Model Complexity, Expectations, and Asset Prices*

Pooya Molavi† Alireza Tahbaz-Salehi‡ Andrea Vedolin§

February 2022

Abstract

This paper analyzes how limits to the complexity of statistical models used by market participants can shape asset prices. We consider an economy in which agents can only entertain models with at most k factors, where k may be distinct from the true number of factors that drive the economy's fundamentals. We first characterize the implications of the resulting departure from rational expectations for return dynamics and relate the extent of return predictability at various horizons to the number of factors in the agents' models and the statistical properties of the underlying data-generating process. We then apply our framework to two applications in asset pricing: (i) violations of uncovered interest rate parity at different horizons and (ii) momentum and reversal in equity returns. We find that constraints on the complexity of agents' models can generate return predictability patterns that are consistent with the data.

*We are grateful to Fernando Alvarez, Andrea Buffa, Charles Engel, Ignacio Esponda, Xavier Gabaix, Tarek Hassan, Ben Hébert, Ali Jadbabaie, Lars Hansen, Jen-Jen La'O, Xinyang Li, Stefan Nagel, Charlie Nathanson, Stavros Panageas, Farzad Saidi, Harald Uhlig, Dimitri Vayanos, Laura Veldkamp, Mike Woodford, Hao Xing, and seminar participants at the University of Chicago, Brandeis University, Boston University, Imperial College, Northwestern University, University of Rochester, UNSW Business School, the Virtual Finance Theory Workshop, the Econometric Society European Winter Meeting, the NBER Behavioral Finance Fall Meeting, the Southern Economic Association Conference, the American Economic Association Meeting, the American Finance Association Meeting, and the World Congress of the Econometric Society for helpful comments and suggestions.

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

‡Northwestern University and CEPR, alirezat@kellogg.northwestern.edu.

§Boston University, NBER, and CEPR, avedolin@bu.edu.

1 Introduction

The rational expectations framework maintains that agents have a complete understanding of their economic environment: they know the structural equations that govern the relationship between endogenous and exogenous variables, have full knowledge of the stochastic processes that determine the evolution of shocks, and are capable of forming and updating beliefs about as many variables as necessary. These assumptions are imposed irrespective of how complex the actual environment is. However, in reality, limits to cognitive and computational abilities mean that market participants are bound to rely on simplified models that may not fully account for the complexity of their environment. Thus, to the extent that agents employ such simplified models, their decisions—and any outcome that depends on those decisions—would depart from predictions obtained under rational expectations.

In this paper, we study how limits to the complexity of statistical models used by market participants shape asset prices. We consider a framework in which the stochastic process that governs the evolution of economic variables may not have a simple representation, and yet, agents are only capable of entertaining statistical models with a certain level of complexity. As a result, they may end up with a low-dimensional approximation to the true data-generating process. We show that this form of model misspecification generates systematic deviations from rational expectations with sharp predictions for the extent and nature of return predictability at different horizons.[1]

We present our results in the context of a simple asset pricing environment, in which a sequence of exogenous fundamentals (say, an asset's dividends) are generated by a stochastic process that can be represented as an n-factor model. While agents can observe the sequence of realized fundamentals, they neither observe nor know the underlying factors that drive them. As a result, they rely on their past observations to estimate a hidden-factor model that would allow them to make predictions about the future. As our main assumption, we assume that agents can only hold and update beliefs about models with at most k factors, where k may be distinct from the true number of factors, n. As in Molavi (2019), this assumption captures the idea that there is a limit to the complexity of statistical models that agents are able to consider, with a larger k corresponding to a more sophisticated agent who can entertain a richer class of models.

Formally, we assume that agents start with a prior belief with full support over the set of k-factor models and update their beliefs over time in a Bayesian fashion. While simple, this formulation has three important features. First, the restriction on the support of agents' priors reflects our behavioral assumption that agents are incapable of holding and updating beliefs about models that are more complex than what they can entertain, while at the same time guaranteeing that they do not, a priori, rule out any model with k or fewer factors. Second, the assumption that agents in our framework are Bayesian ensures that any deviation from the predictions of the rational expectations

[1] While we focus on limits on the complexity of agents' statistical models, one can also consider other dimensions along which decision makers are constrained, such as limited memory (Nagel and Xu, 2019; da Silveira, Sung, and Woodford, 2020), limited capacity for processing information (as in models of rational inattention), an incomplete understanding of general equilibrium effects (e.g., due to level-k reasoning as in Farhi and Werning (2019) or limited "depth of knowledge" as in Angeletos and Sastry (2020)), and restrictions on the number of variables to pay attention to (Gabaix, 2014, 2017).

benchmark is entirely due to the complexity constraint on the agents' models (as captured by the wedge between k and n). Third, this formulation implies that all parameters of the models agents use for forecasting are endogenous outcomes of learning: once we specify the number of factors k, there are no more degrees of freedom on how agents form their subjective expectations. As a result, agents' forecasts (and the resulting price and return dynamics) are fully pinned down by (i) the constraint on the complexity of the agents' models and (ii) the statistical properties of the true data-generating process.

With our behavioral framework in hand, we first characterize the agents' subjective expectations in terms of the primitives of the economy. We establish that, as agents accumulate more observations, their posterior beliefs concentrate on the subset of k-factor models with minimum Kullback-Leibler (KL) divergence to the true data-generating process. This result implies that when agents can contemplate models that are as complex as the true model (that is, when $k \geq n$), they can forecast the future realizations of the fundamental as if they knew the true data-generating process. In other words, when $k \geq n$, our framework reduces to the rational expectations benchmark. Furthermore, the characterization of agents' subjective expectations in terms of the KL divergence to the true process enables us to determine price and return dynamics in the more interesting case when $k < n$.

Having characterized the agents' subjective expectations, we then turn to the main focus of our analysis: how the constraint on the complexity of statistical models used by the agents shapes asset prices. More specifically, we study the extent to which the disparity between the number of factors in the agents' models (k) and that of the true data-generating process (n) can result in (excess) return predictability. We measure the extent of return predictability by relying on two families of linear regressions, which we refer to as the *Fama* and *momentum regressions*. These regressions measure, respectively, the extent to which current fundamentals and excess returns predict future returns at different horizons. Our main theoretical results provide a characterization of the slope coefficients of the aforementioned regressions in terms of the complexity of agents' models and the autocorrelation function (ACF) of the process that drives the fundamentals. Since all parameters in the agents' models, other than the number of factors k, are endogenously determined, our results provide sharp predictions for the extent of return predictability at different horizons, with no remaining degrees of freedom.

We then use our characterization theorem to obtain a series of results that relate the sign pattern of the slope coefficients of the Fama and momentum regressions to the primitives of the environment. In particular, we show that, when the agents are restricted to the class of single-factor models, the slope coefficient of the Fama regression at a given horizon is positive if and only if the ACF of the true data-generating process at that horizon decays at a slower rate compared to the ACF implied by the agents' lower-dimensional model. As our next result, we obtain an alternative characterization of the sign pattern of the slope coefficients of the Fama regression in terms of the spectrum of the matrix that governs the evolution of the economy's fundamental. Finally, we establish that as long as the true data-generating process exhibits short-term persistence, the coefficient of the momentum regression can never take the same sign for all horizons. In other

words, returns are not only predictable, but also exhibit both momentum and reversal.

We conclude our theoretical analysis by extending our framework to a heterogeneous-agent economy in which only a fraction of agents are subject to the constraint on model complexity, while the remaining fraction can entertain models of any order. We show that, in the presence of such heterogeneity, higher-order expectations play a central role in shaping return dynamics, and as a result, the presence of even a relatively small fraction of constrained agents can have an outsized impact on the extent of return predictability.

We then apply our framework to two asset pricing applications. As a first application, we study the implications of our behavioral assumption for violations of the uncovered interest rate parity (UIP) condition in foreign exchange. We show that the constraint on the number of factors in agents' model can generate return predictability patterns that are simultaneously consistent with two well-known, but seemingly contradictory violations of UIP, namely, the forward discount and the predictability reversal puzzles. The forward discount puzzle, which dates back to Fama (1984), is the robust empirical finding that, in short time horizons ranging from a week to a quarter, high interest rate currencies tend to have positive excess returns. The predictability reversal puzzle, more recently documented by Bacchetta and van Wincoop (2010) and Engel (2016), refers to the fact that high interest rate currencies tend to have negative excess returns over longer horizons, that is, the violation of UIP reverses sign after some point. The seemingly contradictory implications of these puzzles for the relationship between currency excess returns and interest rate differentials has led some to argue for the inadequacy of existing models for explaining UIP violations (Engel, 2016).

To test our model's implications for the violation of UIP at different horizons, we use a large cross-section of currency returns and, following Engel (2016), run predictive regressions from a trade-weighted average currency return on the corresponding interest rate differential. We then compare the resulting slope coefficients to those implied by our theoretical framework when agents are constrained in the number of factors in their models. We find that when investors are constrained to using single-factor models, the model-implied slope coefficients line up with the ones from the data: at short horizons, deviations from rational expectations generate UIP violations that imply positive excess returns for high interest rate currencies, whereas at longer horizons the pattern reverses, with high interest rate currencies earning a negative excess return. In other words, our framework generates return predictability patterns that are simultaneously consistent with the forward discount and predictability reversal puzzles. Crucially, the pattern of our model-implied Fama slope coefficients matches its empirical counterpart without using the data on exchange rates or excess returns: the model-implied slope coefficients are constructed solely from the autocorrelation of interest rate differentials.[2] We also show that our findings remain mostly unchanged even if only a relatively small fraction of agents are subject to our behavioral constraint.

As a further illustration, we study violations of UIP in a cross-section of developed and emerging market currencies by running the Fama regressions on a currency-by-currency basis. We find positive and statistically significant relationships between the empirically-estimated slope

[2]We contrast these findings to factor models of higher order and find that as agents become more sophisticated—and in particular, as we increase the number of factors in agents' model to $k = 3$—UIP violations mostly disappear at all horizons.

coefficients and their model-implied counterparts at various horizons.

As a second application of our framework, we study short-run momentum and long-run reversal in equity returns. Time-series momentum is the phenomenon that past returns are predictors of future returns. This pattern usually persists for a year and reverses over the longer term. To apply our framework to this context, we first calculate the autocorrelation function of the fundamental—in this case, the dividend growth process. Using our characterization results, we then obtain the term structure of the model-implied slope coefficients of the momentum regression. We find that the model-implied coefficients track patterns that are consistent with the empirically-estimated coefficients, simultaneously exhibiting short-term momentum and long-run reversal. We note that our framework generates these patterns without using the data on equity prices or returns, as the model-implied slope coefficients are solely constructed from the autocorrelation function of the dividend growth process.

