eusepi, Stefano and Bruce Preston (2011), "Expectations, Learning, and Business Cycle Fluctuations." *American Economic Review*, 101, 2844–2872.

Eusepi, Stefano and Bruce Preston (2018a), "Fiscal Foundations of Inflation: Imperfect Knowledge." *American Economic Review*, 108, 2551–89.

Eusepi, Stefano and Bruce Preston (2018b), "The Science of Monetary Policy: An Imperfect Knowledge Perspective." *Journal of Economic Literature*, 56, 3–59.

Evans, George W. and Seppo Honkapohja (2012), *Learning and Expectations in Macroeconomics*. Princeton University Press.

Farhi, Emmanuel and Iván Werning (2017), "Monetary Policy, Bounded Rationality and Incomplete Markets." Working paper.

Forni, Mario and Marco Lippi (2001), “The Generalized Dynamic Factor Model: Representation Theory.” *Econometric theory*, 17, 1113–1141.

Fuster, Andreas, Benjamin Hebert, and David Laibson (2012), “Investment Dynamics with Natural Expectations.” *International Journal of Central Banking*, 8, 243–265.

Fuster, Andreas, David Laibson, and Brock Mendel (2010), “Natural Expectations and Macroeconomic Fluctuations.” *Journal of Economic Perspectives*, 24, 67–84.

Gabaix, Xavier (2014), “A Sparsity-Based Model of Bounded Rationality.” *Quarterly Journal of Economics*, 129, 1661–1710.

Gabaix, Xavier (2020), “A Behavioral New Keynesian Model.” *American Economic Review*, 110, 2271–2327. Working paper.

Galí, Jordi (2015), *Monetary Policy, Inflation, and the Business Cycle*, second edition. Princeton University Press.

Garcia-Schmidt, Mariana and Michael Woodford (2015), “Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis.” Working paper.

Gáti, Laura (2020), “Monetary Policy & Anchored Expectations: An Endogenous Gain Learning Model.” Working Paper.

Gevers, Michel and Vincent Wertz (1984), “Uniquely Identifiable State-Space and ARMA Parametrizations for Multivariable Linear Systems.” *Automatica*, 20, 333–347.

Grandmont, Jean-Michel (1977), “Temporary General Equilibrium Theory.” *Econometrica*, 45, 535–572.

Grandmont, Jean-Michel (1982), “Temporary General Equilibrium Theory.” *Handbook of Mathematical Economics*, 2, 879–922.

Hagedorn, Marcus and Iourii Manovskii (2008), “The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited.” *American Economic Review*, 98, 1692–1706.

Hansen, Lars Peter and Thomas J. Sargent (2008), *Robustness*. Princeton University Press.

Hicks, John R. (1939), *Value and Capital*. Oxford: Clarendon.

Ho, B. L. and R. E. Kálmán (1966), “Effective Construction of Linear State-Variable Models from Input/Output Functions.” *at-Automatisierungstechnik*, 14, 545–548.

Kleijn, B. J. K. and A. W. Van Der Vaart (2006), “Misspecification in Infinite-Dimensional Bayesian Statistics.” *Annals of Statistics*, 34, 837–877.

Lindahl, Erik (1939), *Studies in the Theory of Money and Capital*. London: Allen & Unwin.

Lucas, Robert E. (1972), “Expectations and the Neutrality of Money.” *Journal of Economic Theory*, 4, 103–124.

Maćkowiak, Bartosz and Mirko Wiederholt (2009), "Optimal Sticky Prices under Rational Inattention." *American Economic Review*, 99, 769–803.

Malmendier, Ulrike and Stefan Nagel (2016), "Learning From Inflation Experiences." *Quarterly Journal of Economics*, 131, 53–87.

Mankiw, N. Gregory and Ricardo Reis (2002), "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics*, 117, 1295–1328.

Marcet, Albert and Juan P. Nicolini (2003), "Recurrent Hyperinflations and Learning." *American Economic Review*, 93, 1476–1498.

Marcet, Albert and Thomas J. Sargent (1989a), "Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information." *Journal of Political Economy*, 97, 1306–1322.

Marcet, Albert and Thomas J. Sargent (1989b), "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models." *Journal of Economic Theory*, 48, 337–368.

Meyn, Sean P. and Richard L. Tweedie (1993), *Markov Chains and Stochastic Stability*. Springer-Verlag.

Molavi, Pooya (2019), "Macroeconomics with Learning and Misspecification: A General Theory and Applications." Working paper.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin (2021), "Model Complexity, Expectations, and Asset Prices." Working Paper.

Orphanides, Athanasios and John C. Williams (2005), "Inflation Scares and Forecast-Based Monetary Policy." *Review of Economic Dynamics*, 8, 498–527.

Preston, Bruce (2005), "Learning about Monetary Policy Rules when Long-Horizon Expectations Matter." *International Journal of Central Banking*.

Rabin, Matthew (2013), "An Approach to Incorporating Psychology into Economics." *The American Economic Review*, 103, 617–622.

Rabin, Matthew and Dimitri Vayanos (2010), "The Gambler's and Hot-Hand Fallacies: Theory and Applications." *Review of Economic Studies*, 77, 730–778.

Ramey, Valerie A. (2016), "Macroeconomic Shocks and Their Propagation." *Handbook of Macroeconomics*, 2, 71–162.

Sargent, Thomas J. (1999), *The Conquest of American Inflation*. Princeton University Press.

Sawa, Takamitsu (1978), "Information Criteria for Discriminating Among Alternative Regression Models." *Econometrica*, 46, 1273–1291.

Shalizi, Cosma Rohilla (2009), "Dynamics of Bayesian Updating with Dependent Data and Misspecified Models." *Electronic Journal of Statistics*, 3, 1039–1074.

Shimer, Robert (2005), "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95, 25–49.

Sims, Christopher A. (2003), “Implications of Rational Inattention.” *Journal of Monetary Economics*, 50, 665–690.

Stock, James H. and Mark Watson (2011), “Dynamic Factor Models.” *Oxford Handbooks Online*.

Stock, James H. and Mark W. Watson (2002), “Forecasting Using Principal Components From a Large Number of Predictors.” *Journal of the American Statistical Association*, 97, 1167–1179.

Vimercati, Riccardo Bianchi, Martin S. Eichenbaum, and Joao Guerreiro (2022), “Fiscal Policy at the Zero Lower Bound without Rational expectations.” Working paper.

Woodford, Michael (2003), “Imperfect Common Knowledge and the Effects of Monetary Policy.” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

Woodford, Michael (2013), “Macroeconomic Analysis Without the Rational Expectations Hypothesis.” *Annual Review of Economics*, 5, 303–346.

Woodford, Michael (2018), “Monetary Policy Analysis When Planning Horizons Are Finite.” Working paper.