Simple Models and Biased Forecasts[*]

Pooya Molavi[†]

February 21, 2025

This paper characterizes the forecasting biases resulting from agents' reliance on simple models and studies the macroeconomic implications of those biases. It considers agents who can only entertain state-space models with no more than d states, where d is a measure of the complexity of a model. Agents are boundedly rational in that they can only consider models that are too simple to capture the true process, yet they use the best model among those considered. The paper establishes that using simple models adds persistence to forward-looking decisions and increases the comovement among them.

This mechanism can narrow the gap between the business-cycle theory and data. In a standard new neoclassical synthesis model, the assumption that agents use simple models fits the data better than the rational-expectations hypothesis. Simple models significantly dampen the response of the economy to monetary shocks, add persistence to its response to TFP shocks, and generate a positive comovement between consumption and investment in response to investment-demand shocks. Moreover, a Bayesian model comparison finds overwhelming posterior odds in favor of simple models against rational expectations.

[*]This paper subsumes and substantially extends the second half of my job-market paper, titled "Macroeconomics with Learning and Misspecification: A General Theory and Applications." I am grateful to Daron Acemoglu, Marios Angeletos, Larry Christiano, Ian Dew-Becker, Théo Durandard, Marty Eichenbaum, Drew Fudenberg, Laura Gáti, Lars Hansen, Ben Hébert, Ali Jadbabaie, Diego Känzig, Chen Lian, Ben Moll, Alessandro Pavan, Giorgio Primiceri, Bruno Strulovici, Alireza Tahbaz-Salehi, Harald Uhlig, Iván Werning, and numerous seminar and conference participants for their helpful comments.

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

1 Introduction

When faced with the difficult task of forecasting in a complex world, people tend to rely on simple models and past experiences. Yet the rational-expectations hypothesis assumes that agents can forecast the future as if they knew the true model of the economy. The unrealistic nature of the rational-expectations assumption would not be of great concern if the predictions of the standard macro models were robust to alternative specifications of expectations. However, the answers to many important questions in macroeconomics, ranging from the power of forward guidance to the response of the economy to aggregate supply and demand shocks, are sensitive to how agents form their expectations.

This paper analyzes the forecasting biases arising from individuals' reliance on simple models and the macroeconomic implications of those biases. It considers a framework in which agents attempt to forecast future values of a vector of observables based on its past realizations. The true data-generating process features complex intertemporal relationships among observables, but agents can only entertain stochastic models with a bound on their intertemporal complexity. Specifically, they only consider stochastic processes that can be represented using a d-dimensional state variable, where d is a parameter that measures the complexity of an agent's model. Agents are boundedly rational in that they can only entertain models that are too simple to capture the true process, but they find the best d-dimensional approximation to it.

Within this framework, the paper makes three main contributions.

Theoretical contributions. The paper's first theoretical contribution is to characterize the low-dimensional models that best approximate the true data-generating process. In the case where $d = 1$ and the true process satisfies an ergodicity assumption, the optimal approximation is Markovian. Such a model reproduces the true *intratemporal* relationships among observables while reducing the *intertemporal* complexity of the true process. In particular, the simplified model tracks the most persistent component of the vector of observables as well as possible, while forgoing the dynamics of the remaining components. The result generalizes to the $d > 1$ case under additional assumptions.

The paper's second theoretical contribution is characterizing the behavioral implications of using low-dimensional models. First, since simple models only keep track of certain components of the vector of observables, agents' forecasts become unresponsive to changes in other components, thus making their forward-looking choices more *persistent.* Second, different agents—with different payoffs, facing different decisions, and using models with different dimensions—all agree on how to pick the components to track using their models. Since forward-looking decisions of such agents are influenced by a limited set of common components, those decisions *comove* more than they would under rational expectations.

These sharp theoretical results rely on two main assumptions. First, I assume that agents are capable of entertaining *any* stochastic model that has a d-dimensional linear-Gaussian state-space representation. Second, following the literature on model misspecification (e.g., Esponda and Pouzo (2016)), the best model is defined as the stochastic process that minimizes a statistical divergence from the true process. Such pseudo-true models arise naturally as the long-run outcome of Bayesian learning or maximum likelihood estimation by misspecified agents (Berk (1966), White (1982)).

Beyond delivering sharp results, the assumptions allow me to abstract from the difficulty of understanding cross-sectional correlations among observables and focus on time-series (or intertemporal) complexity. In particular, I show that expectations formed using simple models are invariant to linear transformations of observables. This linear invariance property means that the framework's predictions do not depend on the exact specification of the variables included in the vector of observables, only on the information they convey.

Applied contribution. The paper's main applied contribution is quantifying the extent to which the additional persistence and comovement delivered by simple models can improve the empirical fit of business-cycle models. Many standard models have difficulty generating empirically plausible degrees of persistence and comovement in endogenous variables in response to shocks. A common solution pursued in the literature is to introduce auxiliary frictions such as habit formation and adjustment costs (e.g., Christiano, Eichenbaum, and Evans (2005)). However, the resulting DSGE models often require implausibly large frictions to generate realistic business cycles. This paper proposes a plausible alternative that relies on the agents' simplification of their environment's intertemporal complexity.

I consider a standard model economy, which combines elements of the New Keynesian and real business cycle models. The economy features price and wage rigidities, endogenous capital formation subject to neoclassical adjustment costs, and realistic monetary and fiscal policies. However, it does not contain any of the add-ons often introduced in DSGE models to increase persistence and comovement: external habit formation in consumption, investment-adjustment costs, price and wage indexation, endogenous capital utilization, or a monetary policy that responds to the level and growth rate of the output gap.

Without these add-ons, and under rational expectations, the model economy does not generate realistic impulse-response functions (IRFs). The IRFs to productivity shocks are essentially monotonic, not hump-shaped. Small monetary expansions lead to unrealistically large increases in output, consumption, investment, and inflation, a manifestation of the forward-guidance puzzle. Investment-demand shocks lead to a negative comovement between consumption and investment. These observations suggest that standard rational-expectations New Keynesian models cannot provide a satisfactory account of business cycles (absent the DSGE add-ons).

Replacing rational expectations with simple models improves the performance of the model. The responses to productivity shocks become hump-shaped. Expansionary monetary-policy shocks lead to much smaller increases in output, consumption, investment, and inflation—within the range of estimated responses to identified monetary-policy shocks. Investment-demand shocks generate a positive comovement between investment and consumption. Moreover, these improvements do not rely on individuals making large forecasting mistakes; the forecasting gains from increasing d are modest in equilibrium. Overall, this paper's framework emerges as a parsimonious substitute for DSGE add-ons. However, the extent to which simple models can narrow the gap between the business-cycle theory and data is ultimately an empirical question.

To answer this question, I use Bayesian estimation techniques. The parameters of the model economy are estimated separately under rational expectations and under the assumption that agents use one-dimensional simple models. A comparison of marginal likelihoods finds overwhelming evidence (by more than 150 log points) in favor of simple models. Moreover, no single standard DSGE add-on can increase the marginal likelihood by as much as simple models. The key to the empirical success of the framework is its ability to reduce the impact of aggregate demand shocks on inflation and to generate comovement in quantities in response to those shocks. Posterior variance decomposition confirms this explanation. At business-cycle frequencies, demand shocks account for the vast majority of the variance in output, consumption, and investment, and a negligible portion of the variance in inflation or interest rates.

Related Literature. This paper contributes to the literature in macroeconomics on deviations from full-information rational expectations (FIRE)—see Woodford (2013) for a survey. A class of deviations commonly studied in the literature involve imperfect common knowledge of payoff-relevant variables. This can be introduced in a variety of ways such as dispersed information, e.g., Lucas (1972), noisy information, e.g., Orphanides (2003), Nimark (2008), Lorenzoni (2009), Angeletos and La'O (2009), Angeletos and Lian

(2018), and Angeletos and Huo (2021), sticky information, e.g., Mankiw and Reis (2002), or costly attention, e.g., Sims (2003), Woodford (2003), Maćkowiak and Wiederholt (2009), Gabaix (2014), and Alvarez, Lippi, and Paciello (2015). By contrast, this paper assumes that agents have full information but forecast based on a simple, misspecified model of the data-generating process. This alternative modeling assumption leads to qualitatively different predictions: 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. Moreover, this framework simultaneously delivers additional persistence and comovement without introducing any new degrees of freedom for the specification of agents' information structure.

The paper also contributes to the large literature on optimally misspecified models. The common thread in this literature is the assumption that agents can only consider statistical models belonging to a misspecified class but choose the best model among those available. For example, in a Restricted Perceptions Equilibrium (see, e.g., Bray and Savin (1986), Marcet and Sargent (1989), and Evans and Honkapohja (2001)), agents perceive only under-parameterized laws of motion and select parameters via optimal linear projection.[1] Similarly, under the Natural Expectations framework of Fuster, Laibson, and Mendel (2010) and Fuster, Hebert, and Laibson (2012), agents' perceived law of motion conditions only on a few lags of observables, with coefficients determined by linear projection. Likewise, agents in Krusell and Smith (1998) use an optimally misspecified model in which current and future prices depend solely on the first few moments of the wealth distribution. While these papers share a common thread, they differ in the class of misspecified models they consider and in the notion of optimality they adopt. In this paper, I focus on the case where agents can consider *any* linear state-space model with up to d state variables and use a Kullback–Leibler divergence as the measure of goodness-of-fit. The main theoretical contribution of the paper is a characterization of the resulting optimally misspecified models.

The paper also contributes to the literature that quantifies the consequences of bounded rationality using DSGE models. Maćkowiak and Wiederholt (2015) consider a calibrated DSGE model where rational inattention is the only source of slow adjustment and show that the model matches the empirical impulse responses to monetary policy shocks and aggregate technology shocks. Angeletos, Collard, and Dellas (2018) estimate a model in which agents experience shocks to their higher-order expectations and show

[1] A large literature studies various solution concepts related to the Restricted Perceptions Equilibrium and conditions under which they arise from learning dynamics. See Branch (2006) for a survey of Restricted Perceptions Equilibrium and related equilibrium concepts and Eusepi and Preston (2018) for a survey of the learning literature.

that those shocks account for a significant fraction of the business-cycle volatility. Closer to this paper, Milani (2007) and Eusepi and Preston (2011) consider New Keynesian and real business cycle models, respectively, in which agents follow a constant-gain learning rule and Slobodyan and Wouters (2012) consider a medium-scale DSGE model with agents forming expectations using small forecasting models updated by the Kalman filter. They find that adaptive learning fits the data better than rational expectations and that it reduces the need for other frictions. This paper adds to this literature in two ways. First, while many deviations from FIRE considered in the literature increase the persistence of macroeconomic variables, the framework presented here simultaneously generates additional persistence and *comovement*. Second, unlike many boundedly-rational models that have multiple degrees of freedom (such as what agents learn about and how they learn), this framework introduces only one parameter: the dimension d. In my quantitative application, I compare a one-dimensional specification ($d = 1$) with its rational-expectations counterpart ($d = 12$). Both versions use the same set of parameters and are estimated under identical priors.

Finally, in a follow-up paper, Molavi, Tahbaz-Salehi, and Vedolin (2024) use a closely related framework to study the implications of model misspecification for asset prices and returns. They show that constraining the complexity of investors' models leads to return and forecast-error predictability and provides a parsimonious account of several puzzles in the asset-pricing literature.

Outline. The rest of the paper is organized as follows: Section 2 presents the framework and formally defines and discusses the notion of fit used in the paper. Section 3 contains the paper's characterization results for simple models. Section 4 discusses the implications of using simple models for agents' forecasts and choices. Section 5 presents a business-cycle application. Section 6 concludes. Proofs of the main results are relegated to Appendix A. Additional theoretical and empirical results can be found in the Online Appendix.