Related Literature Our paper contributes to the literature that studies the asset pricing implications of deviations from rational expectations. The most related branch of this literature considers agents who have a misspecified view of the true data-generating process as a result of behavioral biases. For instance, Barberis, Shleifer, and Vishny (1998) assume agents mistakenly believe the innovations in earnings are drawn from a regime with excess reversals or excess streaks, whereas Gourinchas and Tornell (2004) consider a model in which agents misperceive the relative importance of transitory and persistent shocks. Relatedly, Rabin and Vayanos (2010) show that gambler's fallacy—the belief that random sequences should exhibit systematic reversals, even in small samples—can generate momentum and reversal in returns, while Guo and Wachter (2019) focus on a model in which investors believe that returns are predictable even when they are not. As in these papers, deviations from rational expectations in our framework are due to misspecification in agents' model of the underlying data-generating process. However, we depart from the prior literature along two dimensions. First, we assume that—subject to the constraint on the complexity of their model—agents can estimate a fully flexible linear factor model with no a priori imposed restrictions on its parameters. As a result, any distortion in agents' subjective expectations is purely due to the mismatch between the dimensions of their model and that of the true process. Second, we show that such a mismatch not only can result in momentum and reversal in returns, but can also generate UIP violations that are simultaneously consistent with the forward discount and predictability reversal puzzles.

Another strand of the literature studies how deviations from Bayesian updating shape asset prices. The extrapolative expectations models of Hirshleifer, Li, and Yu (2015), and Barberis, Greenwood, Jin, and Shleifer (2015) and diagnostic expectations models of Bordalo et al. (2018, 2019) are examples of such non-Bayesian models of updating. In contrast to this literature, we maintain the assumption of Bayesian updating and instead formulate our departure from rational expectations by assuming that agents assign zero prior beliefs on complex models with a large number of factors. This formulation allows us to impose constraints on the complexity of agents' models, while preserving other features of the rational expectations framework (such as the internal

consistency of agents' subjective expectations).

Our paper is also related to the broader literature that studies the implications of various kinds of constraints on agents' cognitive and computational abilities, such as imperfect or selective memory (Nagel and Xu, 2019; Bordalo et al., 2020; Wachter and Kahana, 2021), limited capacity for processing information (Van Nieuwerburgh and Veldkamp, 2009, 2010), incomplete understanding of general equilibrium effects (e.g., due to level-k reasoning), and restrictions on the number of variables to pay attention to (Gabaix, 2014). In contrast to these papers, agents in our framework (i) observe the exact realization of all economic variables, (ii) can fully recall the entire past history of their observations, and (iii) have a complete understanding of the relationship between endogenous and exogenous variables.[3]

Our behavioral framework is most closely related to and builds on two prior contributions: the natural expectations model of Fuster et al. (2010, 2012) and the constrained-rational expectations framework of Molavi (2019). Fuster et al. (2010, 2012) study an economy in which decision makers forecast a time series using only its last k realizations and find that when the true process has hump-shaped dynamics, the misspecification in the order of the autoregressive process generates excess returns that are negatively predicted by lagged excess return. We extend this framework along two dimensions. First, rather than restricting agents to forecast a time series using its last k realizations, we allow them to estimate a fully flexible k-factor model. This is equivalent to using any arbitrary k statics of the past realizations to forecast the future.[4] Second, we use our framework to characterize the extent of return predictability at various horizons as a function of the number of factors k in agents' model and the statistical properties of the underlying data-generating process. Also related is the constrained-rational expectations framework of Molavi (2019), who develops a framework to study arbitrary forms of model misspecification in a large class of macroeconomic, general equilibrium environments. He shows that restrictions on the number of factors in agents' models can generate comovement patterns in aggregate variables that resemble business cycle fluctuations. We instead focus on how constraints on agents' model complexity shape the extent and nature of return predictability in asset pricing applications.[5]

Finally, we also contribute to the literature that studies UIP violations and in particular, the reversal in currency return predictability. Engel (2016) develops a model to reconcile the forward discount and predictability reversal puzzles by introducing a non-pecuniary liquidity return on assets, Valchev (2020) proposes a mechanism based on endogenous fluctuations in bond

[3]Also see Martin and Nagel (2020), who show that when the number of *cross-sectional* factors that are potentially relevant for prediction is of the same order of magnitude as the number of assets with available cash-flow data, Bayesian investors use regularization to trade off the costs of downweighting certain pieces of information against the benefit of reduced parameter estimation error. In contrast to this work, we focus on an environment in which agents have access to abundant data but instead are restricted in the number of factors they use to fit the *time series* of their past observations.

[4]The family of k-factor models includes all $\mathrm{ARMA}(p,q)$ processes such that $\max\{p, q+1\} \leq k$ and thus nests the family of $\mathrm{AR}(k)$ processes as a special case.

[5]Methodologically, our paper is also related to the literature on model order reduction in control theory, which is concerned with characterizing lower-order approximations to large-scale dynamical systems. See Antoulas (2005) and Sandberg (2019) for textbook treatments of the subject. In contrast to this literature, which mostly focuses on optimal Hankel-norm approximations, Bayesian learning in our framework implies that agents end up with a model that minimizes the KL divergence to the true data-generating process.

convenience yields, and Bacchetta and van Wincoop (2020) argue for the role of delayed portfolio adjustments. Different from these papers, we show that constraints on the complexity of investors' statistical models—and the resulting departures from rational expectations—generate patterns that are simultaneously consistent with the forward discount and predictability reversal puzzles.[6]

Outline The rest of the paper is organized as follows. Section 2 presents the environment and specifies our behavioral assumption. Section 3 contains our main theoretical results, where we illustrate how the constraint on the complexity of agents' models shapes return predictability at various horizons. Section 4 presents our two empirical applications. All proofs and some additional mathematical details are provided in the Appendix.

2 Model

We start by presenting a reduced-form asset pricing framework that serves as the basis of our analysis. This framework allows us to focus on the main ingredients of our model, while abstracting from details that are not central to the analysis. As we discuss in Section 4, our framework is general enough to nest various asset pricing applications.

2.1 Reduced-Form Framework

Consider a discrete-time economy consisting of a unit mass of identical agents. Agents form subjective expectations about an exogenous sequence $\{x_t\}_{t=-\infty}^{\infty}$ of variables, which we refer to as the economy's *fundamentals*. Depending on the context, the fundamental may correspond to an asset's dividend stream over time, interest rate differential between two countries, or any sequence of variables that can be treated as exogenous in that specific application.

The sequence of fundamentals $\{x_t\}_{t=-\infty}^{\infty}$ is generated by a stationary n-factor model given by

$$\begin{aligned} z_t &= \mathbf{A}^* z_{t-1} + \mathbf{B}^* \epsilon_t \\ x_t &= c^{*\prime} z_t, \end{aligned} \tag{1}$$

where $z_t \in \mathbb{R}^n$ denotes the vector of factors that drive the dynamic of fundamentals, $c^* \in \mathbb{R}^n$ is a vector of constants that captures the fundamental's loading on each of the n factors, and $\mathbf{A}^*$ and $\mathbf{B}^*$ are square matrices that govern the evolution of factors over time. To ensure stationarity, we assume that all eigenvalues of $\mathbf{A}^*$ are inside the unit circle. The noise terms $\epsilon_t \in \mathbb{R}^n$ are independent over time and across factors and are normally distributed with mean zero and unit variance, i.e., $\epsilon_{it} \sim \mathcal{N}(0,1)$. The process that generates the fundamentals can thus be summarized by the n underlying factors $(z_1, z_2, \ldots, z_n)$ and the collection of parameters $\theta^* = (\mathbf{A}^*, \mathbf{B}^*, c^*)$. Note that the formulation in equation (1) is flexible enough to nest the entire family of stationary ARMA processes (Aoki, 1983).

[6]The expectations-based violations of UIP in our framework is also related to Froot and Frankel (1989) and Chinn and Frankel (2020), who document that most of the forward premium bias can be attributed to errors in agents' subjective expectations, as opposed to risk premia.

Throughout, we assume that there are no redundant factors in (1), in the sense that the dynamics of the fundamental cannot be represented by a factor model with fewer factors than n. The number of factors n thus measures the complexity of the data-generating process (1), with a larger n corresponding to a more complex process.[7]

Whereas the fundamental sequence $\{x_t\}_{t=-\infty}^{\infty}$ is assumed to be exogenous, we are interested in how the fundamentals and the agents' subjective expectations about them jointly determine an endogenous sequence of variables $\{y_t\}_{t=-\infty}^{\infty}$, which we refer to as *prices*. More specifically, we assume that the price at time t satisfies the recursive equation

$$y_t = x_t + \delta \mathbb{E}_t[y_{t+1}], \tag{2}$$

where $\mathbb{E}_t[\cdot]$ denotes the agents' time t subjective expectation and $\delta \in [0, 1]$ is a constant. Equation (2) is akin to a standard no-arbitrage condition with a natural interpretation. For example, if the sequence $\{x_t\}_{t=-\infty}^{\infty}$ represents an asset's dividend stream, then y_t corresponds to the price of the asset at time t, with equation (2) simply capturing the fact that the asset's price is the sum of its dividend at that time and its expected future price, discounted at some rate δ. We also note that the agents' subjective expectations, $\mathbb{E}_t[\cdot]$, may be distinct from the expectations arising from the true data-generating process (1), which we denote by $\mathbb{E}_t^*[\cdot]$.

Given the sequence of fundamentals and prices $\{(x_t, y_t)\}_{t=-\infty}^{\infty}$, we define *excess returns* at time $t+1$ as the sum of the fundamental and the change in price between t and $t+1$, properly discounted:

$$\mathrm{rx}_{t+1} = \delta y_{t+1} - y_t + x_t. \tag{3}$$

Using equation (2), we can also express excess returns as the difference between the realized and the expected price: $\mathrm{rx}_{t+1} = \delta(y_{t+1} - \mathbb{E}_t[y_{t+1}])$.

Together with the specification of how the agents' subjective expectations are formed, equations (1)–(3) fully describe our environment. Note that while we have abstracted from certain details of the economy (such as preferences, endowments, and the market structure), our subsequent analysis applies to any model with the same reduced-form representation as the above framework. As already mentioned, in Section 4, we show that this reduced-form representation is general enough to nest various asset pricing applications.

2.2 Constraints on Model Complexity

While at any given time t, agents observe the sequence of realized fundamentals up to that time, we assume that they observe neither the underlying factors $(z_1, \ldots, z_n)$ nor the collection of parameters $\theta^* = (\mathbf{A}^*, \mathbf{B}^*, c^*)$ that governs the data-generating process. As a result, they use the past realizations of the fundamental to learn about the underlying process that generates $\{x_t\}_{t=-\infty}^{\infty}$, which they then use to make forecasts about the future.

[7] In other words, a larger n means that describing the dynamics of the fundamental requires a system of linear difference equations of a higher dimension. The minimum number of factors necessary to represent the dynamics of $\{x_t\}_{t=-\infty}^{\infty}$ in the form of equation (1) is given by $n = \mathrm{rank}(\mathbf{Q}^*)$, where $\mathbf{Q}^*$ is the infinite-dimensional matrix of auto-covariances of the fundamental with typical element $q_{ij}^* = \mathbb{E}^*[x_t x_{t+i-j}]$ and $\mathbb{E}_t^*[\cdot]$ denotes the expectation operator corresponding to the data-generating process in (1)(Kailath, 1980, p. 363).

As our main behavioral assumption, we assume that agents face a constraint on the complexity of statistical models they can consider. Specifically, as in Molavi (2019), we assume that they can only entertain models with at most k factors, where k may be distinct from the true number of factors, n, in equation (1). The number of factors, k, thus indexes the agents' degree of sophistication, with a larger k corresponding to agents who can entertain a richer class of models.

Formally, we assume that agents can only hold and update beliefs over the set of k-factor models $\Theta_k = \{(\mathbf{A}, \mathbf{B}, c) : \mathbf{A}, \mathbf{B} \in \mathbb{R}^{k\times k} \text{ and } c \in \mathbb{R}^k\}$ of the form

$$\begin{aligned} \omega_t &= \mathbf{A}\omega_{t-1} + \mathbf{B}\varepsilon_t \\ x_t &= c'\omega_t, \end{aligned} \tag{4}$$

where $\omega_t \in \mathbb{R}^k$ denotes the vector of k underlying factors, $(\mathbf{A}, \mathbf{B}, c) \in \Theta_k$ parameterizes the process that govern the factors' evolution and the fundamental's loading on each of the factors, and the noise terms $\varepsilon_{it} \sim \mathcal{N}(0, 1)$ are independent over time and across factors. Note that, with some abuse of notation, we can write $\Theta_k \subseteq \Theta_{k+1}$, thus capturing the fact that agents with a higher k can contemplate a larger class of models.

A few remarks are in order. First, note that we impose no restrictions, other than the number of factors, on the agents' models: the k factors $(\omega_1, \ldots, \omega_k)$ in the agents' model may overlap with a subset of the n factors $(z_1, \ldots, z_n)$ that drive the fundamental, may be linear combinations of the underlying n factors, or can be constructed in an entirely different way altogether. Second, as we discuss in further detail below, the k-factor model agents use for forecasting is an endogenous outcome of learning: the agents will rule out any model $\theta \in \Theta_k$ that is inconsistent with their past observations. Third, when $k < n$, the set of models entertained by the agents does not contain the true n-factor data-generating process in (1), i.e., $\theta^* \notin \Theta_k$. In such a case, our behavioral assumption implies that irrespective of which k-factor model they use, the agents will end up with a misspecified model of the world. This observation also clarifies the bite of our behavioral assumption: whereas more sophisticated agents with $k \geq n$ can recover the model that generates the fundamentals (at least in principle), those with $k < n$ can at best construct lower-dimensional approximations to the true data-generating process.

2.3 Subjective Expectations and Learning

Agents form their subjective expectations by learning from the past realizations of the fundamental and updating their beliefs in a Bayesian fashion. Specifically, we assume that they start with a common prior belief, $\mu_{t_0} = \bar{\mu} \in \Delta\Theta_k$, with full support over the set of k-factor models at some initial period t_0 and form Bayesian posterior beliefs $\mu_t \in \Delta\Theta_k$ after observing the sequence of fundamentals $\{x_{t_0}, \ldots, x_{t-1}, x_t\}$.

This formulation has a few important implications. First, it implies that agents assign a zero prior belief on all models with more than k factors, thus reflecting our assumption that they are incapable of holding and updating beliefs about models that are more complex than what they can entertain. Second, the assumption that the prior $\bar{\mu}$ has full support over $\Delta\Theta_k$ guarantees that, a priori, agents do not rule out any model that has a representation with k or fewer factors. Therefore,

aside from the constraint on the number of factors, we allow agents to estimate a fully flexible model with arbitrary factor dynamics ($\mathbf{A}$), volatility of and covariance between the noise terms ($\mathbf{B}$), and loading of the fundamental on each of the factors (c). Third, and most importantly, agents in our framework are fully Bayesian with an internally-consistent system of beliefs, in the sense that their subjective expectations satisfy the law of iterated expectations: $\mathbb{E}_t\mathbb{E}_{t+h}[\cdot] = \mathbb{E}_t[\cdot]$ for all $h \geq 0$ and all t. This means that, under agents' subjective expectations, all forecast errors at time $t+h$ are unpredictable given their information set at time t. When coupled with equations (2) and (3), this observation also implies that excess returns in our framework are equal to the discounted sum of the agents' forecast revisions:[8]

$$\mathrm{rx}_{t+1} = \sum_{\tau=1}^{\infty} \delta^{\tau} \left(\mathbb{E}_{t+1}[x_{t+\tau}] - \mathbb{E}_t[x_{t+\tau}] \right). \tag{5}$$

Throughout the rest of the paper, we restrict our attention to the agents' long-run beliefs by assuming that they have already observed a long sequence of the past realizations of the fundamental. Formally, we assume that at any given period t, agents observe the sequence $\{x_{t_0}, \ldots, x_{t-1}, x_t\}$, and consider the limit as $t_0 \to -\infty$. This assumption allows us to study the systematic implications of the complexity constraint on the agents' models (i.e., the disparity between k and n), while abstracting from the temporary fluctuations in beliefs that arise after observing a finite sample.

2.4 Discussion

We conclude this section with a brief discussion of our framework and its key ingredients.

The assumption that agents in our framework are Bayesian with access to a long history of observations is meant to ensure that any deviation from the predictions of textbook rational expectations is entirely due to the complexity constraint on their models. Indeed, as we show in subsequent sections, when the agents' set of models is flexible enough to contain the true data-generating process (i.e., when $k \geq n$), Bayesian learning guarantees that agents can forecast future realizations of the fundamental as if they knew the true process. In such a case, our framework coincides with the benchmark rational expectations framework.

Bayesian updating also implies that the agents' posterior beliefs at any given time only depend on the sequence of realized fundamentals up to that time and is independent of other characteristics of the economy.[9] Thus, holding their information sets constant, agents in our framework end up

[8] Equation (5) also implies that returns are unpredictable under agents' subjective expectations, i.e., $\mathbb{E}_t[\mathrm{rx}_{t+h}] = 0$ for all $h \geq 1$. However, as we discuss in further details below, that does not mean that returns are also unpredictable under the expectations corresponding to the true data-generating process in (1), i.e., $\mathbb{E}_t^*[\mathrm{rx}_{t+h}] \neq 0$ in general.

[9] When agents have access to a finite collection of observations, their posterior beliefs at any time t depends not only on their history of observations $\{x_{t_0}, \ldots, x_t\}$ but also their prior belief, $\bar{\mu}$. However, as $t_0 \to -\infty$, the support of the limit set of their posteriors becomes independent of the specific choice of the prior, as long as the latter has full support over $\Delta\Theta_k$, an assumption we maintain throughout. We also note that, in view of equations (2) and (3), all realizations of prices and returns up to time t are measurable with respect to the information set generated by $\{x_\tau\}_{\tau=-\infty}^{t}$. Therefore, irrespective of whether past prices and returns are observable or not, it is sufficient to consider expectations conditioned on the past realizations of the fundamental.

with the same posterior beliefs irrespective of how preferences, endowments, or other features of the economy are specified.

Another important feature of our framework is that all parameters of the model agents use for forecasting are endogenous outcomes of learning: once we specify the number of factors k, there are no more degrees of freedom on how agents form their expectations. As a result, agents' forecasts (and the resulting price and return dynamics) are fully pinned down by (i) the limit on the complexity of their model, as captured by k, and (ii) the true data-generating process (1) as summarized by $\theta^* = (\mathbf{A}^*, \mathbf{B}^*, c^*)$.

The endogeneity of the agents' model also implies that how their forecasts respond to new information cannot be decoupled from the environment they live in. In particular, as we show in Appendix B.2, agents' forecasts may exhibit systematic over- or under-reaction to news depending on the statistical properties of the underlying data-generating process. This is in contrast to the literature that hardwires over- and under-reaction of expectations into the agents' models.

As a final remark, we note that our focus on the simple no-arbitrage condition in (2) abstracts from potentially important dimensions—such as time-varying discount rates— that matter for price and return dynamics. This choice allows us to isolate, in the most transparent manner, how the agents' inability to entertain complex models distorts their subjective expectations and shapes return dynamics.

3 Return Predictability

With the framework in Section 2 in hand, we now turn to the main focus of our analysis: how the constraint on the complexity of statistical models used by market participants can shape asset prices. More specifically, we study the extent to which the disparity between the number of factors in the agents' models (k) and that of the true data-generating process (n) can result in (excess) return predictability. Focusing on return predictability allows us to compare the implications of our behavioral framework to the benchmark of rational expectations, according to which future returns should be unpredictable.