2 Framework

2.1 Environment

Time is discrete and is indexed by $t \in \mathbb{Z}$. An agent observes a sequence of variables over time and uses her past observations to forecast their future values. I let $y_t \in \mathbb{R}^n$ denote the time-t value of the vector of observables, or simply the *observable*. Vector y_t follows a mean-zero stochastic process $\mathbb{P}$ with the corresponding expectation operator $\mathbb{E}[\cdot]$. I start by taking $\mathbb{P}$ as a primitive, but the process will be an endogenous outcome of agents' actions in the business-cycle application studied in Section 5.

I make several technical assumptions on the true process. First, $\mathbb{P}$ is purely non-deterministic, stationary, and ergodic, and has a finite second moment. Second, there exists a subspace $\mathcal{W}$ of $\mathbb{R}^n$ (possibly equal to $\mathbb{R}^n$ itself) such that y_t is supported on $\mathcal{W}$ with density $\mathbb{f}$.[2] Finally, the true process has finite entropy rate: $\lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[-\log \mathbb{f}(y_1, \ldots, y_t)\right] < \infty$. These assumptions are all quite weak. For instance, they are satisfied if y_t follows a stationary vector ARMA process with Gaussian innovations.

The agent has perfect information about the past realizations of the observable; her time-t information set is given by $\{y_t, y_{t-1}, \ldots\}$. However, she may use a misspecified model to map her information to her forecasts. This model misspecification leads to deviations in the agent's forecasts from those that arise in the rational-expectations benchmark.

2.2 Simple Models

As the paper's main behavioral assumption, I assume that the agent is constrained to use state-space models with a small number of state variables to forecast the vector of observables. She can only entertain models of the form

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

where z_t is the d-dimensional vector of *subjective latent states*, $A \in \mathbb{R}^{d\times d}$, $w_t \in \mathbb{R}^d$ 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 w_t and v_t are independent. While the integer d is a primitive of the model that parameterizes the dimension of the agent's models, matrices A, B, Q, and R are parameters that are determined endogenously by maximizing the fit to the true process. Formally, I define a *d-state model* as a stationary stochastic process over $\{y_t\}_{t=-\infty}^{\infty}$ that has a representation of the form (1) such that (i) the dimension of vector z_t is d, (ii) A is a convergent matrix, (iii) Q is positive definite, and (iv) R is positive semidefinite.[3] I let P^θ denote the d-state model parameterized by the collection of matrices $\theta \equiv (A, B, Q, R)$, let $E^\theta[\cdot]$ denote the corresponding expectation operator, and let Θ_d denote the set of all d-state models. Whenever there is no risk of

[2] This assumption is weaker than the assumption that $\mathbb{P}$ has full support over $\mathbb{R}^n$ because 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.

[3] A matrix is *convergent* if all of its eigenvalues are smaller than one in magnitude. A being convergent and Q being positive definite are sufficient for a model (A, B, Q, R) to define a stationary ergodic process.

confusion, I use the term d-state model to refer both to the stochastic process P^θ for y_t and the parameters $\theta = (A, B, Q, R)$ of its state-space representation.

The integer d captures the agent's sophistication in modeling the stochastic process for the vector of observables, with larger values of d indicating agents who can entertain more complex models. When d is sufficiently large, the agent can approximate the unconditional and conditional second moments of any purely non-deterministic covariance-stationary process arbitrarily well using a model in her set of models. On the other hand, when d is small relative to the number of states required to model the true process, no model in the agent's set of models will provide a good approximation to $\mathbb{P}$. The agent then necessarily ends up with a misspecified model of the true process and biased forecasts—regardless of which model in the set Θ_d she uses to make her forecasts. Characterizing this bias is the focus of the next section of the paper.

2.3 The Notion of Fit

I assume that the agent forecasts using a model in the family of d-state models that provides 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. 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 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^θ denotes the density of P^θ; 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 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 a model's fit. Similarly, $\mathrm{KLDR}(\theta)$ 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 $t \to \infty$ (Shalizi, 2009). The KLDR is also tightly linked to asymptotics of Bayesian learning, as discussed in the following subsection.

Model $\theta \in \Theta_d$ is a *pseudo-true* d-state model if $\mathrm{KLDR}(\theta) \leq \mathrm{KLDR}(\tilde{\theta})$ for all $\tilde{\theta} \in \Theta_d$.[4] If the agent's set of models contains a model θ such that $f^\theta(y_1,\ldots,y_t) = \mathbb{f}(y_1,\ldots,y_t)$ almost

[4]Two closely related notions are the *Restricted Perceptions Equilibrium* (e.g., Evans and Honkapohja (2001)) and the *Berk–Nash Equilibrium* (e.g., Esponda and Pouzo (2016)), which both describe settings in which agents rely on optimally misspecified models. Because the data-generating process here is exogenous,

everywhere and for all t, then any pseudo-true d-state model is observationally equivalent to the true process. The set of models Θ_d is then correctly specified. When no such d-state model exists, $\text{KLDR}(\theta) > 0$ for any model $\theta \in \Theta_d$, and the set of models is misspecified. The following proposition states that pseudo-true models are observationally equivalent to the true process when the set of models is correctly specified:

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

The paper's focus is the misspecified case, where d is small relative to the number of states required to capture the true process. This statement is about d being smaller than the "true d," not necessarily smaller than n, the dimension of y_t. However, it is often natural to also think of d as much smaller than n. Approximating the true process by a pseudo-true d-state model then corresponds to using a parsimonious time-series model to capture the essential features of a large data set. Unless otherwise specified, I assume throughout the paper that $d \le n$. However, the paper's characterization results easily generalize to the $d > n$ case.

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 evolve over time. 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 a probability of one to the set of pseudo-true models.[5]

This result offers an "as if" interpretation of the pseudo-true d-state models. One can assume that the 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, as long as the agent's prior is supported on the set of d-state models, she will forecast the observable in the long run *as if* she were using a pseudo-true d-state model.

I avoid using the equilibrium terminology. Instead, I adopt the econometric term *pseudo-true model* (e.g., Sawa (1978)) to refer to the optimally misspecified stochastic process.

[5]While Berk (1966) only covers i.i.d. observations, Shalizi (2009)'s extension covers hidden Markov models.

Focusing on pseudo-true models allows me to abstract away from learning dynamics and focus on the asymptotic bias caused by misspecification.[6]

3 Characterization of Pseudo-True Models

In this section, I characterize the set of pseudo-true d-state models, beginning with the case where $d = 1$. As a preliminary step, I discuss a useful property of pseudo-true models, which is of independent interest.

3.1 The Invariance Property

I begin with a result that shows the invariance of the pseudo-true d-state models to linear transformations of the observable. Consider an agent who, instead of observing vector $y_t \in \mathbb{R}^n$, observes vector $\tilde{y}_t = Ty_t \in \mathbb{R}^m$, where T denotes an $m \times n$ matrix. As long as T is a rank-n matrix, y_t and $\tilde{y}_t$ convey the exact same information. Thus, one might expect that the agent's beliefs when she observes y_t are consistent with her beliefs when she instead observes $\tilde{y}_t$.

The following definition formalizes the notion that two probability distributions are consistent with each other given a linear transformation of the observable. Let $T \in \mathbb{R}^{m\times n}$ be a matrix and P be a probability distribution over infinite sequences in $\mathbb{R}^n$. The probability distribution over infinite sequences in $\mathbb{R}^m$ induced by T and P is denoted by $T(P)$ and defined as $T(P)(\mathcal{Y}) \equiv P\left(\{y_t\}_{t=-\infty}^{\infty} : \{Ty_t\}_{t=-\infty}^{\infty} \in \mathcal{Y}\right)$ for any measurable set $\mathcal{Y} \subseteq \mathbb{R}^{m\mathbb{Z}}$. If the observable y_t follows the stochastic process $\mathbb{P}$, then its linear transformation $\tilde{y}_t = Ty_t$ follows the transformed process $T(\mathbb{P})$. The following result establishes that transforming the observable by a rank-n matrix leads the set of pseudo-true models to be transformed accordingly:

Theorem 1 (linear invariance)**.** *Suppose $T \in \mathbb{R}^{m\times n}$ is a rank-n matrix. Then P^θ is a pseudo-true d-state model given true model $\mathbb{P}$ if and only if $T\left(P^\theta\right)$ is a pseudo-true d-state model given true model $T(\mathbb{P})$.*

The linear invariance property makes the predictions of the framework invariant to the exact specification of the variables included in the vector of observables. The agent's pseudo-true models and forecasts only depend on the observables' information content,

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

not on how that information is presented. For instance, whether the agent observes the nominal interest rate and the inflation rate or the real interest rate and the inflation rate is immaterial to how she forms her expectations. Likewise, the agent's expectations remain unchanged if the vector of observables is augmented with linear combinations of variables already in her information set.

The theorem suggests that it is without loss to assume that the vector of observables is free of redundant variables. Define the lag-l autocovariance matrix of the true process as follows:

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

When y_t includes redundant variables, the variance-covariance matrix Γ_0 is singular, and the true process is degenerate.[7] In such cases, a lower-dimensional vector $\tilde{y}_t$ and a full-rank matrix T exist such that $\mathbb{E}[\tilde{y}_t \tilde{y}'_t]$ is non-singular and $y_t = T\tilde{y}_t$. Therefore, by Theorem 1, the pseudo-true models given y_t can be found by first finding the pseudo-true models given $\tilde{y}_t$ and then applying transformation T. This observation implies that there is no loss of generality in assuming that the variance-covariance matrix Γ_0 is non-singular and that the agent only considers subjective models with non-singular variance-covariance matrices.[8] I maintain these assumptions throughout the rest of the paper.

3.2 Pseudo-True One-State Models

The agent's pseudo-true one-state forecasts turn out to depend on the true process only through the unconditional variance and the autocorrelation structure of the vector of observables. The autocorrelations are measured by a novel set of objects (to my knowledge), which I refer to as autocorrelation matrices. I define the lag-l *autocorrelation matrix* of the observable under the true process as follows:[9]

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

The concept of autocorrelation matrices naturally extends the idea of autocorrelation functions. If the observable y_t is a scalar, C_l simplifies to the standard autocorrelation

[7] A probability distribution on a space is said to be *degenerate* if it is supported on a manifold of lower dimension.

[8] 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 the 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.

[9] Here and throughout the paper, I follow the usual convention that, for a symmetric positive definite matrix X, the square-root matrix $X^{\frac{1}{2}}$ is the unique symmetric positive definite matrix that satisfies $X^{\frac{1}{2}}X^{\frac{1}{2}} = X$.

function at lag l. However, when the observable is an n-dimensional vector, C_l is an $n \times n$ real symmetric matrix with eigenvalues inside the unit circle.[10] Autocorrelation matrices capture the extent of serial correlation in the vector of observables. When the spectral radius of C_l is close to zero for all l, the process is close to being i.i.d., whereas when the spectral radius of C_l is close to one, then the process is close to being unit root.[11]

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

Theorem 2. *Under any pseudo-true one-state model θ, the agent's s-period-ahead forecast is given by*

$$E_t^{\theta}[y_{t+s}] = a^s(1-\eta)qp' \sum_{\tau=0}^{\infty} a^{\tau}\eta^{\tau} y_{t-\tau}, \tag{4}$$

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

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

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 $\lambda_{max}(\Omega(a,\eta))$, normalized so that $u'u = 1$.

The endogenous variables a, η, p, and q have intuitive meanings. The scalar a represents the persistence of the subjective latent state. If $a = 0$, the subjective state is i.i.d., whereas if $a = 1$, it follows a unit-root process.[12] The scalar η captures the perceived noise in the agent's observations of the subjective state. When η is small, the agent believes recent observations to be highly informative of the value of the subjective state. As a result, her expectations respond more to recent observations and discount old observations more. The vector p determines the agent's relative attention to different components of the vector of observables. When p_i is larger than p_j, the agent puts more weight on $y_{i,t-\tau}$ relative to $y_{j,t-\tau}$ for all τ when forming her estimate of the subjective state. Finally, the vector q captures the relative sensitivity of the agent's forecasts of different observables to changes in her estimate of the subjective state. When q_i is larger than q_j, then a change in the estimated value of the state at time t leads the agent to change her forecast of $y_{i,t+s}$ by more than her forecast of $y_{j,t+s}$ for all s.