3.1 Return Predictability Regressions

We measure the extent of return predictability by relying on two families of linear regressions. The first family of regressions, which we refer to as *Fama regressions*, measures the extent to which the current realization of the fundamental predicts future excess returns at different horizons:

$$\mathrm{rx}_{t+h} = \alpha_h^{\text{Fama}} + \beta_h^{\text{Fama}} x_t + \epsilon_{t,h}, \tag{6}$$

where $h \geq 1$ and rx denotes excess returns, as defined in (3). If returns are unpredictable—as would be the case under rational expectations—then $\beta_h^{\text{Fama}} = 0$ for all horizons h. The family of slope coefficients $(\beta_1^{\text{Fama}}, \beta_2^{\text{Fama}}, \dots)$ thus provides not only a measure for departures from the rational

expectations benchmark, but also the extent to which such departures vary with the prediction horizon.[10]

Our second family of regressions, which we refer to as *momentum regressions*, measures the extent to which current returns predict future returns:

$$\mathrm{rx}_{t+h} = \alpha_h^{\mathrm{mom}} + \beta_h^{\mathrm{mom}} \mathrm{rx}_t + \epsilon_{t,h}. \tag{7}$$

Once again, the unpredictability of future returns under rational expectations requires that $\beta_h^{\mathrm{mom}} = 0$ for all $h \geq 1$. Hence, the term structure of coefficients $(\beta_1^{\mathrm{mom}}, \beta_2^{\mathrm{mom}}, \dots)$ provides us with a natural measure for assessing the implications of our behavioral framework at different time horizons compared to the rational expectations benchmark, with a positive (negative) β_h^{mom} corresponding to the extent of time-series momentum (reversal) in returns at horizon h.

The following lemma relates the large-sample limits of slope coefficients in regressions (6) and (7) to agents' subjective expectations and the expectations with respect to the true data-generating process in (1).

Lemma 1. *The slope coefficients of the Fama and momentum regressions in* (6) *and* (7) *are given by*

$$\beta_h^{\mathrm{Fama}} = \frac{1}{\mathbb{E}^*[x_t^2]} \sum_{\tau=0}^{\infty} \delta^{\tau+1} \mathbb{E}^*[x_t(\mathbb{E}_{t+h}[x_{t+h+\tau}] - \mathbb{E}_{t+h-1}[x_{t+h+\tau}])] \tag{8}$$

$$\beta_h^{\mathrm{mom}} = \frac{\sum_{\tau,s=1}^{\infty} \delta^{\tau+s} \mathbb{E}^*[(\mathbb{E}_{t+1}[x_{t+\tau}] - \mathbb{E}_t[x_{t+\tau}])(\mathbb{E}_{t+h+1}[x_{t+h+s}] - \mathbb{E}_{t+h}[x_{t+h+s}])]}{\mathbb{E}^*[\sum_{\tau=1}^{\infty} \delta^{\tau}(\mathbb{E}_{t+1}[x_{t+\tau}] - \mathbb{E}_t[x_{t+\tau}])]^2}, \tag{9}$$

where $\mathbb{E}[\cdot]$ *and* $\mathbb{E}^*[\cdot]$ *denote the agents' subjective expectations and the expectations with respect to the true data-generating process, respectively.*

This result has a few immediate implications. First, it is straightforward to verify that, if $\mathbb{E}[\cdot] = \mathbb{E}^*[\cdot]$, then $\beta_h^{\mathrm{Fama}} = \beta_h^{\mathrm{mom}} = 0$ for all $h \geq 1$. Thus, as expected, under rational expectations, neither the fundamental nor excess returns are predictive of future returns. Second, when agents' subjective expectations do not coincide with the expectation under the true data-generating process—as would be the case when the complexity constraint on the agents' model binds—the coefficients in the return-predictability regressions may in general be different from zero. This is despite the fact that agents' expectations are internally consistent and satisfy the law of iterated expectations. Finally, Lemma 1 illustrates that the wedge between the agents' subjective expectations and rational expectations may have differential impacts on the slope coefficients at different horizons h.

Taken together, equations (8) and (9) illustrate that to determine the extent of return predictability (as captured by the term structures of β_h^{Fama} and β_h^{mom}), it is sufficient to characterize the agents' subjective expectations, $\mathbb{E}[\cdot]$, in terms of the primitives of the economy. Thus, as the first step in our analysis, we characterize how the constraint on the complexity of agents' models shapes their subjective expectations about the process that generates the fundamental.

[10] We refer to equation (6) as the Fama regression because of its similarity to Fama's (1984) regression specification for testing deviations from the uncovered interest rate parity condition. In that context, the fundamental, x_t, corresponds to the log interest rate differential between two countries, y_t is the log of the foreign exchange rate, and rx_t is the currency excess return at time t. See Section 4 for a more detailed discussion of the application of our framework to foreign exchange.

3.2 Subjective Expectations Under Complexity Constraint

Recall from Section 2 that the true data-generating process can be represented by the collection of parameters $\theta^* = (\mathbf{A}^*, \mathbf{B}^*, c^*)$, all of which are of dimension n, whereas agents' models are restricted to the set of k-dimensional models, Θ_k. Given an arbitrary model $\theta \in \Theta_k$ for the agents, denote the Kullback–Leibler (KL) divergence of model θ from the true process by

$$\mathrm{KL}(\theta^* \| \theta) = \mathbb{E}^*[-\log f^\theta(x_{t+1}|x_t, \dots)] - \mathbb{E}^*[-\log f^*(x_{t+1}|x_t, \dots)], \tag{10}$$

where f^* is the density of the fundamental under the true data-generating process, f^θ is the agents' subjective density under model θ, and $\mathbb{E}^*[\cdot]$ is the expectation with respect to the true process. It is well known that the KL divergence is always non-negative and obtains its minimum value of zero if and only if densities f^θ and f^* coincide almost everywhere. As such, $\mathrm{KL}(\theta^* \| \theta)$ can be interpreted as a measure for the disparity between agents' subjective expectations under model θ and the true process.[11]

The next result characterizes the agents' subjective expectations in terms of the KL divergence to the true data-generating process. As discussed in Subsection 2.3, we focus on the outcome of learning under the assumption that the agents have access to a long history of the past realizations of the fundamental. This allows us to study the systematic implications of the complexity constraint for the agents' subjective expectations, while abstracting from the temporary fluctuations in beliefs that arise due to a finite sample.

Proposition 1. *Let $\tilde{\Theta} \subseteq \Theta_k$ denote an arbitrary collection of k-factor models with a positive measure under the agents' prior beliefs, i.e., $\bar{\mu}(\tilde{\Theta}) > 0$. If*

$$\operatorname*{ess\,inf}_{\theta \in \tilde{\Theta}} \ \mathrm{KL}(\theta^* \| \theta) > \operatorname*{ess\,inf}_{\theta \in \Theta_k} \ \mathrm{KL}(\theta^* \| \theta),$$

then

$$\lim_{t_0 \to -\infty} \mu_{t+t_0}(\tilde{\Theta}) = 0 \qquad \mathbb{P}^*\text{-almost surely.}$$

The above result, which is in line with the literature on learning under model misspecification (such as Berk (1966), Esponda and Pouzo (2016, 2020), Rabin and Vayanos (2010), and Molavi (2019)), states that the agents' posterior beliefs concentrate on the subset of k-factor models that have minimum KL divergence to the true data-generating process. Specifically, when agents can contemplate models that are as complex as the true model (that is, when $k \geq n$ and as a result, $\theta^* \in \Theta_k$), Proposition 1 implies that they assign a posterior belief of zero on all models with a non-zero KL divergence to the true process θ^*. In other words, the agents can forecast the future realizations of the fundamental as if they knew the true process (i.e., $\mathbb{E}[\cdot] = \mathbb{E}^*[\cdot]$), in which case our framework coincides with the benchmark of rational expectations. In contrast, when the model complexity

[11] To see this, note that $\mathbb{E}^*[-\log f^\theta(x_{t+1}|x_t, \dots)]$ and $\mathbb{E}^*[-\log f^*(x_{t+1}|x_t, \dots)]$ are entropy-like terms that capture the extent of uncertainty regarding one-step-ahead predictions conditional on the history of past observations under, respectively, model θ and the true model, θ^*. Therefore, the right-hand side of (10) measures the additional uncertainty—and hence, the resulting degradation in prediction quality—when agents use model θ as opposed to the true model.

constraint binds (i.e., when $k < n$ so that $\theta^* \notin \Theta_k$), Proposition 1 implies that agents become convinced, mistakenly, that the fundamental is generated according to the stochastic process that is closest to the true process as measured by KL divergence. In juxtaposition with Lemma 1, this result provides us with the necessary ingredients to characterize the implications of the complexity constraint on the agents' model for return predictability.

Before proceeding any further, however, it is instructive to study the implications of Proposition 1 in the context of a simple example.

Example 1. Suppose the stochastic process that generates the fundamental is governed by n independent-evolving factors, each following an $\mathrm{AR}(1)$ process:

$$\begin{aligned} z_t &= \mathrm{diag}(a_1^*, \ldots, a_n^*) z_{t-1} + \mathrm{diag}(b_1^*, \ldots, b_n^*) \epsilon_t \\ x_t &= [c_1^* \ \ldots \ c_n^*] \, z_t, \end{aligned} \tag{11}$$

where $\mathrm{diag}(p)$ denotes a diagonal matrix whose diagonal entries are given by vector p, $a_i^* \in (-1, 1)$ represents the persistence of the i-th factor, b_i^* parameterizes the volatility of innovations, and c_i^* is the fundamental's loading on that factor.[12] Thus, in the language of equation (1), the process that generates the fundamental is summarized by $\theta^* = (\mathbf{A}^*, \mathbf{B}^*, c^*)$, where $\mathbf{A}^* = \mathrm{diag}(a_1^*, \ldots, a_n^*)$ and $\mathbf{B}^* = \mathrm{diag}(b_1^*, \ldots, b_n^*)$.

Now, suppose the economy is populated by agents who can only entertain models with a single factor, i.e., $k = 1$. Therefore, as long as $n \geq 2$, agents cannot recover the dynamics that govern the fundamental and instead, end up with a one-dimensional approximation to the true process. More specifically, Proposition 1 implies that if agents start with a prior with full support over the set of one-factor models, $\Delta\Theta_1$, as they accumulate more observations, they become increasingly convinced that the fundamental is generated according to $\omega_t = a\omega_{t-1} + b\varepsilon_t$ and $x_t = c\omega_t$, where $\omega_t \in \mathbb{R}$ denotes a single hidden factor and the persistence parameter a is a scalar given by[13]

$$a = \left(\sum_{i=1}^{n} \frac{(b_i^* c_i^*)^2}{1 - a_i^{*2}} \, a_i^* \right) \Bigg/ \left(\sum_{i=1}^{n} \frac{(b_i^* c_i^*)^2}{1 - a_i^{*2}} \right). \tag{12}$$

From the above, it is immediate that when $n = 1$, the agents' posterior beliefs concentrate on models that are observationally equivalent to the true data-generating process. Thus, as expected, the agents' subjective expectations coincide with expectations under the true model. More generally, when $n \geq 2$, equation (12) implies that agents become convinced that the fundamental is generated according to a one-factor model with a persistence parameter that falls in the convex hull of the true persistence parameters, $\{a_1^*, \ldots, a_n^*\}$, with a larger weight on factors that (i) are more persistent (i.e., a larger $|a_i^*|$), (ii) are subject to more volatile innovations (larger $|b_i^*|$), and (iii) are more important drivers of the fundamental (larger $|c_i^*|$).