[10] See Lemma E.1 in the Online Appendix for a proof.

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

[12] The theorem does not rule out the possibility that $|a| = 1$, in which case the corresponding state-space model might not be stationary ergodic. However, Lemma E.2 in the Online Appendix establishes that any pseudo-true one-state model inherits the stationarity and ergodicity of the true process.

It follows standard Kalman filter results that the agent's forecasts take the form of equation (4) for *some* a, η, p, and q. The substance of the result is rather characterizing the (a, η, p, q) tuple that lead to a model with minimal KLDR from the true process. The theorem suggests a tractable way of computing the pseudo-true one-state forecasts in any stationary and ergodic environment given only the knowledge of the true autocorrelation matrices.

The next result characterizes the perceived variance-covariance matrix of the observable under the pseudo-true one-state models:

Theorem 3. *Given any pseudo-true one-state model* θ*, the subjective variance-covariance of the vector of observables,* $E^{\theta}[y_t y_t']$*, coincides with the true variance-covariance matrix,* $\Gamma_0 \equiv \mathbb{E}[y_t y_t']$.

The theorem hinges on two main assumptions: First, there are no constraints on the agent's set of models other than the bound on the number of subjective state variables; put differently, matrices A, B, Q, and R of representation (1) are unrestricted other than the constraint on their dimension. This flexibility allows the agent to represent any cross-sectional correlation pattern by an appropriate selection of matrices A, B, Q, and R. Second, the agent uses a model that minimizes the KLDR from the true process. This leads her to a set of such matrices that perfectly capture the true cross-sectional correlations.

Theorems 2 and 3 fully characterize the pseudo-true one-state models in terms of the true variance-covariance matrix Γ_0 and the tuple (a, η, p, q), which in turn only depends on Γ_0 and the true autocorrelation matrices $\{C_l\}_{l=1}^{\infty}$. Any unconditional or conditional moment of the pseudo-true one-state model can, in turn, be found in terms of Γ_0 and (a, η, p, q).

3.3 Pseudo-True One-State Models Under Exponential Ergodicity

The pseudo-true one-state models can be found in closed form given a class of true stochastic processes that naturally arise in applications. The appropriate class turns out to be the following:

Definition 1. A stationary ergodic process $\mathbb{P}$ is *exponentially ergodic* if $\rho(C_l) \le \rho(C_1)^l$ for all $l \ge 1$, where $\rho(C_l)$ denotes the spectral radius of C_l.

Exponential ergodicity is stronger than ergodicity. Ergodicity requires that the serial correlation at lag l decays to zero as $l \to \infty$. Exponential ergodicity requires the rate of decay to be faster than $\rho(C_1)$. Although exponentially-ergodic processes only constitute a subset of the class of stationary ergodic processes, many standard processes are exponentially ergodic. For instance, the vector of observables follows an exponentially-ergodic process if it is a spanning linear combination of n independent AR(1) shocks.

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

Theorem 4. *Suppose the true process is exponentially ergodic. Under any pseudo-true one-state model θ, the agent's s-period-ahead forecast is given by*

$$E_t^{\theta}[y_{t+s}] = a^s q p' y_t, \tag{5}$$

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 $q = \Gamma_0^{\frac{1}{2}} u$.

A remarkable feature of the characterization in Theorem 4 is that the agent's forecasts only depend on the last realization of the observable (and not its lags). In other words, the pseudo-true one-state model is *Markovian* if the true process is exponentially ergodic. This property might come as a surprise in light of the fact that, in the correctly-specified case, forecasts obtained using the stationary Kalman filter generically use the entire history of the observable. The seeming discrepancy between the two results is due to misspecification of the agent's set of models in Theorem 4, as illustrated by the following example:

Example 1. Suppose the observable is scalar and follows an AR(∞) process: $y_{t+1} = \sum_{\tau=1}^{\infty} \phi_\tau y_{t+1-\tau}$.[13] It is then immediate that the one-step-ahead forecast of the observable under the true, correctly-specified model is given by

$$\mathbb{E}_t[y_{t+1}] = \sum_{\tau=1}^{\infty} \phi_\tau y_{t+1-\tau}.$$

Contrast this with what an agent can do when she is constrained to use (misspecified) one-state models. Under any such model θ, the agent's one-step-ahead forecast takes a similar form:

$$E_t^{\theta}[y_{t+1}] = \sum_{\tau=1}^{\infty} \alpha_\tau y_{t+1-\tau}.$$

[13] Such a representation exists for generic processes in the class of mean-zero, purely non-deterministic, and stationary processes.

However, the restriction to one-state models constrains coefficients $\{\alpha_\tau\}_{\tau=1}^{\infty}$ to be given by $\alpha_\tau = (1-\eta)a^\tau \eta^{\tau-1}$ for some $a \in [-1,1]$, some $\eta \in [0,1]$, and all τ. Therefore, the pseudo-true one-state model is the model that picks $\{\alpha_\tau\}_{\tau=1}^{\infty}$ to minimize the KLDR subject to the constraint that $\alpha_\tau = (1-\eta)a^\tau \eta^{\tau-1}$ for all τ. The agent wants to set α_τ to a value that is related to the correlation of y_{t+1} and $y_{t-\tau}$, but the constraint prevents her from fine-tuning the $\{\alpha_\tau\}_{\tau=1}^{\infty}$ coefficients. When the true process is exponentially ergodic, y_{t+1} is much more correlated with y_t than it is with lags of y_t. Then, the best such a constrained agent can do is to fine-tune the coefficient of y_t and entirely disregard its lags. In other words, the constrained minimizer of the KLDR is Markovian even though the unconstrained minimizer is not.

The next example shows the use of Theorem 4 in the context of a canonical process:

Example 2. Suppose the true process $\mathbb{P}$ has the following representation:

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

where $\epsilon_t \sim \mathcal{N}(0, \Sigma)$, $F = \operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\Sigma = \operatorname{diag}\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\right)$, $H \in \mathbb{R}^{n\times n}$ 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. Therefore, Theorem 4 can be used to characterize the pseudo-true one-state forecasts. The persistence, 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 $q = (H'VH)^{\frac{-1}{2}} H' V e_1$, where $V \equiv (I - F^2)^{-1}\Sigma$ is the variance-covariance matrix of f_t and e_1 denotes the first coordinate vector.

The agent's forecasts take a particularly simple form when H is the identity matrix, i.e., $y_{it} = f_{it}$ for $i = 1, \ldots, n$. Then, p and q are both multiples of the first coordinate vector e_1, and the agent's forecasts simplify to

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

The agent's forecast of the most persistent element of the vector of observables coincides with its rational-expectations counterpart, but she forecasts every other element of the observable as if it were i.i.d.

The example illustrates that the agent exhibits a form of *persistence bias.* She forecasts the most persistent component of the vector of observables as accurately as under rational expectations but misses the dynamics of the other components. The intuition for the result is easiest to see when the most persistent component is close to being unit root. In that case, poorly tracking the most persistent component would lead to persistent mistakes in the agent's 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 component close to unit root as best possible, even if doing so results in errors in forecasting the other components. In Section 4.1, I generalize the insight of this example by formally establishing persistence bias in a more general context.[14]

This example can be generalized by relaxing the assumption that matrices F and Σ are diagonal and allowing for non-Gaussian innovations. The key requirement for the process to be exponentially ergodic is that matrix H in representation (6) is full rank. This assumption can be seen as a full-information (or spanning) assumption. If the agent observes an observable of the form (6) with a full-rank matrix H, then she has enough information to forecast the observable as well as in the full-information rational-expectations benchmark—even if she fails to do so due to the constraint on her set of models. See Appendix E of Molavi (2024) for a detailed discussion of this generalization and a sufficient condition for exponential ergodicity.

3.4 Pseudo-True d-State Models

I end this section by discussing how the insights from the $d = 1$ case generalize when $d > 1$. To characterize the pseudo-true d-state models, one needs to find models $\theta = (A, B, Q, R)$ that minimize the KLDR from the true process. Doing so requires minimizing a non-convex function over a non-compact set, consisting of all the matrices A, B, Q, and R of appropriate dimensions. This problem does not lend itself to an analytical solution without further restrictions.

To address this issue, I restrict the models the agent considers to be Markovian. A d-state model θ is *Markovian* if P^θ satisfies the Markov property, i.e., $P^\theta(y_{t+1}|y_t, y_{t-1}, \dots) = P^\theta(y_{t+1}|y_t)$. An agent who believes the observable follows a Markovian d-state model believes that (a) the current realization of the observable contains all the information required for forecasting, and (b) all the relevant information contained in y_t can be sum-

[14] Bidder and Dew-Becker (2016) and Dew-Becker and Nathanson (2019) propose an alternative reason agents might focus on tracking the most persistent components of a payoff-relevant variable. Bidder and Dew-Becker (2016) show that long-run risk is the worst case scenario for ambiguity-averse agents. Dew-Becker and Nathanson (2019) show that, as a result, ambiguity-averse agents will learn most about dynamics at the lowest frequencies.

marized by a d-dimensional state variable. The following proposition provides a necessary and sufficient condition for a model to be Markovian:

Proposition 2. *Let $Var_t^\theta(y_{t+1})$ denote the variance-covariance matrix of y_{t+1} given model θ and conditional on the history $\{y_\tau\}_{\tau\leq t}$ of the observable, and let $Var^\theta(y_{t+1}|z_t)$ denote the corresponding variance-covariance matrix conditional on the time-t realization of the subjective latent state. $Var_t^\theta(y_{t+1}) \geq Var^\theta(y_{t+1}|z_t) = B'QB + R$ for any d-state model θ, with equality if and only if θ is Markovian.*[15]

The proposition highlights an intuitive property of Markovian models. Note that the agent can observe the history $\{y_\tau\}_{\tau\leq t}$ of the observable but not the subjective latent state z_t. The first part of the proposition shows that the agent cannot forecast any better than if she knew the realization of the latent subjective state. In other words, the forecast error given z_t provides a lower bound on the forecast error given $\{y_\tau\}_{\tau\leq t}$. The second part of the proposition shows that the agent can achieve this lower bound when her model of the world is Markovian. She can then forecast as well as an agent who knows the latent subjective state, because all the relevant information in the latent state can be extracted from the realized history of the observable. Markovian models can thus be seen as models that feature *full information*.

Markovian models constitute only a subset of the class of all state-space models of a given dimension. However, the pseudo-true one-state models are Markovian when the true process is exponentially ergodic, as shown in Theorem 4. Even with the flexibility to choose non-Markovian models, an agent who is attempting to minimize the KLDR from an exponentially ergodic process settles on a Markovian model. 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 Markovian models as an assumption and characterizing the resulting pseudo-true models. A Markovian d-state model θ is a *pseudo-true Markovian d-state model* if $\text{KLDR}(\theta) \leq \text{KLDR}(\tilde{\theta})$ for any Markovian d-state model $\tilde{\theta}$.

Pseudo-true Markovian models have a number of appealing properties. They satisfy a version of the linear-invariance result of Theorem 1. They share Bayesian and quasi-maximum-likelihood learning foundations with other pseudo-true models. Perhaps most importantly, they can be fully characterized in closed-form in some useful cases:

Theorem 5. *Suppose either $d = 1$ or the lag-one autocovariance matrix is symmetric. Then the following statements hold:*

[15]I use the usual convention that $X \geq Y$ for symmetric positive semidefinite matrices X and Y if $X - Y$ is positive semidefinite.

(a) *Under any pseudo-true Markovian d-state model θ, the agent's s-period-ahead forecast is given by*

$$E_t^\theta[y_{t+s}] = \sum_{i=1}^{d} a_i{}^s q_i p_i' y_t, \tag{7}$$

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

(b) *Under any pseudo-true Markovian d-state model θ, the subjective variance-covariance of the vector of observables, $E^\theta[y_t y_t']$, coincides with the true variance-covariance matrix, $\Gamma_0 \equiv \mathbb{E}[y_t y_t']$.*

The result shows that the insights from the analysis of one-state simple models carry over to d-state ones. In particular, agents who are restricted to Markovian d-state models exhibit a form of persistence bias. They focus on perfectly forecasting the d most persistent components of the vector of observables at the expense of the other components. Moreover, agents who are constrained to use Markovian d-state models uncover the true variance-covariance matrix of the observable.