While simple, the above example highlights the fact that the outcome of the agents' learning process is endogenous to the characteristics of the underlying data-generating process. It also

[12] To ensure that there are no redundant factors, we assume that $a_i^* \neq a_j^*$ for all $i \neq j$ and $b_i^* c_i^* \neq 0$ for all i.

[13] See the appendix for the detailed derivation of equation (12), as well as the expressions for constants b and c in the agents' model.

illustrates that, even when agents are restricted to models that may be significantly less complex than the true data-generating process—and as a result end up with incorrect forecasts about the realizations of the fundamental—they nonetheless recover a lower-dimensional representation that approximates salient features of the true model.

3.3 Characterization

As our next step, we use Lemma 1 and Proposition 1 to characterize how the constraint on the complexity of the agents' models shapes the extent of return predictability, as captured by the term structures of slope coefficients, β_h^{Fama} and β_h^{mom}.

Recall from Section 2 that all parameters of the model agents use for forecasting are endogenous outcomes of learning. As a result, once we specify the maximum number of factors k in the agents' model, there are no more degrees of freedom on how they form their expectations. This, in turn, implies that the coefficients of return-predictability regressions can be expressed only in terms of the number of factors k in the agents' model and the statistical properties of the true data-generating process. We have the following result:

Theorem 1. *Suppose the agents are constrained to k-factor models. Then, the slope coefficients of the Fama and momentum regressions* (6) *and* (7) *are given by*

$$\beta_h^{\text{Fama}} = \delta u'(\mathbf{I} - \delta \mathbf{M})^{-1} u \left(\xi_h^* - \sum_{\tau=1}^{\infty} \phi_\tau \xi_{h-\tau}^* \right) \tag{13}$$

$$\beta_h^{\text{mom}} = \frac{\xi_h^* - \sum_{s=1}^{\infty} \phi_s (\xi_{h-s}^* + \xi_{h+s}^*) + \sum_{s,\tau=1}^{\infty} \phi_s \phi_\tau \xi_{h+s-\tau}^*}{1 - 2\sum_{s=1}^{\infty} \phi_s \xi_s^* + \sum_{s,\tau=1}^{\infty} \phi_s \phi_\tau \xi_{s-\tau}^*}, \tag{14}$$

where $\xi_\tau^* = \mathbb{E}^*[x_t x_{t+\tau}]/\mathbb{E}^*[x_t^2]$ is the autocorrelation of the fundamental, $\phi_s = u'[\mathbf{M}(\mathbf{I} - uu')]^{s-1}\mathbf{M}u$, and $\mathbf{M}$ and u are, respectively, a $k \times k$ weakly stable matrix and a k-dimensional unit vector that minimize

$$H(\mathbf{M}, u) = 1 - 2\sum_{s=1}^{\infty} \phi_s \xi_s^* + \sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty} \phi_s \phi_\tau \xi_{\tau-s}^*. \tag{15}$$

This theorem, which is the main characterization result of the paper, relates the term structure of return predictability coefficients β_h^{Fama} and β_h^{mom} to (i) the number of factors in the agents' model and (ii) the statistical properties of the underlying data-generating process. Importantly, it illustrates that the coefficients of the Fama and momentum regressions depend on the underlying process only through the autocorrelation function (ACF) of the fundamental, $\xi_\tau^* = \mathbb{E}^*[x_t x_{t+\tau}]/\mathbb{E}^*[x_t^2]$, and are independent of the volatility of the fundamental. In other words, irrespective of the relationship between k and n, the agents' posterior beliefs concentrate on models that match the volatility of the fundamental.

Theorem 1 follows from our earlier result that the agents' posterior beliefs concentrate on the subset of k-factor models that have minimum KL divergence to the true data-generating process. Even though the objective function in (15) is in terms of matrix $\mathbf{M}$ and vector u, these objects are closely linked to the parameters of the agents' model, $\theta = (\mathbf{A}, \mathbf{B}, c)$.[14] Finally, we note that the expression $\phi_s = u'[\mathbf{M}(\mathbf{I} - uu')]^{s-1}\mathbf{M}u$, which appears in (13) and (14), has a structural interpretation: $(\phi_1, \phi_2, \dots)$ are the coefficients of the autoregressive representation of the k-factor model that the agents' beliefs converge to. More specifically, the agents' subjective forecast for the realization of the fundamental in the next period is given by $\mathbb{E}_t[x_{t+1}] = \sum_{\tau=1}^{\infty} \phi_\tau x_{t+1-\tau}$.[15]

The characterization result in Theorem 1 serves two distinct purposes in our analysis. First, it allows us to directly apply our framework to various asset pricing applications. In particular, in any context in which endogenous and exogenous variables are related to one another via equation (2), we can use the autocorrelation function of the exogenous variables and the expressions in (13) and (14) to compute the implied coefficients of the Fama and momentum regressions. This is the approach we take in the next section. Second, Theorem 1 also enables us to perform comparative static analyses with respect to the primitives of the economy and to compare the extent of return predictability at different horizons, as we do in the remainder of this section.

We start with a simple result that considers the case where the agents' models are sufficiently rich to fully capture the statistical properties of the data-generating process.

Proposition 2. *Suppose $\delta > 0$ and let k and n denote the number of factors in, respectively, the agents' model and the true data-generating process.*

(a) If $k \geq n$, then $\beta_h^{\text{Fama}} = \beta_h^{\text{mom}} = 0$ for all horizons h.

(b) If $k < n$, then there exists h and $\tilde{h}$ such that $\beta_h^{\text{Fama}} \neq 0$ and $\beta_{\tilde{h}}^{\text{mom}} \neq 0$.

Statement (a) establishes that if agents can contemplate models that are as complex as the true model, then there is no return predictability at any horizon. This is a consequence of the fact that agents in our framework are Bayesian with access to a long history of observations. Bayesian updating implies that agents rule out models that are inconsistent with their past observations, while the large sample size guarantees that they do not suffer from finite-sample problems such as overfitting. They recover the underlying data-generating process even if they use models that have too many parameters relative to the true model (i.e., when $k > n$). Statement (b) of Proposition 2 then shows that the converse implication is also true: if there is no return predictability, the set of models entertained by the agents has to be rich enough to contain the true underlying model. Therefore, return predictability and model misspecification in our framework are one and the same.

Our next result concerns the extent of return predictability at long horizons.

[14] In particular, as we show in the proof of Theorem 1, the eigenvalues of matrix $\mathbf{M}$ that solves (15) coincide with the eigenvalues of matrix $\mathbf{A}$ that minimizes the KL divergence between the two models.

[15] This latter observation also clarifies the reason behind the expression for β_h^{Fama} in (13). It implies that the covariance between the realization of the fundamental at time t and the agents' forecast error at time $t+h$ (i.e., $x_{t+h} - \mathbb{E}_{t+h-1}[x_{t+h}]$) is proportional to $\xi_h^* - \sum_{\tau=1}^{\infty} \phi_\tau \xi_{h-\tau}^*$, which in turn is proportional to β_h^{Fama}, consistent with equation (13). A similar logic is behind the expression in (14) for the momentum regression.

Proposition 3. *Excess returns are not predictable in the long run:*

$$\lim_{h\to\infty} \beta_h^{\text{Fama}} = \lim_{h\to\infty} \beta_h^{\text{mom}} = 0.$$

This result is a direct consequence of the assumptions that (i) the true data-generating process is stationary and (ii) agents only entertain stationary factor models. These stationarity assumptions imply that the effect of time-t variables (including the fundamental x_t and excess returns rx_t) on returns at time $t+h$ die out eventually as h increases, irrespective of whether the complexity constraint binds or not.

3.4 Single-Factor Models

While the optimization problem in Theorem 1—which relates return predictability to the primitives of the economy—does not have a closed-form characterization for a general k, it is possible to obtain such a characterization for the case in which agents are restricted to the class of single-factor models. Despite being only a special case, this closed-form characterization is a transparent and easy-to-use result that is informative about how the extent of return predictability varies with horizon h and the statistical properties of the true data-generating process.

Proposition 4. *If agents are constrained to single-factor models, then the slope coefficients of the Fama and momentum regressions are given by*

$$\beta_h^{\text{Fama}} = \frac{\delta}{1-\delta\xi_1^*}(\xi_h^* - \xi_{h-1}^*\xi_1^*) \tag{16}$$

$$\beta_h^{\text{mom}} = \frac{(1+\xi_1^{*2})\xi_h^* - \xi_1^*(\xi_{h-1}^* + \xi_{h+1}^*)}{1-\xi_1^{*2}}, \tag{17}$$

respectively, where ξ_h^* is the autocorrelation of the fundamental at lag h.

The closed-form characterization in Proposition 4 allows us to explore the determinants of the term structure of the return predictability coefficients in further detail. We are particularly interested in the sign patterns of the Fama and momentum coefficients, i.e., whether higher realizations of the fundamental or returns at time t predict higher or lower excess returns at time $t+h$.

Proposition 5. *Suppose $\xi_h^* > 0$ for all horizon $h \le \bar{h}$. Then, $\beta_h^{\text{Fama}} > 0$ at horizon $h \le \bar{h}$ if and only if*

$$\Delta \log \xi_h^* > \Delta \log \xi_h, \tag{18}$$

where ξ_h is the autocorrelation implied by the agents' model at lag h and $\Delta \log \xi_h = \log \xi_h - \log \xi_{h-1}$ denotes its growth rate.

According to this result, what determines the sign of the slope coefficient of the Fama regression at some horizon h is whether the ACF of the true data-generating process at that horizon grows at a faster rate (or, equivalently, decays at a slower rate) compared to the autocorrelation implied by the agents' lower-dimensional model.[16]

[16] The assumption that $\xi_h^* > 0$ is to ensure that the logarithms in (18) are well defined.

The characterization in Proposition 5 is in terms of an endogenous object that depends on the outcome of the agents' learning process, namely the autocorrelation implied by the model their beliefs converge to.[17] Therefore, as our next result, we provide a characterization of the sign pattern of the term structure of β_h^{Fama} in terms of the primitives of the economy. In particular, we show that the spectrum of matrix $\mathbf{A}^*$ that governs the evolution of the fundamental plays a central role in shaping the signs of the coefficients of the Fama regression.

Proposition 6. *Suppose $k = 1$ and $n = 2$. Furthermore, suppose $\mathbf{A}^*$ has two distinct eigenvalues, λ_1 and λ_2, such that $|\lambda_1| \geq |\lambda_2|$.*

(a) If both eigenvalues are real and $\lambda_1 > 0$, then β_h^{Fama} never changes sign.

(b) If both eigenvalues are real and $\lambda_1 < 0$, then β_h^{Fama} changes sign at every horizon.

(c) If both eigenvalues are complex, then β_h^{Fama} changes sign at horizons $h = r\pi/|\varphi|$ for $r \in \mathbb{N}$, where $\varphi = \arg(\lambda_1)$ is the argument (or the phase) of λ_1 in its polar representation.

Our next result concerns the sign pattern of the coefficients of the momentum regression. It establishes that as long as the process that drives the fundamental exhibits some persistence in the very short run—and irrespective of any of its other characteristics—returns are not only predictable, but also exhibit both momentum and reversal.

Proposition 7. *Suppose $k = 1$ and $n \geq 2$. Furthermore, suppose $\xi_1^* > 0$. Then, there exist h and $\tilde{h}$ such that $\beta_h^{\text{mom}} > 0 > \beta_{\tilde{h}}^{\text{mom}}$.*

3.5 Heterogenous-Agent Economy

Our results thus far relied on the assumption that the economy consists of a unit mass of identical agents, all of whom are restricted to using models with the same maximum number of factors, k. In this subsection, we extend our previous results by assuming that only a fraction $1 - \gamma$ of the agents are subject to our behavioral constraint, while the remaining γ fraction can entertain models with any number of factors. For simplicity, we refer to the two groups of agents as behavioral and rational agents, respectively.[18]

The heterogeneity in the agents' ability to entertain statistical models of different complexities results in heterogenous subjective expectations. Therefore, as in Allen, Morris, and Shin (2006), we assume that the relationship between endogenous prices and exogenous fundamentals is given by the following generalization of equation (2):

$$y_t = x_t + \delta \bar{\mathbb{E}}_t[y_{t+1}], \tag{19}$$

[17] Though, see Appendix B.1, where we characterize the ACF implied by the agents' model in terms of the solution of optimization problem (15).

[18] Despite the terminology, recall that all agents in this economy are Bayesian, with the only difference between the two groups being that the "behavioral" agents assign zero prior beliefs to models consisting of more than k factors. Furthermore, note that when $k \geq n$, agents in both groups end up with the same exact subjective expectations. The subjective expectations of the two groups differ only when $k < n$.

where $\bar{\mathbb{E}}[\cdot] = \gamma\mathbb{E}^*[\cdot]+(1-\gamma)\mathbb{E}[\cdot]$ denotes the cross-sectional average of agents' expectations.[19] Iterating on the above, we can also obtain the following counterpart to equation (5) for excess returns in terms of agents' expectations:

$$\mathrm{rx}_{t+1} = \sum_{\tau=1}^{\infty} \delta^{\tau} \left(\bar{\mathbb{E}}_{t+1}\bar{\mathbb{E}}_{t+2} \cdots \bar{\mathbb{E}}_{t+\tau}[x_{t+\tau}] - \bar{\mathbb{E}}_{t}\bar{\mathbb{E}}_{t+1} \cdots \bar{\mathbb{E}}_{t+\tau}[x_{t+\tau}] \right). \tag{20}$$

The key observation is that, even though subjective expectations of each group of agents satisfy the law of iterated expectations, the cross-sectional average expectation $\bar{\mathbb{E}}[\cdot]$ may not. Therefore, unlike the representative-agent framework of Section 2, excess returns in the heterogenous-agent economy also depend on higher-order expectations whenever $k < n$. The failure of the law of iterated expectations with respect to $\bar{\mathbb{E}}[\cdot]$ in our framework resembles a similar phenomenon in differential-information economies.[20]

Our next result characterizes the slope coefficient of the Fama regression (6) in the heterogenous-agent economy in terms of the corresponding family of coefficients in the representative-agent economy consisting of only behavioral agents (i.e., $\gamma = 0$). Let $\beta_h^{\text{Fama}}(\gamma)$ denote the slope coefficient of the Fama regression in an economy with γ and $1-\gamma$ fraction of rational and behavioral agents, respectively. We obtain this result under the assumption that while rational agents can recover the model used by behavioral agents, behavioral agents behave as if they live in a representative-agent economy only consisting of agents with k-factor models. We make this assumption to capture the idea that, given their priors, behavioral agents are convinced—mistakenly so when $k < n$—that a k-factor model is sufficient to capture the process that drives the fundamental.

Proposition 8. *The slope coefficient of the Fama regression in the heterogenous-agent economy is given by*

$$\beta_h^{\text{Fama}}(\gamma) = (1-\gamma) \sum_{s=0}^{\infty} (\delta\gamma)^s \beta_{h+s}^{\text{Fama}}(0), \tag{21}$$

where $\beta_h^{\text{Fama}}(0)$ is the slope coefficient of the representative-agent economy and is given by (13).

The above result relates the term structure of β_h^{Fama} in the heterogenous-agent economy to that of the economy only consisting of behavioral agents. This allows us to use our previous results to characterize the extent and nature of return predictability in economies consisting of both rational and behavioral agents. The key observation is that $\beta_h^{\text{Fama}}(\gamma)$ is not simply a weighted average of $\beta_h^{\text{Fama}}(0)$ and $\beta_h^{\text{Fama}}(1) = 0$. Rather, the fact that cross-sectional average expectations $\bar{\mathbb{E}}[\cdot]$ do not satisfy the law of iterated expectations implies that return predictability at horizon

[19] We assume that $\bar{\mathbb{E}}[\cdot]$ in equation (19) is the unweighted cross-sectional average of the agents' expectations. See the overlapping generations model of Allen, Morris, and Shin (2006) for a simple micro-foundation. More generally, depending on the underlying micro-founded model, $\bar{\mathbb{E}}[\cdot]$ is a weighted average of the agents' subjective expectations, with endogenous (and potentially state-dependent) weights (Panageas, 2020). Given our focus on a reduced-form framework, we abstract from these issues.

[20] See, for example, Allen, Morris, and Shin (2006), Barillas and Nimark (2017), Angeletos and Lian (2018), and Angeletos and Huo (2020), where agents have access to private signals about fundamentals. In contrast to these papers, all information in our framework is public and it is the heterogeneity in the maximum number of factors in agents' models that results in heterogenous expectations and the potential violation of the law of iterated expectations.

h in the heterogenous-agent economy also depends on the extent of return predictability in the representative-agent economy at all horizons $\tau \geq h$.

4 Applications

In this section, we apply our framework to two asset pricing applications: the violation of uncovered interest rate parity in foreign exchange and time-series momentum and reversal in equity returns. More specifically, we use the characterization result in Theorem 1 to test our model's predictions for the slope coefficients of return predictability regressions (6) and (7) in these contexts.

4.1 Reversal of Uncovered Interest Rate Parity

One of the central tenets of international finance is the uncovered interest rate parity (UIP) condition, which maintains that high interest rate currencies should depreciate vis-à-vis those with low interest rates. Yet—in what has become known as the "forward discount puzzle"—a vast empirical literature documents that, over short time horizons (ranging from a week to a quarter), high interest rate currencies tend to appreciate. In other words, short-term deposits of high-interest rate currencies tend to earn a predictively positive excess return.

More recently, however, Bacchetta and van Wincoop (2010) and Engel (2016) document a distinct but related puzzle, known as the "predictability reversal puzzle." They find that UIP violations reverse sign over longer horizons, with high interest rate currencies earning negative excess returns at horizons from four to seven years. The seemingly contradictory implications of the forward discount and predictability reversal puzzles for the relationship between currency excess returns and interest rate differentials has led some to argue for the inadequacy of existing models for explaining UIP violations. For example, Engel (2016) argues that, risk-based explanations of the forward discount puzzle—which attribute the violations of UIP to the relative riskiness of holding short-term deposits in the high-interest rate country—cannot account for the predictability reversal puzzle.

In this subsection, we apply our theoretical results to study the implications of constraints on investors' model complexity for the pattern of UIP violations at different horizons and investigate the extent to which our behavioral framework can jointly explain the forward discount and predictability reversal puzzles.

We map this context to our framework in Section 2 by letting the fundamental denote the log interest rate differential between the U.S. and a foreign country, i.e., $x_t = i_t^* - i_t$, where i_t and i_t^* are nominal interest rates on deposits held in U.S. dollars and the foreign currency, respectively. We also let y_t denote the log of the foreign exchange rate, expressed as the U.S. dollar price of the foreign currency. With the discount rate set to $\delta = 1$, equation (2) is then nothing but the interest rate parity condition, according to which an increase in the U.S. to foreign short-term interest rate differential is associated with an exchange rate appreciation, whereas a higher expected future exchange rate implies a depreciation. Note that, as in Section 2, the expectation in (2) denotes the