4 Behavioral Implications

In this section, I apply the characterization results from the previous section to develop the behavioral implications of the framework. Throughout the section, I maintain the assumption that at least one of the following is satisfied for every agent: (a) the agent is constrained to use one-state models and the true process is exponentially ergodic; (b) the agent is constrained to use Markovian one-state models; or (c) the agent is constrained to use Markovian d-state models and the lag-one autocovariance matrix, Γ_1, is symmetric.

To elaborate on the behavioral implications of the framework, I embed it in a reduced-form economy. Consider a finite set of agents, indexed by $j = 1, \ldots, J$. In every period t, each agent j takes a purely forward-looking decision x_{jt}, which depends on her forecasts via the best-response function

$$x_{jt} = E_{jt}\left[\sum_{s=1}^{\infty} c_{js}' y_{t+s}\right], \tag{8}$$

where $y_t \in \mathbb{R}^n$ is as before the vector of observables, $E_{jt}[\cdot]$ denotes agent j's subjective forecasts, and $c_{js} \in \mathbb{R}^n$ are preference parameters satisfying $\sum_{s=1}^{\infty} \|c_{js}\|_2 < \infty$ for all j.[16] I continue to take the true process $\mathbb{P}$ as a primitive of the economy and assume that agent j can entertain state-space models with no more than d_j states.

The reduced-form specification in (8) allows the derivation of sharp theoretical results, which highlight the role of simple models and biased forecasts and are independent of the specifics of agents' decision problems. These results are valid up to first order for purely forward-looking decisions that depend non-linearly on the forecasts of the observable. They also hold arbitrarily well when decisions are sufficiently forward-looking (e.g., when the discount factor is close to one). In the next section, I further develop the implications of the general framework in the context of a microfounded general equilibrium macro model.

4.1 Persistence Bias

Decomposing the observable into its more and less persistent components will be useful for the subsequent discussions:

Proposition 3. *Let a_i denote the ith largest eigenvalue of the first autocorrelation matrix, C_1, in magnitude, and let u_i denote the corresponding eigenvector, normalized such that $u_i' u_k = \mathbb{1}_{\{i=k\}}$ for all i and k. The observable can be decomposed as follows:*

$$y_t = \sum_{i=1}^{n} y_t^{(i)} q_i, \tag{9}$$

where $y_t^{(i)} \equiv p_i{}' y_t$, $p_i \equiv \Gamma_0^{\frac{-1}{2}} u_i$, $q_i \equiv \Gamma_0^{\frac{1}{2}} u_i$, u_i is as in Theorem 5, and scalars $y_t^{(i)}$ all have unit variance. If ρ_i denotes the lag-one autocorrelation of $y_t^{(i)}$, then $|\rho_1| \geq |\rho_2| \geq \cdots \geq |\rho_n|$.

This proposition represents the observable in terms of the basis vectors $\{q_i\}_{i=1}^n$, with $y_t^{(i)}$ denoting the components (or coordinates) of y_t with respect to this basis. The components of y_t are sorted by their persistence, with $y_t^{(1)}$ representing the *most persistent component* and $y_t^{(n)}$ the *least persistent component* of the observable. This decomposition is valid for arbitrary stationary stochastic processes and is independent of agents' forecasting and decision problems. However, the way agents' choices respond to changes in the observable neatly aligns with the decomposition in (9). This is shown in the following corollary of Theorems 4 and 5:

[16]The assumption that each agent takes a single action is without loss of generality. The analysis would be identical if one instead assumed that agent j makes multiple choices in each period, with the kth action of agent j given by $x_{jkt} = E_{jt}\left[\sum_{s=1}^{\infty} c'_{jks} y_{t+s}\right]$.

Corollary 1 (persistence bias). *Agent j's time-t forecasts and forward-looking actions only respond to changes in the d_j most persistent components of y_t.*

Agents who use pseudo-true d-state models treat the most persistent and least persistent components of y_t in qualitatively different ways. A change in the current value of the observable can be decomposed into changes in the components $y_t^{(i)}$ of y_t. Agents do not change their forecasts in response to changes in the least persistent components of y_t. Consequently, their forward-looking actions also remain unresponsive to current changes in these less persistent components.

It is worth noting that agents' forecasts and actions are unresponsive to changes in the less persistent components of the observable only on impact. In general, different components of y_t do not evolve independently. Therefore, a change in the current value of $y_t^{(i)}$ could lead to changes in the values of $y_{t+s}^{(j)}$ for some $j \neq i$ and $s > 0$. This can result in a delayed response of agents' forecasts and actions to changes in the observable's less persistent components.

4.2 Increased Comovement

Constraining agents to use simple models increases the comovement between their forward-looking choices. The argument for this prediction is best seen by considering agents j and k, both of whom are constrained to use one-state models. Because of persistence bias, the agents' time-t actions can be written as time-invariant linear functions of the observable's most persistent component. More specifically, $x_{jt} = g_j^{(1)} y_t^{(1)}$ and $x_{kt} = g_k^{(1)} y_t^{(1)}$, where $g_j^{(1)}$ and $g_k^{(1)}$ are constants that depend on the true process and the agents' preferences. Thus, one agent's actions can be expressed as a constant multiple of the other agent's actions. In other words, the agents' actions comove perfectly. The following proposition formalizes and extends this conclusion:

Proposition 4. *Let $D \equiv \max_j d_j$ denote the largest value of d_j among agents and $x_t \equiv (x_{jt})_j \in \mathbb{R}^J$ denote the vector containing agents' time-t actions. Given generic true processes, x_t has the factor structure*

$$x_t = G y_t^{(1:D)},$$

where G is a $J \times D$ matrix of loadings and $y_t^{(1:D)}$ is the D-dimensional vector consisting of the D most persistent components of the observable.[17]

[17] The result requires all pseudo-true models of a given dimension to be observationally equivalent, a condition that holds for generic true processes.

The proposition establishes that the J-dimensional vector of all the forward-looking actions of all the agents in the economy moves with the D factors collected in $y_t^{(1:D)}$. The number of factors depends solely on the complexity of agents' models, while the composition of these factors depends on the properties of the true process. The loadings of actions on different factors depend on the preference parameters c_{js}. If D is much smaller than J, then a large number of actions comove with movements in a small number of factors.

Agents' actions exhibit comovement not only among those using models with the same value of d but also among those using models of different dimensions. To see the intuition for this result, consider agents j and k who use models of dimensions d_j and $d_k > d_j$, respectively. While the agents disagree on the number of state variables needed to forecast the observable, they agree on what d_j of those state variables ought to be. The d_j states used by agent j are a subset of the d_k states used by agent k (up to linear transformations). This strong form of comovement is a unique prediction of the framework studied in this paper. It relies on the fact that pseudo-true d-state models rank the components of y_t consistently across d: As d increases, pseudo-true forecasts condition on additional components of y_t but without altering the components already being used to forecast.

A low-dimensional factor structure is one natural expression of comovement. Another commonly used comovement measure is the Pearson correlation coefficient between two variables. The following corollary of Proposition 4 shows that constraining any two agents to one-dimensional models increases the correlation between their actions:

Corollary 2. *Consider actions j and k, both of the form (8), taken by agents j and k with $d_j = d_k = 1$. Generically,*

$$1 = \left|Corr\left(x_{jt}^{1d}, x_{kt}^{1d}\right)\right| > \left|Corr\left(x_{jt}^{RE}, x_{kt}^{RE}\right)\right|,$$

where x_{jt}^{1d} and x_{jt}^{RE} denote agent j's time-t action when using a pseudo-true one-state model and the true model, respectively.

The time-t actions of agents using a pseudo-true one-state model depend solely on the current realization of $y_t^{(1)}$, the most persistent component of y_t. Consequently, an econometrician who analyzes those actions will conclude that the actions are driven by a single shock to the economy. This conclusion holds regardless of the specifics of preferences, technology, or market structure. However, the single shock recovered by the econometrician is not a true shock. It is an endogenous index whose statistical properties depend on the primitives of the economy, the stochastic properties of the shocks that hit it, and the parameters of policy rules.

5 A Business-Cycle Application

In this section, I use this paper's framework to study how bounded rationality changes the response of an economy to supply, demand, and policy shocks and the propagation of those shocks to endogenous variables. I do so in the context of a standard business-cycle model economy, which combines elements of the New Keynesian and real business cycle models. This exercise demonstrates that the macroeconomic model's empirical fit is improved when bounded rationality is introduced in the form of dimensionality reduction. Moreover, simple models can serve as a parsimonious substitute for add-ons, such as external habit formation, investment-adjustment costs, and endogenous capital utilization, which are used to increase the persistence and comovement of endogenous variables.

5.1 The Model Economy

The model economy is a standard New Keynesian economy with price and wage rigidities and endogenous capital formation. Alternatively, the model can be viewed as a DSGE model à la Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2007), and Justiniano, Primiceri, and Tambalotti (2010) but without the following add-ons: (i) external habit formation in consumption, (ii) investment-adjustment costs, (iii) price and wage indexation, (iv) endogenous capital utilization, (v) a monetary policy that responds to the level and growth rate of the output gap. The first three add-ons increase the persistence of consumption, investment, inflation, and wages. The last two enhance the comovement properties of the economy.

Primitives. The economy is populated by seven groups of agents: final-good producers, intermediate-goods producers, investment firms, employment agencies, households, labor unions, and the government. In what follows, I briefly describe each group's problem—for more details, see Online Appendix B.

The final good is produced by competitive firms by combining a continuum of intermediate goods according to a CES production function with elasticity of substitution λ_p. Intermediate goods are produced by monopolists using a Cobb–Douglas production function, with capital share α, labor share $1-\alpha$, and a fixed cost of production F. The labor productivity and fixed costs both grow at the deterministic rate γ. Intermediate-goods producers are subject to nominal frictions à la Calvo, with ξ_p denoting the degree of nominal rigidity. The capital stock of the economy is owned by competitive investment firms. There are neoclassical capital adjustment costs à la Hayashi (1982), where the convexity of the adjustment-cost function is governed by ς_k. Capital depreciates at rate δ.

There is a continuum of households, each of which is a monopolistic supplier of a specialized type of labor. A competitive employment agency combines specialized labor into a homogeneous labor input using a CES function with elasticity λ_w. Households supply labor, consume the final good, and save in a short-term nominal government bond. Households have standard preferences, with a unit intertemporal elasticity of substitution and an inverse Frisch elasticity equal to ν. Wages are set by a continuum of labor unions, each representing a household. Wages are subject to nominal rigidities with Calvo parameter ξ_w. However, as is common in the literature, I assume that a competitive insurance agency fully insures households against fluctuations in their labor income resulting from nominal frictions.

Monetary policy sets the nominal interest rate following a Taylor rule with smoothing parameter ρ_R and coefficient ϕ_π on the inflation rate. Government spending is exogenous. In the baseline specification, I assume that government spending grows at the same rate as the deterministic growth rate of GDP. The government finances its spending by issuing short-term nominal bonds and levying lump-sum taxes on households. Taxes follow a tax rule that ensures that the real value of public debt grows at the deterministic growth rate of the economy.

To focus on the core mechanisms, I initially consider a version of the model with only three shocks: a total-factor productivity (TFP) shock, an investment-demand shock, and a monetary-policy shock. All three shocks follow AR(1) processes with Gaussian innovations. The TFP shock is the main supply shock in DSGE models, while the investment-demand shock is the demand shock that explains the largest variance shares of real variables at business-cycle frequencies (Justiniano et al., 2010). Monetary-policy shocks contribute little to the variance of nominal and real variables at business-cycle frequencies. However, they have clear empirical counterparts that can be identified using vector autoregression (VAR), narrative, and high-frequency approaches. Comparing model-implied impulse-response functions (IRFs) to monetary-policy shocks with their empirical counterparts provides a powerful test of the model economy's internal propagation mechanism (Christiano et al., 2005). Later, I enrich the economy with a full suite of standard DSGE shocks, estimate it using Bayesian techniques—both under simple models and under rational expectations—and perform Bayesian model selection.