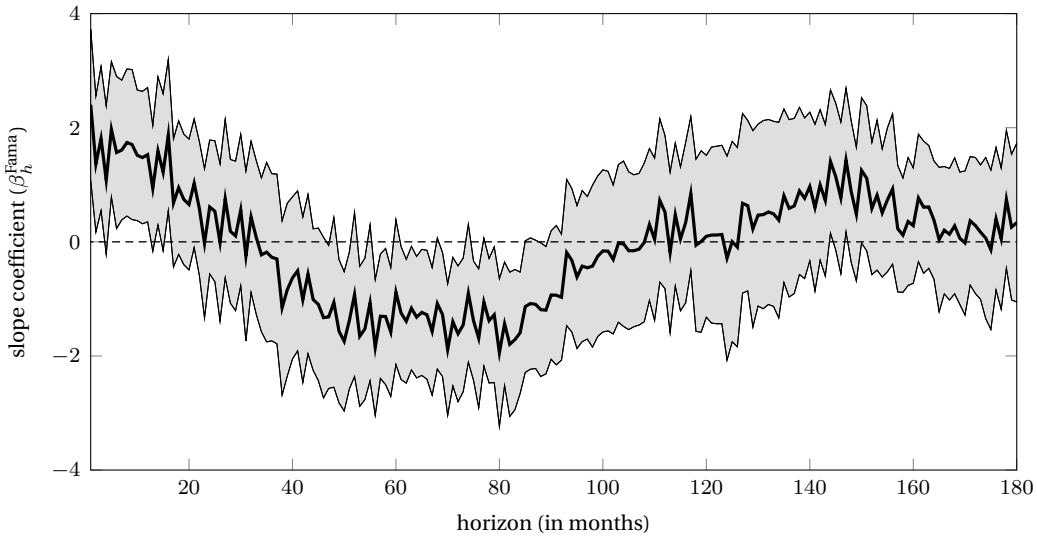


Figure 1. Violation of Uncovered Interest Rate Parity at Different Horizons

Notes: This figure plots estimated slope coefficients together with 90 percent confidence intervals of the Fama regression: $\text{rx}_{t+h} = \alpha_h^{\text{Fama}} + \beta_h^{\text{Fama}}(i_t^* - i_t) + \epsilon_{t,h}$, where $h = 1, \ldots, 180$ denotes the horizon in months, rx_{t+h} is the currency excess return on a trade-weighted average of seven currencies vis-à-vis the U.S. dollar, and $i_t^* - i_t$ is the log interest rate differential between the trade-weighted average interest rate and the U.S. interest rate. Data is monthly and runs from January 1985 to December 2019. Confidence intervals are calculated using Newey and West (1987) standard errors.

agents' subjective expectations, which may differ from those arising from the true data-generating process. Finally, equation (3) is simply the definition of currency excess returns:

$$\text{rx}_{t+1} = y_{t+1} - y_t + (i_t^* - i_t).$$

We start by reproducing the empirical findings on UIP violations at different horizons. Following Engel (2016), we build a trade-weighted average exchange rate and interest rate differential relative to the U.S. for the following countries: Australia, Canada, Euro (Germany before its introduction), New Zealand, Japan, and the United Kingdom. The weights are constructed as the value of each country's exports and imports as a fraction of the average value of trade over the six countries. Monthly exchange rate data is from Datastream and interest rate differentials are calculated using covered interest rate parity from forward rates, $i_t^* - i_t = f_t - y_t$, also available from Datastream. We then run the following family of regressions:

$$\text{rx}_{t+h} = \alpha_h^{\text{Fama}} + \beta_h^{\text{Fama}}(i_t^* - i_t) + \epsilon_{t,h}, \qquad h = 1, \ldots, 180,$$

where h is the horizon measured in months. This regression is, of course, identical to the Fama regression (6) in Section 2.

Figure 1 plots the estimated slope coefficients at various horizons. For $h = 1$, we find the slope coefficient to be positive, thus recovering the classic forward discount puzzle: at short time horizons, higher interest rate differentials (relative to the U.S.) lead to higher excess returns. This pattern remains the same up to a horizon of three years, but then reverses its sign, illustrating the predictability reversal puzzle: for horizons between four to seven years, higher interest rates predict

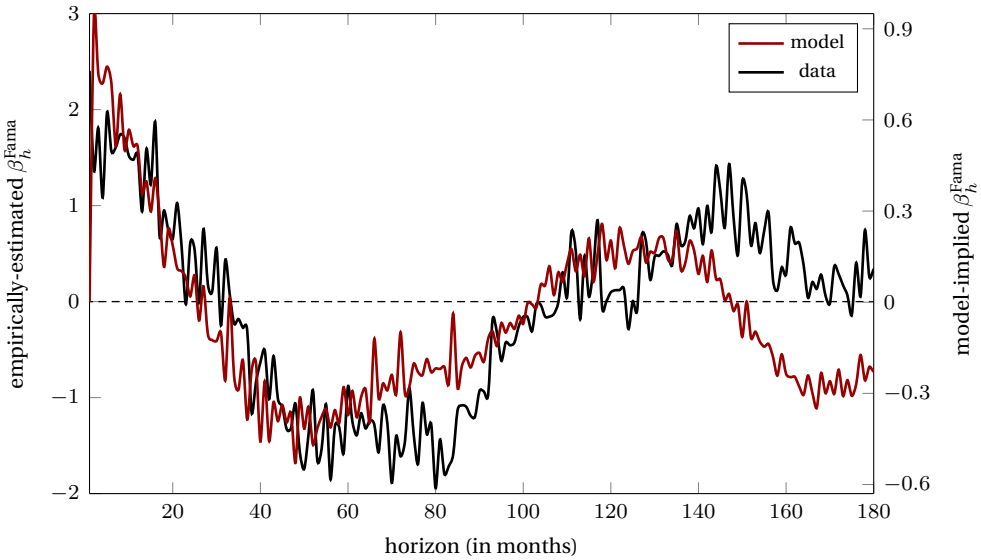


Figure 2. One-Factor Model-Implied Fama Coefficient

Notes: This figure plots estimated Fama slope coefficients from Figure 1 (left axis) together with model-implied (right axis) betas from a one-factor model given in Proposition 4 for horizons 1 to 180 months. Data are monthly and run from January 1985 to December 2019.

lower excess returns. Finally, as the figure indicates, the estimated coefficient of the Fama regression becomes indistinguishable from zero at even longer horizons.

Turning to our framework's predictions, we first calculate the autocorrelation function $\{\xi_\tau^*\}_{\tau\geq 1}$ of the interest rate differential for the trade-weighted basket of currencies against the U.S. dollar. Taking this autocorrelation as our primitive, we then calculate the term structure of the model-implied coefficients of the Fama regression using expression (13) in Theorem 1 for different number of factors, k, in the agents' model.

As our first exercise, we consider the case in which agents can only entertain single-factor models, i.e., $k = 1$. Recall that in this special case, we can use the closed-form expression (16) in Proposition 4 to calculate the model-implied slope coefficients of the Fama regression. Figure 2 plots the term structure of the model-implied coefficients for $k = 1$ together with the coefficients obtained from the data from Figure 1. As the figure indicates, the pattern of model-implied coefficients tracks the pattern observed in the data fairly closely (though not its magnitude). Most importantly, we see a reversal in the slope coefficient: model-implied coefficients are positive for horizons up to 26 months and reverse to a negative sign thereafter. Figure 2 thus suggests that if agents are constrained to rely on the family of single-factor models to make forecasts about the evolution of the interest rate differential process, their subjective expectations result in return predictability patterns that are simultaneously consistent with the forward discount and predictability reversal puzzles.

It is important to emphasize that while the estimated coefficients in Figure 1 are obtained from regressing returns on interest rate differentials, the model-implied coefficients in Figure 2 do not use the data on exchange rates or excess returns. Rather, they are simply obtained by plugging the autocorrelation of the interest rate differential into equation (16).

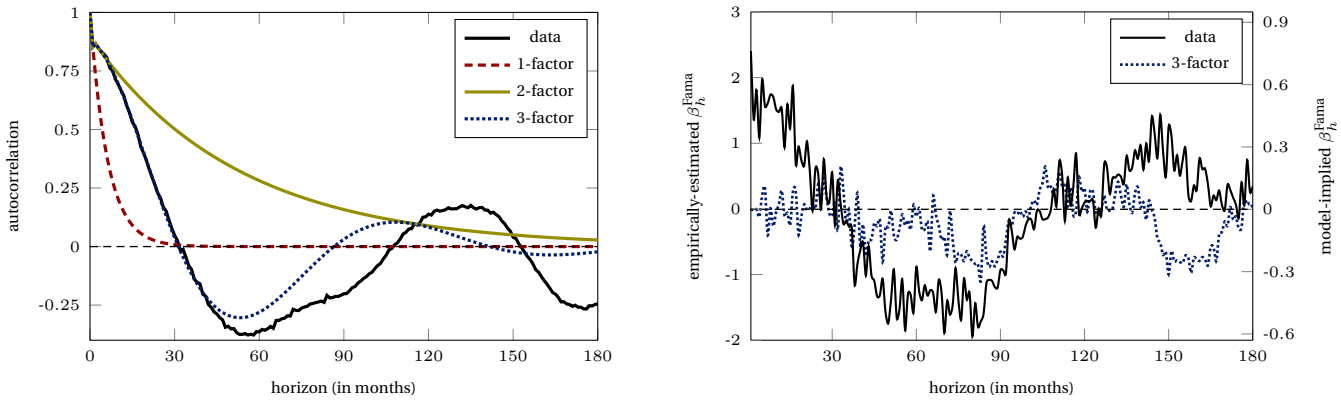


Figure 3. Autocorrelation Function and Three-Factor-Model-Implied Fama Coefficients