Equilibrium. The analysis proceeds in two steps. The first is to characterize the *temporary equilibrium*, which imposes individual optimality and market clearing but not rational expectations.[18] The second step is to supplement the temporary equilibrium with the model of expectation formation and characterize the resulting (full) equilibrium.

The first step of the analysis is relatively straightforward. I start by deriving agents' first-order optimality conditions. I then characterize the balanced-growth path along which inflation is constant and equal to the monetary authority's inflation target, and output, consumption, investment, government spending, capital stock, real wages, and the public debt all grow at rate γ, the deterministic growth rate of labor productivity. Finally, I log-linearize the optimality conditions around the balanced-growth path. The resulting temporary-equilibrium conditions can be found in Online Appendix C.

I next describe agents' subjective expectations. For simplicity, I assume that households, investment firms, intermediate-goods producers, and labor unions all face identical constraints on the models they can entertain, thus ending up with identical subjective expectations. Every agent has perfect foresight about the balanced-growth path of the economy. Agents have full information about the log-deviations of all endogenous and exogenous variables from the balanced-growth path. In particular, agents' time-t information set is given by the history $\{\omega_s\}_{s\le t}$ of vector

$$\omega_s \equiv \left(\hat{a}_s, \hat{m}_s, \hat{\mu}_s, \hat{k}_s, \hat{y}_s, \hat{x}_s, \hat{c}_s, \hat{i}_s, \hat{L}_s, \hat{\rho}_s, \hat{\pi}_s, \hat{R}_s, \hat{w}_s, \hat{\tau}_s\right)',$$

consisting of the time-s realizations of TFP, monetary-policy, and investment shocks as well as log-deviations of capital stock, GDP, pre-tax income, consumption, investment, hours, rental rate of capital, inflation, nominal interest rate, wages, and taxes.[19] Instead of imposing rational expectations, I assume that agents are constrained to use one-dimensional state-space models of the form (1) to forecast future values of ω.

The equilibrium definition is straightforward. An equilibrium consists of a stochastic process $\mathbb{P}^*$ for $\{\omega_t\}_t$ and a model θ^* for agents such that (i) $\mathbb{P}^*$ is derived from market-clearing conditions and optimal behavior by all agents in the economy given subjective model θ^*, and (ii) θ^* is a pseudo-true one-state model given the stochastic process $\mathbb{P}^*$.[20]

[18] The notion of temporary equilibrium goes back to Grandmont (1977). See Woodford (2013) for a discussion of temporary equilibria in the context of modern monetary models and Farhi and Werning (2019) for an application to heterogeneous-agent New Keynesian economies.

[19] In equilibrium, some elements of ω_t are linear combinations of other elements of ω_t. By the linear invariance result, dropping the redundant variables from ω_t does not change any of the equilibrium outcomes. Similarly, I can add additional variables to ω_t that, in equilibrium, are linearly dependent on variables already included in ω_t without changing anything.

[20] This equilibrium notion is closely related to the Restricted Perceptions Equilibrium (e.g., Branch (2006)). However, whereas the Restricted Perceptions Equilibrium selects the best misspecified model via an optimal

5.2 Impulse-Response Functions

I begin my analysis of the business-cycle economy by studying its impulse-response functions (IRFs) to TFP, investment-demand, and monetary-policy shocks—both under rational expectations and under the assumption that agents use pseudo-true one-state models. I use the same calibration for primitive parameters and shock processes in both variants. This makes it possible to transparently see how bounded rationality changes the internal propagation mechanism of the economy.[21]

The model parameters are calibrated as follows: A period represents a quarter. I set $\beta = 0.99$, $\gamma = 1.005$, $\delta = 0.025$, $\nu = 1$, $\alpha = 1/3$, $\lambda_p = 0.5$, $\rho_R = 0.8$, and $\phi_\pi = 1.5$. I set the steady-state ratio of government spending to GDP and public debt to GDP both equal to zero. The persistence parameters of the shocks are calibrated to $\rho_a = 0.95$ for TFP, $\rho_m = 0.4$ for monetary policy, and $\rho_\mu = 0.7$ for investment-demand shocks. The standard deviations of shocks are set to $\sigma_a = 1$ for TFP, $\sigma_m = 0.5$ for monetary policy, and $\sigma_\mu = 2$ for investment-demand shocks. These values are in the ballpark of both the values chosen in the literature and those obtained from Bayesian estimation of the model. Three parameters—price rigidity ξ_p, wage rigidity ξ_w, and capital-adjustment cost ς_k—greatly influence the persistence of endogenous variables and the propagation of shocks. I assume flexible wages and set $\xi_p = 0.6$ and $\varsigma_k = 0.5$. These conservative values are chosen to demonstrate that the model can generate realistic IRFs without relying on the counterfactually large degrees of nominal and real frictions often assumed in business-cycle models.

There are no free parameters for agents' expectations (other than d, which I have set equal to one). Agents' models, beliefs, and forecasts are all pinned down by structural parameters of the economy and the stochastic processes of the shocks. In equilibrium, agents' forecasts of elements of vector ω_s are given by

$$E_t[\omega_{t+s}] = a^{*s} q^* p^{*\prime} \omega_t,$$

where the perceived persistence is given by

$$a^* = 0.997,$$

linear projection, here optimality is defined by minimizing the Kullback–Leibler divergence rate from the true process.

[21] Figure D.1 in the Online Appendix plots the posterior on IRFs from a Bayesian estimation of the fully flexible version of the model (enriched with a full suite of DSGE shocks). The posterior concentrates its mass on parameter configurations with implied IRFs that are similar to those obtained here.

and the relative-attention vector p^* and the relative-sensitivity vector q^* are given by[22]

$$
\begin{array}{rcccccccccccccc}
 & & \hat{a} & \hat{m} & \hat{\mu} & \hat{k} & \hat{y} & \hat{x} & \hat{c} & \hat{\imath} & \hat{L} & \hat{\rho} & \hat{\pi} & \hat{R} & \hat{w} \\
p^{*\prime} & = & (0.01 & 0.02 & 0.01 & 0.92 & 0.07), & & & & & & & & \\
q^{*\prime} & = & (0.47 & -0.00 & 0.19 & 1.00 & 1.02 & 1.08 & 1.08 & 0.85 & -0.19 & -0.31 & 0.14 & 0.13 & 0.88).
\end{array}
$$

Agents forecast following a three-step procedure. First, they project the vector of observables on the relative-attention vector p^* to form their estimate $\hat{z}_t \equiv p^{*\prime}\omega_t$ of the current value of the latent subjective state—I refer to $\hat{z}_t$ as agents' "nowcast." Then, they form their forecasts of future values of the subjective state given its perceived persistence: $E_t[z_{t+s}] = a^{*s}\hat{z}_t$. Finally, they multiply their forecasts of the subjective state by the relative-sensitivity vector q^* to form their forecasts of observables: $E_t[\omega_{t+s}] = q^* E_t[z_{t+s}]$.

Agents perceive the subjective state as highly persistent but not unit root ($a^* = 0.997$). The nowcast is much more sensitive to changes in the capital stock ($p_k^* = 0.92$) than to changes in GDP ($p_y^* = 0.07$), and it barely responds to innovations in the three exogenous shocks ($|p^*| \leq 0.02$). This is a manifestation of persistence bias: In equilibrium, the capital stock is more persistent than output and TFP, monetary-policy, and investment-demand shocks. Agents' forecasts of capital stock, output, gross income, consumption, investment, and the real wage move almost one-for-one with changes in agents' nowcast ($|q^*| \geq 0.88$), whereas their forecasts of TFP, investment-demand shock, hours, and the rental rate of capital exhibit much less sensitivity. The observables whose forecasts are least sensitive to new information ($|q^*| \leq 0.14$) are the nominal variables: Agents' expectations of the monetary-policy shock, inflation, and nominal interest rate are somewhat "anchored" to their steady-state values.

In equilibrium, switching to a higher-dimensional model yields only modest improvements in forecasting accuracy. To quantify this improvement, I fix the aggregate laws of motion at their equilibrium values under simple models. I then compute the reduction in the present discounted value of mean squared forecast errors for key prices when an agent unilaterally switches to the true, equilibrium model. The analysis shows that using the true model improves forecast accuracy by about 2% for the rental rate of capital, 3% for inflation, 6% for the nominal interest rate, and 15% for the real wage. In other words, the individual gain from adopting more complex models is relatively small.[23] This observa-

[22]Vector p^* is identified only up to a set of linear transformations. Since ω_s contains redundant variables, $\tilde{p}'\omega_s = p^{*\prime}\omega_s$ for all s and a set of vectors $\tilde{p}$ belonging to a subspace. By the linear invariance result, all such vectors lead to identical forecasts and actions for agents at all times.

[23]Computing the welfare costs of bounded rationality requires a second-order approximation to the temporary equilibrium conditions due to the issues highlighted by Kim and Kim (2003) and Woodford (2002) and is beyond the scope of the current analysis.

tion aligns with the findings of Akerlof and Yellen (1985) that even insignificant deviations from full rationality can have first-order implications for equilibrium outcomes.

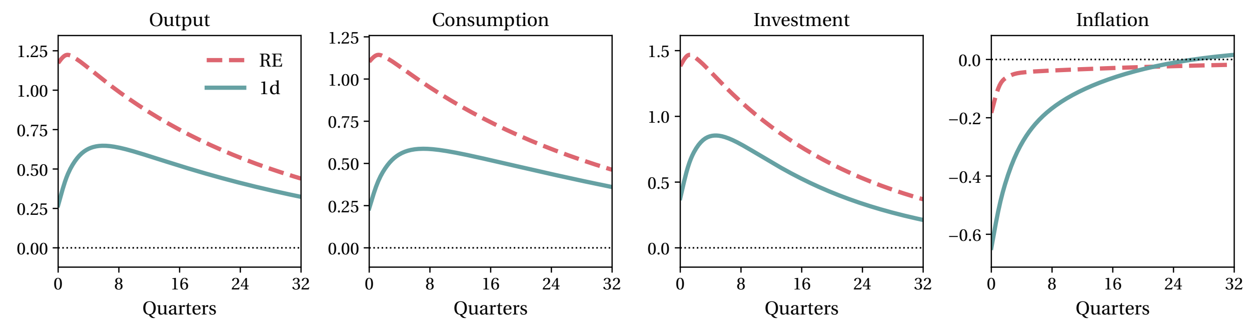

Figure 1. Impulse-response functions to a positive TFP shock.

Notes: Baseline calibration. One-dimensional simple models in solid blue. Rational expectations in dashed red. Responses to a one-percent increase in TFP. Output, consumption, and investment in percents; inflation in percentage points.

Figure 1 plots the IRFs to a positive TFP shock. With one-dimensional simple models, the IRFs of real variables mimic the hump-shaped responses found in models with features that serve to increase the sluggishness of aggregate variables, such as consumption-habit formation and investment-adjustment costs. The response of output on impact is 77% smaller with simple models than under rational expectations (RE). The corresponding figures for consumption and investment are 79% and 73%, respectively. The responses of real variables peak after one quarter under RE. With simple models, the peak ranges from six quarters after impact for consumption to eight quarters after impact for investment. Simple models thus provide an account of the hump-shaped responses of aggregate variables to TFP shocks in empirical studies, e.g., Ramey (2016, pp. 135–151), which does not rely on auxiliary frictions. The fact that quantities respond less with simple models requires more of the increase in TFP to be absorbed by a fall in prices. The result is a larger decrease in inflation in response to the increase in TFP. Nevertheless, the response of inflation is more muted and more transitory than those of real variables (both under RE and with simple models).

Figure 2 plots the IRFs to an expansionary monetary-policy shock. Agents' use of simple models dampens the responses of real variables by about 98% on impact. However, here, simple models also dampen the response of inflation on impact by 98.6%. Bounded rationality also increases the persistence of responses to monetary-policy shocks. The response of all aggregate variables decrease monotonically and rapidly under rational expectations. In contrast, the responses are hump-shaped and significantly more persistent when agents use simple models.

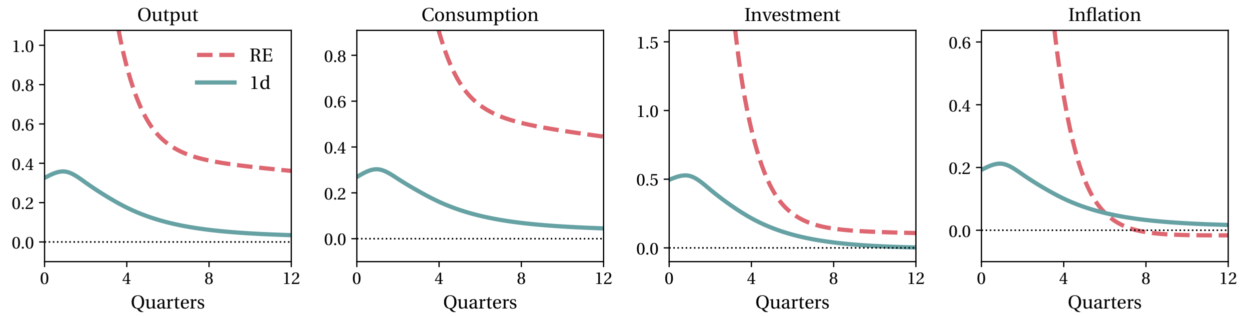

Figure 2. Impulse-response functions to an expansionary monetary-policy shock.

Notes: Baseline calibration. One-dimensional simple models in solid blue. Rational expectations in dashed red. The shock is normalized to reduce the nominal rate by 25 basis points on impact in each variant. Output, consumption, and investment in percents; inflation in percentage points.

The responses of aggregate variables to monetary-policy shocks are counterfactually strong under rational expectations. This is because changes in future real interest rates have the same effect on current output as changes in current real interest rates. With nominal rigidities and rational expectations, monetary-policy shocks change agents' forecasts of future real interest rates. These expected changes pass through almost one-for-one to aggregate output and consumption.[24] The additional persistence resulting from agents' use of simple models allows the model to generate realistic IRFs to monetary-policy shocks—despite the fact that the economy has no wage rigidity, wage or price indexation, habit formation, or investment-adjustment costs, and it only has moderate degrees of price rigidity and capital-adjustment costs.

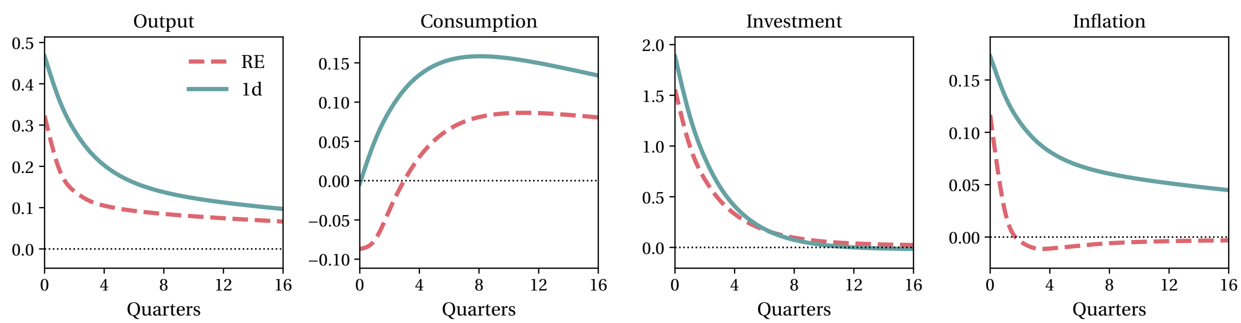

Figure 3. Impulse-response functions to a positive investment-demand shock.

Notes: Baseline calibration. One-dimensional simple models in solid blue. Rational expectations in dashed red. Responses to a one-percent increase in the demand for investment. Output, consumption, and investment in percents; inflation in percentage points.

[24] This is the forward guidance puzzle of Del Negro, Giannoni, and Patterson (2023). See Angeletos and Lian (2018), García-Schmidt and Woodford (2019), Farhi and Werning (2019), and Gabaix (2020) for other resolutions based on relaxations of the rational-expectations assumption.

Figure 3 plots the IRFs to an investment shock. Under rational expectations, the economy produces a negative comovement between consumption and investment in response to the shock. This is due to the fact that investment shocks increase the marginal productivity of investment and the rate of return, incentivizing households to save more and postpone consumption.[25] DSGE models overturn this prediction through frictions such as nominal-wage rigidity, non-time-separable preferences, endogenous capital utilization, and investment-adjustment costs. This paper's framework offers an alternative solution that relies on the anchoring of expectations. Since p_{μ}^{*} is close to zero in equilibrium, agents' expectations do not move much in response to the positive investment shocks. This dampens the initial response of forward-looking variables such as consumption to the shock. As time passes, the investment boom increases the capital stock and aggregate output, leading to an increase in the value of agents' nowcast. This improvement in agents' nowcast makes them optimistic about their future income, thus leading to an increase in consumption through the permanent-income equation.

These impulse-response functions offer suggestive evidence that simple models could improve the empirical fit of the business-cycle model economy. However, to establish this conclusively requires the use of a statistical model-selection criterion. I do so by estimating the parameters of the economic model separately under rational expectations and under simple models and performing Bayesian model selection.

5.3 Bayesian Inference

The model is estimated with Bayesian estimation techniques using seven key macroeconomic quarterly US time series as observable variables: the log differences of real GDP, real consumption, real investment, and the real wage; log hours worked; the log difference of the GDP deflator; and the federal funds rate. The construction of the time series closely follows Justiniano et al. (2010). In particular, consumption includes services and non-durables but excludes consumer durables, whereas investment includes consumer durables. The sample period is 1954:III–2007:IV. Online Appendix D includes more details on the data used to construct the likelihood function as well as the prior densities and posterior estimates of model parameters.

Since the likelihood uses seven macroeconomic series, the theoretical economic model needs seven exogenous shocks to avoid issues with stochastic singularity. I enrich the model with four additional shocks: an intertemporal preference shock, price- and

[25]This observation goes back at least to Barro and King (1984). Also see Justiniano et al. (2010) for a discussion.

wage-markup shocks, and a government spending shock, ending up with the seven shocks commonly assumed in the DSGE literature. The intertemporal preference and government spending shocks follow AR(1) processes with Gaussian innovations. Following Smets and Wouters (2007) and Justiniano et al. (2010), I assume that markup shocks follow ARMA(1,1) processes with Gaussian innovations. The moving average component of these shocks help capture high-frequency fluctuations in price and wage inflation.

I partition the model parameters into two groups. The first group consists of β, γ, δ, φ, F, g/y, and b/y. These parameters are set using level information not used in the Bayesian estimation step. I set $\beta = 0.99$, which implies a steady-state annualized real interest rate of about four percent. I set $\gamma = 1.005$, which implies an annual real GDP growth rate of two percent. I set $\delta = 0.025$, implying an annual rate of depreciation on capital equal to 10 percent. No value is picked for φ because the value of φ does not affect anything other than the steady-state working hours. The fixed cost of production F is set to guarantee that profits are zero along the balanced-growth path. I set the steady-state ratio of government spending to GDP g/y to 0.21 and the steady-state ratio of public debt to GDP b/y to 0.39. These values correspond to the average ratios of government spending to GDP and public debt to GDP, respectively, in the sample used in Bayesian estimation.

The priors on the remaining parameters are fairly diffuse and in line with those adopted in Smets and Wouters (2007) and Justiniano et al. (2010). Following those papers, the intertemporal preference, the price-markup, and the wage-markup shocks are normalized to enter with a unit coefficient in the consumption, inflation, and wage equations, respectively. The prior distributions of all the persistence parameters are beta, with mean 0.6 and standard deviation 0.15. The priors on the standard deviations of innovations are disperse and chosen to generate volatilities for the endogenous variables broadly in line with the data. Specifically, the priors for the standard deviations of innovations are inverse gamma, with mean 0.5 and standard deviation 1.0.

Table 1 reports the contribution of each shock to the variance of each macroeconomic variable at business cycle-frequencies. The first three columns make clear that the three aggregate demand shocks account for the largest share of the fluctuations in aggregate quantities: 94% for output, 82% for consumption, more than 99% for investment, and 52% for hours. The only other shock with a significant contribution to fluctuations in these variables is the technology shock, which explains 45% of fluctuations in hours but almost no part of fluctuations in the other three. The aggregate demand shocks are non-inflationary. Together they explain less than 4% of fluctuations in inflation. This does not rely on the monetary policy aggressively leaning against them. The three aggregate demand shocks explain a negligible part of the fluctuations in the nominal interest rate.

Table 1. Posterior variance decomposition at business-cycle frequencies.

Series\shock	Government	Preference	Investment	Technology	Price markup	Wage markup	Monetary
GDP	14.2 [12.0, 15.7]	14.5 [12.3, 15.9]	65.5 [62.5, 69.6]	0.0 [0.0, 0.0]	0.2 [0.1, 0.3]	0.1 [0.0, 0.1]	5.5 [4.2, 6.3]
Consumption	0.0 [0.0, 0.0]	70.2 [61.9, 73.6]	12.3 [8.0, 22.1]	0.1 [0.0, 0.1]	1.0 [0.5, 1.3]	0.1 [0.1, 0.2]	16.2 [11.6, 19.1]
Investment	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	99.5 [99.3, 99.7]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.1 [0.0, 0.1]	0.4 [0.2, 0.6]
Hours	7.4 [6.5, 8.2]	7.5 [6.6, 8.4]	36.9 [33.2, 39.9]	45.2 [41.8, 48.2]	0.1 [0.1, 0.2]	0.0 [0.0, 0.0]	2.9 [2.2, 3.4]
Inflation	0.1 [0.1, 0.1]	0.1 [0.1, 0.1]	3.6 [0.1, 6.6]	7.6 [4.0, 9.8]	76.6 [66.3, 85.1]	12.0 [6.4, 14.8]	0.0 [0.0, 0.0]
Real wage	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	4.8 [3.1, 5.7]	1.2 [0.4, 1.8]	13.8 [9.2, 16.9]	80.2 [76.1, 85.5]	0.0 [0.0, 0.0]
Interest rate	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.3 [0.0, 0.5]	0.5 [0.1, 0.6]	4.1 [2.1, 5.2]	0.9 [0.3, 1.2]	94.3 [92.3, 96.8]

Notes: One-dimensional simple models. Business-cycle frequencies correspond to periodic components with cycles between 6 and 32 quarters, as in Stock and Watson (1999). Variance decomposition is performed at the posterior mode. 68 percent HPDIs computed using Laplace's approximation in brackets. HPDI bounds need not add up to one.

These findings are consistent with Angeletos, Collard, and Dellas (2020)'s anatomy of the US business cycles in the post-war period. They identify non-inflationary aggregate demand shocks as the main drivers of business cycles. These are shocks that lead to increases in output, consumption, investment, and hours with essentially no effect on inflation and no movement in TFP. Standard DSGE shocks cannot simultaneously meet all these requirements under rational expectations: Technology shocks move TFP while expansionary aggregate demand shocks generate inflation through the New Keynesian Phillips curve. Simple models increase the persistence of subjective expectations, thus significantly weakening the impact of the feedback embedded in the New Keynesian Phillips curve from expectations to current inflation. This allows demand-driven fluctuations in aggregate quantities with essentially no movement in productivity, inflation, or the nominal interest rate.

The observations made in this section give some credence to the idea that models featuring boundedly rational agents can better account for various aspects of business-cycle fluctuations. I end this section by comparing the fit of the estimated model to the data under rational expectations and simple models. I do so by comparing the marginal likelihoods of the Bayesian posteriors, reported in columns two and three of Table 2. The marginal likelihood is more than 150 log points higher with simple models than rational expectations, implying overwhelming posterior odds in favor of the former. Columns four to seven report the marginal likelihood given alternative specifications of the economic model, which feature additional add-ons but maintain rational expectations. While many of these add-ons can improve the fit, none of them do so as much as bounded rationality—

although consumption-habit formation comes close. This finding suggests that this paper's framework is not only favored by the data over rational expectations but also can act as a parsimonious substitute for the frictions commonly assumed in DSGE models.

Table 2. Marginal likelihoods.

	Simple models	Rational expectations					
Add-on	—	—	Indexation	Utilization	Investment adjustment	Taylor	Habit
Log marginal likelihood	−1319.5	−1470.4	−1470.8	−1468.5	−1457.1	−1400.0	−1321.8

Notes: Log marginal likelihoods are computed using Laplace's approximation. The specifications in columns two and three only feature price and wage rigidities. The specifications in columns four to eight each feature a single addition: price and wage indexation, endogenous capital utilization, investment-adjustment costs (instead of neoclassical capital-adjustment costs), a Taylor rule that responds to the level and growth rate of the output gap, and external habit formation in consumption, respectively.

6 Conclusion

This paper characterizes the implications of agents' reliance on low-dimensional models for forecasting economic variables. It shows that agents who use such models focus on accurately forecasting the most persistent components of the observable while neglecting the dynamics of the less persistent ones. Consequently, their forward-looking actions exhibit greater persistence and comovement than under rational expectations.

I illustrated the framework in a medium-scale new neoclassical synthesis economy. The additional persistence and comovement generated by simple models enhance the propagation of TFP shocks, enable investment-demand shocks to produce the empirically observed comovement between investment and consumption, and significantly dampen the response to monetary shocks—all while improving the model's empirical fit.

Although the analysis here is conducted within a representative-agent framework, the rational expectations assumption is even more heroic in heterogeneous-agent macro models with large state spaces (*cf.* Moll (2024)). Incorporating simple models into such heterogeneous-agent frameworks would allow one to study how bounded rationality interacts with idiosyncratic risk in shaping aggregate dynamics. Moreover, heterogeneity in model complexity (captured by differences in d) may itself play a crucial role in shaping aggregate outcomes. These promising directions are left for future research.

A Proofs of the Main Results

I first state two lemmas that are used in the proofs of the main results. The following is a corollary of standard results on the Kalman filter:

Lemma 1. *The Kullback–Leibler divergence rate of model $\theta = (A, B, Q, R)$ from the true process can be written as*

$$\begin{aligned} KLDR(\theta) = & -\frac{1}{2}\log\det{}^*\left(\hat{\Sigma}_y^\dagger\right) + \frac{n}{2}\log(2\pi) + \frac{1}{2}\operatorname{tr}\left(\hat{\Sigma}_y^\dagger\Gamma_0\right) - \frac{1}{2}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(\hat{\Sigma}_y^\dagger\Phi_\tau\Gamma_\tau'\right) \\ & -\frac{1}{2}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(\hat{\Sigma}_y^\dagger\Gamma_\tau\Phi_\tau'\right) + \frac{1}{2}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(\hat{\Sigma}_y^\dagger\Phi_s\Gamma_{\tau-s}\Phi_\tau'\right) + constant, \end{aligned} \tag{10}$$

where the constant contains terms that do not depend on θ, $\Phi_\tau \equiv B'(A-KB')^{\tau-1}K$, $K \in \mathbb{R}^{d\times n}$ is the Kalman gain matrix and $\hat{\Sigma}_y \equiv Var_t^\theta(y_{t+1})$ is the subjective conditional variance of y_{t+1}, given by

$$K = A\hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger, \tag{11}$$

$$\hat{\Sigma}_y = B'\hat{\Sigma}_z B + R, \tag{12}$$

and $\hat{\Sigma}_z \equiv Var_t^\theta(z_{t+1})$ is the subjective conditional variance of z_{t+1}, which solves the following (generalized) algebraic Riccati equation:[26]

$$\hat{\Sigma}_z = A\left(\hat{\Sigma}_z - \hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger B'\hat{\Sigma}_z\right)A' + Q. \tag{13}$$

Furthermore, the s-period-ahead forecast under model θ is given by

$$E_t^\theta[y_{t+s}] = B'A^{s-1}\sum_{\tau=0}^{\infty}(A-KB')^\tau K y_{t-\tau}. \tag{14}$$

The next lemma is obtained from the previous lemma by a change of variables:

Lemma 2. *Model $\theta = (A, B, Q, R)$ is a pseudo-true d-state model given true autocovariance matrices $\{\Gamma_l\}_l$ with Γ_0 invertible 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, N) is a tuple that minimizes $KLDR(\tilde{M}, \tilde{D}, \tilde{N})$*

[26] The dagger denotes the Moore–Penrose pseudo-inverse, and $\det^*$ denotes the pseudo-determinant, i.e., the product of all non-zero eigenvalues of a square matrix. These objects are the appropriate counterparts of the matrix inverse and the determinant for the case where $\mathcal{W}$ does not equal $\mathbb{R}^n$ and model θ is degenerate. See Chapter 4 of Anderson and Moore (2005) for a treatment in the non-singular case and Silverman (1976) for the case where $B'\hat{\Sigma}_z B + R$ may be singular.

subject to the constraints that $\tilde{M}$ is a $d \times d$ convergent matrix, $\tilde{D}$ is an $n \times d$ diagonal matrix with elements in the $[0,1]$ interval, $\tilde{N}$ is an $n \times n$ invertible matrix, and $\|\tilde{M}(I - \tilde{D}\tilde{D}')\tilde{M}'\|_2 < 1$, where KLDR$(\tilde{M}, \tilde{D}, \tilde{N})$ is given by

$$-\tfrac{1}{2}\log\det\left(\tilde{N}\tilde{N}'\right) + \tfrac{1}{2}\operatorname{tr}\left(\tilde{N}'\Gamma_0\tilde{N}\right) - \textstyle\sum_{\tau=1}^{\infty}\operatorname{tr}\left(\left(\tilde{M}\left(I - \tilde{D}'\tilde{D}\right)\right)^{\tau-1}\tilde{M}\tilde{D}'\tilde{N}'\Gamma_\tau'\tilde{N}\tilde{D}\right)$$
$$+\tfrac{1}{2}\textstyle\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\operatorname{tr}\left(\tilde{D}\left(\tilde{M}\left(I - \tilde{D}'\tilde{D}\right)\right)^{s-1}\tilde{M}\tilde{D}'\tilde{N}'\Gamma_{\tau-s}\tilde{N}\tilde{D}\tilde{M}'\left(\left(I - \tilde{D}'\tilde{D}\right)\tilde{M}'\right)^{\tau-1}\tilde{D}'\right). \quad (15)$$

Furthermore, the subjective expectation of y_{t+s} and the subjective variance of y_{t+1} conditional on the time-t information can be written in terms of matrices M, D, and N as

$$E_t^\theta[y_{t+s}] = N'^{-1}DM^{s-1}\sum_{\tau=0}^{\infty}\left(M\left(I - D'D\right)\right)^\tau MD'N'y_{t-\tau}, \quad (16)$$

$$\hat{\Sigma}_y = N^{-1\prime}\left(I + \sum_{\tau=1}^{\infty}DM^\tau D'DM'^\tau D'\right)N^{-1}, \quad (17)$$

and $\hat{\Sigma}_z \equiv Var_t^\theta(z_{t+1})$, the subjective conditional variance of z_{t+1}, satisfies

$$\hat{\Sigma}_z^{\frac{1}{2}}B\left(B'\hat{\Sigma}_zB + R\right)^{-1}B'\hat{\Sigma}_z^{\frac{1}{2}} = VD'DV' \quad (18)$$

for some orthogonal matrix V.

The proofs of these two lemmas are straightforward. They can be found in Appendix G of Molavi (2024).

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

Let $\theta \in \Theta_d$ denote an arbitrary pseudo-true d-state model when the true process is $\mathbb{P}$ and $\tilde{\theta} \in \widetilde{\Theta}_d$ denote an arbitrary pseudo-true d-state model when the true process is $\widetilde{\mathbb{P}}$. I first show that $T(P^\theta)$ and $P^{\tilde{\theta}}$ are both supported on $\widetilde{\mathcal{W}}$. Note that there always exists a d-state model for which the KLDR is finite—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, P^θ is supported on $\mathcal{W}$, and so, $T(P^\theta)$ is supported on $\widetilde{\mathcal{W}}$. On the other hand, since the true distribution $\mathbb{P}$ is supported on $\mathcal{W}$, the transformed distribution $\widetilde{\mathbb{P}}$ is supported on $\widetilde{\mathcal{W}}$. Consequently, by the above argument, $P^{\tilde{\theta}}$ 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 model $T(\theta) \in \widetilde{\Theta}_d$ as $T(\theta) \equiv \big(A, BT', Q, TRT'\big)$. 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:

$$\hat{\Sigma}_z = A\left(\hat{\Sigma}_z - \hat{\Sigma}_z B\left(B'\hat{\Sigma}_z B + R\right)^\dagger B'\hat{\Sigma}_z\right)A' + Q, \tag{19}$$

$$\widetilde{\hat{\Sigma}}_z = A\left(\widetilde{\hat{\Sigma}}_z - \widetilde{\hat{\Sigma}}_z BT'\left(TB'\widetilde{\hat{\Sigma}}_z BT' + TRT'\right)^\dagger TB'\widetilde{\hat{\Sigma}}_z\right)A' + Q. \tag{20}$$

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 $\Phi_\tau \equiv B'(A - KB')^{\tau-1}K$, and let $\widetilde{\Phi}_\tau$ denote the corresponding matrix given model $T(\theta)$. Then, $\widetilde{\Phi}_\tau \equiv TB'(A - KT^\dagger TB')^{\tau-1}KT^\dagger = T\Phi_\tau T^\dagger$. 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 $T(\theta)$. Then, $\widetilde{\hat{\Sigma}}_y = TB'\widetilde{\hat{\Sigma}}_z BT' + TRT' = T\hat{\Sigma}_y T'$. On the other hand, $\widetilde{\Gamma}_l \equiv \widetilde{\mathbb{E}}[\tilde{y}_t\tilde{y}_{t-l}'] = T\mathbb{E}[y_t y_{t-l}]T' = T\Gamma_l T'$. Therefore, Lemma 1 and the fact that $T^\dagger T = I$ imply that $\widetilde{\text{KLDR}}(T(\theta)) = \text{KLDR}(\theta)$, 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 $\widetilde{\Theta}_d$: 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 $T(\mathbb{P})$, 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 $T(\mathbb{P})$. This conclusion immediately implies that the set of pseudo-true d-state models under true process $\mathbb{P}$ is identified with the set of pseudo-true d-state models under true process $T(\mathbb{P})$.

It only remains to show that $P^{T(\theta)} = T(P^\theta)$ for any model $\theta \in \Theta_d$. Since $P^{T(\theta)}$ and $T(P^\theta)$ 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 $T(P^\theta)$. $\square$

Proof of Theorem 2. Let M, D, and N be as in Lemma 2. When $d = 1$, then $M = a$ for some $a \in [-1, 1]$ and $D = d_1 e_1$ for some $d_1 \in [0, 1]$, where e_1 denotes the first coordinate vector. Define $\eta \equiv 1 - d_1^2$ and $S \equiv \Gamma_0^{\frac{1}{2}} N$. Then KLDR, defined in (15), can be written (with slight abuse of notation) as a function of a, η, and S:

$$\mathrm{KLDR}(a, \eta, S) = -\frac{1}{2} \log\det(SS') + \frac{1}{2}\operatorname{tr}(S'S) - \frac{1}{2} e_1' S' \Omega(a, \eta) S e_1 + \text{constant},$$

where

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

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

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

Therefore,

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

By Lemma 2, minimizing the KLDR with respect to A, B, Q, and R is equivalent to minimizing $\mathrm{KLDR}(M, D, N)$ with respect to M, D, and N. But for any a, η, and S, one can construct a corresponding M, D, and N, and vice versa. Therefore, I can instead minimize $\mathrm{KLDR}(a, \eta, S)$ with respect to a, η, and S.

I first minimize $\mathrm{KLDR}(a, \eta, S)$ with respect to S taking a and η as given. The first-order optimality condition with respect to S is given by $S^{-1} = S' - e_1 e_1' S' \Omega(a, \eta)$, which implies that

$$S'S - e_1 e_1' S' \Omega(a, \eta) S = I. \tag{22}$$

Therefore, for any solution to the problem of minimizing $\mathrm{KLDR}(a, \eta, S)$,

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

Thus, minimizing $\mathrm{KLDR}(a, \eta, S)$ with 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

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

I proceed by first characterizing $S(a,\eta)$. Note that the necessary first-order optimality conditions for problem (23) are given by matrix equation (22). The following claim is proved in the Online Appendix:

Claim 1. *For any matrix S that solves equation (22), the necessary first-order optimality condition for problem (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 $u'u = 1$.

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

I next find parameters A, B, Q, and R representing the a, η, and S that minimize KLDR(a,η,S). First, note that $M = a$, $D = \sqrt{1-\eta}e_1$, and $N = \Gamma_0^{\frac{-1}{2}}S$. The representation in Lemma 2 is thus given by $A = a$, $B = \sqrt{1-\eta}e_1'S^{-1}\Gamma_0^{\frac{1}{2}}$, $Q = 1 - a^2\eta$, and $R = \Gamma_0^{\frac{1}{2}}S^{-1\prime}\left(I - (1-\eta)e_1e_1'\right)S^{-1}\Gamma_0^{\frac{1}{2}}$.[27] By Claim 1 and the argument above,

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

[27] For this (A,B,Q,R) tuple to represent a one-state model, I need A to be convergent, Q to be positive definite, and R to be positive semidefinite. That R is always positive semidefinite is immediate. Showing that A is convergent and Q is positive definite takes more work. I do so in Lemma E.2.

Thus, $B = \sqrt{(1-\eta)\,(1-\lambda_{\max}(\Omega))}u'_{\max}(\Omega)\Gamma_0^{\frac{1}{2}}$, and

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

Finally, note that $M = a$, $D = \sqrt{1-\eta}e_1$, and $N = \Gamma_0^{\frac{-1}{2}}S$. Therefore, by equation (16), the subjective forecasts are given by

$$E_t^{\theta}[y_{t+s}] = a^s(1-\eta)\Gamma_0^{\frac{1}{2}}S'^{-1}e_1e_1'S'\Gamma_0^{\frac{-1}{2}}\sum_{\tau=0}^{\infty}a^{\tau}\eta^{\tau}y_{t-\tau}. \tag{24}$$

Using Claim 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 definitions of p and q. □

Proof of Theorem 4. Let λ denote the eigenvalue of C_1 largest in magnitude.[28] If $\rho(C_1) = 0$, then $\rho(C_\tau) = 0$ for all $\tau \geq 1$. Since C_τ are symmetric matrices, this implies that $C_\tau = 0$ for all $\tau \geq 1$. Therefore,

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

The above expression is maximized by setting $(1-\eta)a = 0$. Therefore, by Theorem 2, for any pseudo-true one-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 $\rho(C_1) = 0$.

In 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 $\rho(C_\tau) < 1$, established in Lemma E.1. Function $\overline{f}(a,\eta)$ has two maximizers given by $(\overline{a}^*, \overline{\eta}^*) = (-\rho(C_1), 0)$ and $(\overline{a}^*, \overline{\eta}^*) = (\rho(C_1),$

[28] 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 $|\lambda| = |\lambda'| = \rho(C_1)$.

0) with the maximum given by $\overline{f}^* = \rho(C_1)^2$. I establish the theorem by showing that $\lambda_{\max}(\Omega(a,\eta)) \leq \overline{f}(a,\eta)$ for all a and η, $\lambda_{\max}(\Omega(\lambda,0)) = \overline{f}(\lambda,0) = \overline{f}^*$, and $\lambda_{\max}(\Omega(-\lambda,0)) \leq \overline{f}(-\lambda,0) = \overline{f}^*$ 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 τ,

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

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

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

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

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

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

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

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

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

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

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

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

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

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 pseudo-true one-state model, $\eta = 0$, $a = \lambda$, and

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

By Theorem 2, 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. □

References

AKERLOF, GEORGE A. AND JANET L. YELLEN (1985): "A Near-Rational Model of the Business Cycle, With Wage and Price Inertia," *The Quarterly Journal of Economics*, 100 (Supplement), 823–838.

ALVAREZ, FERNANDO E., FRANCESCO LIPPI, AND LUIGI PACIELLO (2015): "Monetary Shocks in Models with Inattentive Producers," *The Review of Economic Studies*, 83 (2), 421–459.

ANDERSON, BRIAN D.O. AND JOHN B. MOORE (2005): *Optimal Filtering*, Mineola, N.Y.: Dover Publications.

ANGELETOS, GEORGE-MARIOS, FABRICE COLLARD, AND HARRIS DELLAS (2018): "Quantifying Confidence," *Econometrica*, 86 (5), 1689–1726.

——— (2020): "Business-Cycle Anatomy," *American Economic Review*, 110 (10), 3030–70.

ANGELETOS, GEORGE-MARIOS AND ZHEN HUO (2021): "Myopia and Anchoring," *American Economic Review*, 111 (4), 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 (2018): "Forward Guidance without Common Knowledge," *American Economic Review*, 108 (9), 2477–2512.

BARRO, ROBERT J. AND ROBERT G. KING (1984): "Time-Separable Preferences and Intertemporal-Substitution Models of Business Cycles," *The Quarterly Journal of Economics*, 99 (4), 817–839.

BERK, ROBERT H. (1966): "Limiting Behavior of Posterior Distributions When the Model is Incorrect," *Annals of Mathematical Statistics*, 37 (1), 51–58.

BIDDER, RHYS AND IAN DEW-BECKER (2016): "Long-Run Risk Is the Worst-Case Scenario," *American Economic Review*, 106 (9), 2494–2527.

BRANCH, WILLIAM A. (2006): "Restricted Perceptions Equilibria and Learning in Macroeconomics," *Post Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model. Cambridge University Press, New York*, 135–160.

BRAY, M. M. AND N. E. SAVIN (1986): "Rational Expectations Equilibria, Learning, and Model Specification," *Econometrica*, 54 (5), 1129–1160.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113 (1), 1–45.

Del Negro, Marco, Marc P. Giannoni, and Christina Patterson (2023): "The Forward Guidance Puzzle," *Journal of Political Economy Macroeconomics*, 1, 43–79.

Dew-Becker, Ian and Charles G. Nathanson (2019): "Directed Attention and Nonparametric Learning," *Journal of Economic Theory*, 181, 461–496.

Douc, Randal and Eric Moulines (2012): "Asymptotic Properties of the Maximum Likelihood Estimation in Misspecified Hidden Markov Models," *Annals of Statistics*, 40 (5), 2697–2732.

Esponda, Ignacio and Demian Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 84 (3), 1093–1130.

Eusepi, Stefano and Bruce Preston (2011): "Expectations, Learning, and Business Cycle Fluctuations," *American Economic Review*, 101 (6), 2844–2872.

——— (2018): "The Science of Monetary Policy: An Imperfect Knowledge Perspective," *Journal of Economic Literature*, 56 (1), 3–59.

Evans, George W. and Seppo Honkapohja (2001): *Learning and Expectations in Macroeconomics*, Princeton University Press.

Farhi, Emmanuel and Iván Werning (2019): "Monetary Policy, Bounded Rationality, and Incomplete Markets," *American Economic Review*, 109 (11), 3887–3928.

Fuster, Andreas, Benjamin Hebert, and David Laibson (2012): "Investment Dynamics with Natural Expectations," *International Journal of Central Banking*, 8 (SUPPL. 1), 243–265.

Fuster, Andreas, David Laibson, and Brock Mendel (2010): "Natural Expectations and Macroeconomic Fluctuations," *Journal of Economic Perspectives*, 24 (4), 67–84.

Gabaix, Xavier (2014): "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 129 (4), 1661–1710.

——— (2020): "A Behavioral New Keynesian Model," *American Economic Review*, 110, 2271–2327, working paper.

García-Schmidt, Mariana and Michael Woodford (2019): "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis," *American Economic Review*, 109 (1), 86–120.

Grandmont, Jean-Michel (1977): "Temporary General Equilibrium Theory," *Econometrica*, 45 (3), 535–572.

Hayashi, Fumio (1982): "Tobin's Marginal Q and Average Q: A Neoclassical Interpretation," *Econometrica*, 213–224.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti (2010): "Investment Shocks and Business Cycles," *Journal of Monetary Economics*, 57 (2), 132–145.

Kim, Jinill and Sunghyun Henry Kim (2003): "Spurious Welfare Reversals in International Business Cycle Models," *Journal of International Economics*, 60 (2), 471–500.

Krusell, Per and Anthony A. Smith, Jr. (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106 (5), 867–896.

Lorenzoni, Guido (2009): "A Theory of Demand Shocks," *American Economic Review*, 99 (5), 2050–84.

Lucas, Robert E. (1972): "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 4 (2), 103–124.

Maćkowiak, Bartosz and Mirko Wiederholt (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99 (3), 769–803.

——— (2015): "Business Cycle Dynamics under Rational Inattention," *Review of Economic Studies*, 82 (August), 1502–1532.

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 (4), 1295–1328.

Marcet, Albert and Thomas J. Sargent (1989): "Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models," *Journal of Economic Theory*, 48 (2), 337–368.

Milani, Fabio (2007): "Expectations, Learning and Macroeconomic Persistence," *Journal of Monetary Economics*, 54 (7), 2065–2082.

Molavi, Pooya (2024): "Simple Models and Biased Forecasts," Working paper, arXiv:2202.06921v5.

Molavi, Pooya, Alireza Tahbaz-Salehi, and Andrea Vedolin (2024): "Model Complexity, Expectations, and Asset Prices," *The Review of Economic Studies*, 91 (4), 2462–2507.

Moll, Benjamin (2024): "The Trouble with Rational Expectations in Heterogeneous Agent Models: A Challenge for Macroeconomics," Working paper.

Nimark, Kristoffer (2008): "Dynamic Pricing and Imperfect Common Knowledge," *Journal of Monetary Economics*, 55 (2), 365–382.

Orphanides, Athanasios (2003): "Monetary Policy Evaluation With Noisy Information," *Journal of Monetary Economics*, 50 (3), 605–631.

Ramey, Valerie A. (2016): "Macroeconomic Shocks and Their Propagation," *Handbook of Macroeconomics*, 2, 71–162.

Sawa, Takamitsu (1978): "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46 (6), 1273–1291.

Shalizi, Cosma Rohilla (2009): "Dynamics of Bayesian Updating with Dependent Data and Misspecified Models," *Electronic Journal of Statistics*, 3, 1039–1074.

Silverman, Leonard M. (1976): "Discrete Riccati Equations: Alternative Algorithms, Asymptotic Properties, and System Theory Interpretations," in *Control and Dynamic Systems*, Elsevier, vol. 12, 313–386.

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

Slobodyan, Sergey and Raf Wouters (2012): "Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models," *American Economic Journal: Macroeconomics*, 4 (2), 65–101.

Smets, Frank and Rafael Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97 (3), 586–606.

Stock, James H. and Mark W. Watson (1999): "Business Cycle Fluctuations in US Macroeconomic Time Series," in *Handbook of Macroeconomics*, Elsevier, vol. 1, 3–64.

White, Halbert (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50 (1), 1–25.

Woodford, Michael (2002): "Inflation Stabilization and Welfare," *Contributions to Macroeconomics*, 2 (1), 1009.

——— (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 25.

——— (2013): "Macroeconomic Analysis Without the Rational Expectations Hypothesis," *Annual Review of Economics*, 5 (1), 303–346.