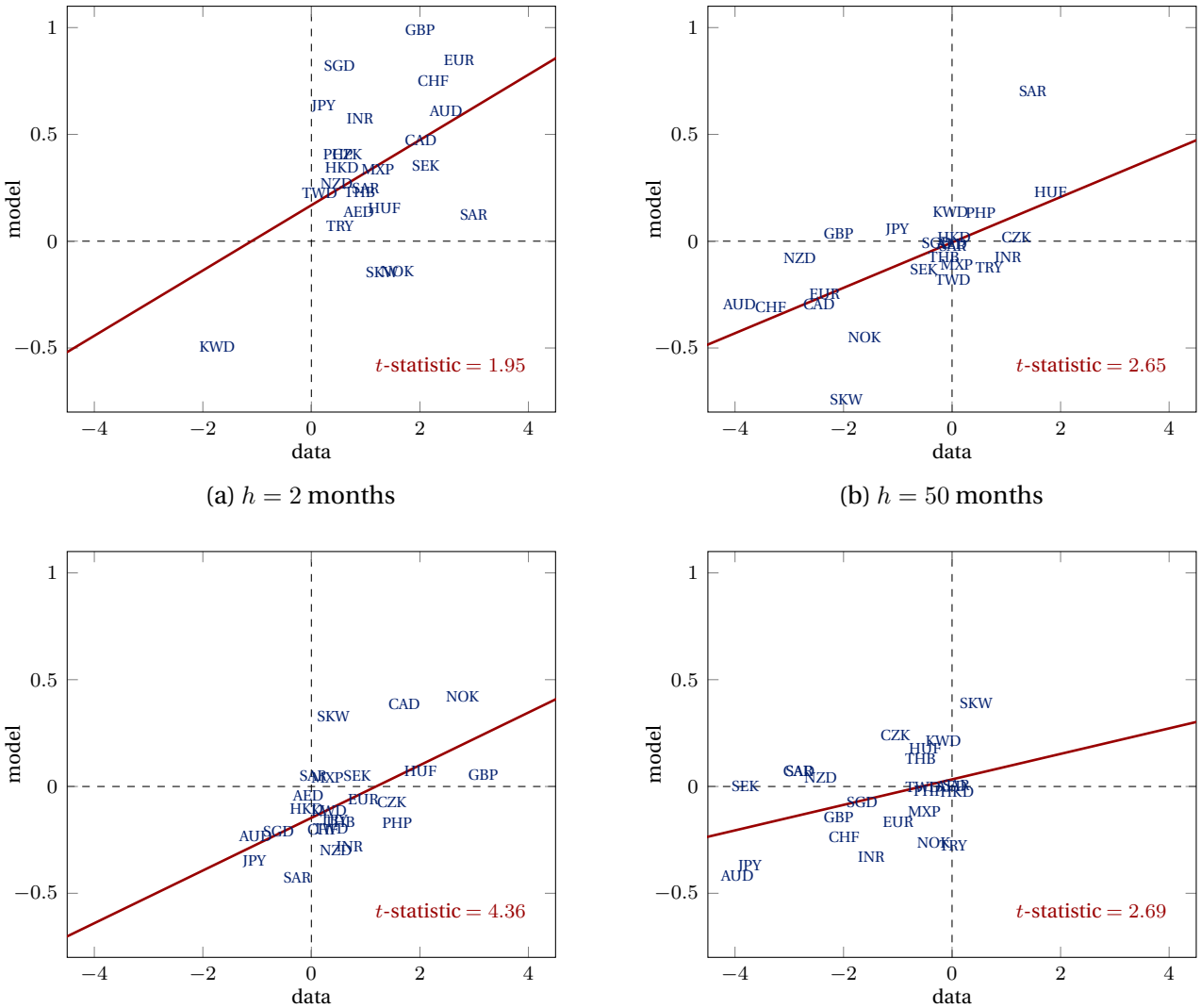
Notes: The left panel plots the autocorrelation function of the trade-weighted interest rate differential (data) together with one-, two-, and three-factor-model-implied ACFs. The right panel plots the model-implied Fama slope coefficients for horizons 1 to 180 months for the three-factor model. Data are monthly and run from January 1985 to December 2019.

Next, we investigate how increasing the number of factors, k, in the agents' models impacts the term structure of model-implied slope coefficients. To this end, we once again use the empirical autocorrelation function of the interest rate differential as an input to calculate β_h^{Fama} for $k = 2$ and 3. However, when $k > 1$, there is no closed-form expression for the model-implied slope coefficients. As a result, we use equation (13) and the characterization result in Theorem 1 to solve for β_h^{Fama} numerically. Importantly, as a by-product, we also obtain the model-implied autocorrelation function, i.e., the autocorrelation function from the (potentially incorrect) perspective of agents who use a k-factor model.[21]

The results are reported in Figure 3. The left panel depicts the model-implied ACF for $k = 1, 2, 3$, together with the empirical ACF of the trade-weighted average interest rate differential for horizons 0 to 180 months. As the figure indicates, the model-implied ACFs can differ quite substantially across various levels of agents' sophistication. For example, while the three-factor model exhibits patterns that are fairly similar to the empirical ACF, the ACF implied by the single-factor model looks significantly different. Crucially, this is reflected in the model-implied slope coefficients as illustrated in the right panel of Figure 3: the three-factor model, which generates a model-implied ACF that tracks the empirical ACF very closely, also results in model-implied slope coefficients that are significantly smaller at all horizons. This, of course, is to be expected in view of our results in Section 3. As agents are able to entertain richer and more complex statistical models, they end up with models that better fit the empirical ACF, which in turn results in less significant deviations from the rational expectations benchmark and hence less return predictability.

Comparing the ACF of the true interest rate differential process to the ACF implied by the agents' model can also shed light on why our framework can generate return predictability patterns that are simultaneously consistent with the forward discount and predictability reversal puzzles. Recall from Proposition 5 that when agents are constrained to single-factor models, $\beta_h^{\text{Fama}} > 0$ if and only

[21] See Appendix B.1 for the details of how the model-implied autocorrelation function can be calculated in terms of the solution of optimization problem (15).



(a) $h = 2$ months

(b) $h = 50$ months

(c) $h = 100$ months

(d) $h = 120$ months

Figure 4. Cross-Section of Fama Coefficients

Notes: This figure plots the empirically-estimated Fama regression coefficients against the corresponding model-implied coefficients for 21 currencies against the U.S. dollar at $h = 2$, 50, 100, and 120 month horizons. The red line in each panel indicates the corresponding least-square fit. The currencies are Australia, Canada, Czech Republic, Euro, Hong Kong, Hungary, India, Japan, Kuwait, Mexico, New Zealand, Norway, Philippines, Singapore, South Africa, South Korea, Sweden, Switzerland, Taiwan, Thailand, Turkey, the United Arab Emirates, and the United Kingdom. Data are monthly and run from October 1997 to December 2019.

if $\Delta \log \xi_h^* > \Delta \log \xi_h$. Therefore, as the left panel of Figure 3 illustrates, the slope coefficient of the Fama regression has to be positive for short horizons—when the dashed red line that represents the model-implied ACF decays at a faster rate than the solid black line that plots the true ACF—but should change sign at horizon $h = 26$ months (when the model-implied ACF decays more slowly). This is of course consistent with the pattern we documented in Figure 2.

We can further explore the relationship between the interest rate differentials and the model-implied slope coefficients by using the result in Proposition 6, according to which the sign pattern of model-implied coefficients for $k = 1$ is tightly linked to the spectrum of matrix $\mathbf{A}^*$ that governs

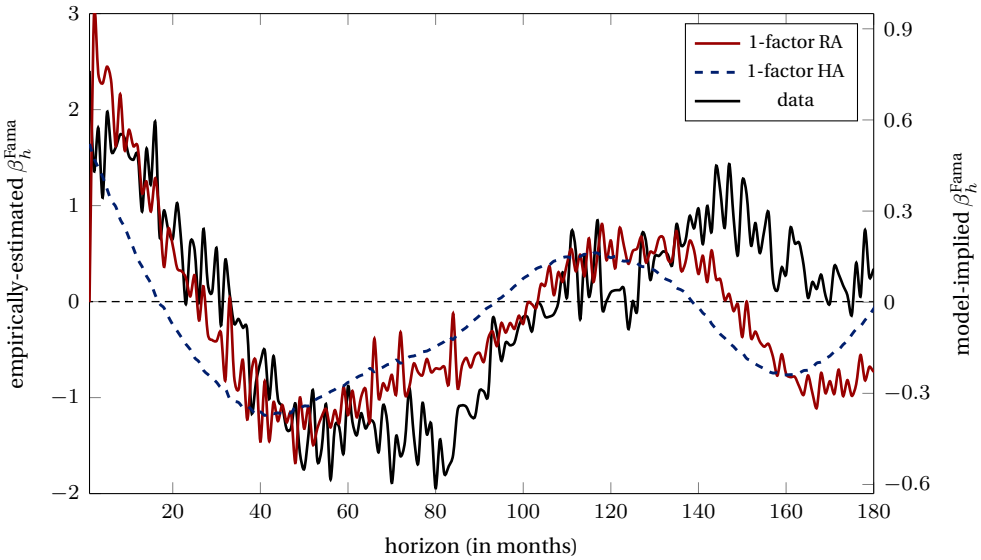


Figure 5. Fama Coefficients for Representative- and Heterogeneous-Agent Economies

Notes: This figure plots the single-factor model-implied slope coefficients of the Fama regression for horizons 1 to 180 months in a representative- and heterogeneous-agent economies (right axis) together with the corresponding coefficients estimated from the data (left axis). The fraction of behavioral agents in the heterogenous-agent economy is 10%, i.e., $\gamma = 0.9$. Data are monthly and run from January 1985 to December 2019.

the dynamics of the fundamental (in this case, the interest rate differential process). Given the close similarity between the ACFs of the three-factor model and the interest rate differential process (Figure 3), we extract the eigenvalues of the estimated three-factor model and treat them as approximations to the top three eigenvalues of $\mathbf{A}^*$. We find that the eigenvalues with the largest moduli form a complex conjugate pair with arguments $\varphi = \pm 0.06$. Therefore, to the extent that the three-factor model provides a close approximation to the true interest rate differential process, Proposition 6 implies that β_h^{Fama} should change signs roughly every $\pi/\varphi = 52$ months. This is consistent with the pattern in Figure 2, which exhibits three zero crossings within the 150-month horizon.

As a further test of our model's predictions, we calculate the model-implied slope coefficients of the Fama regression for a larger cross-section of countries and plot them against the corresponding slope coefficients estimated from the data.[22] Figure 4 depicts the results for $k = 1$ at different horizons. Recall from Theorem 1 and Proposition 4 that model-implied slope coefficients depend on the shape of the ACF of the fundamental (in this case, the interest rate differential between the corresponding country and the U.S.). Therefore, as the ACFs differ in the cross-section of currencies, so should the model-implied betas. Nonetheless, as Figure 4 illustrates, the model-implied and empirically-estimated slope coefficients have a positive and statistically-significant relationship.

[22] We use currency pairs for the following countries (all against the U.S. dollar): Australia, Canada, Czech Republic, Euro, Hong Kong, Hungary, India, Japan, Kuwait, Mexico, New Zealand, Norway, Philippines, Singapore, South Africa, South Korea, Sweden, Switzerland, Taiwan, Thailand, Turkey, United Arab Emirates, and the United Kingdom. Due to data quality issues in the beginning of the sample, we start in October 1997 when all countries have exchange rate and interest rate differential data available.

This positive relationship holds both at short horizons (such as two months) when most Fama coefficients are positive, as well as for the longer horizons (such as 120 months) when coefficients tend to be mostly negative.

As a final exercise, we test whether the above findings are robust to the introduction of heterogeneity in the number of factors in the agents' models. To this end, we use the characterization in Proposition 8 to calculate the model-implied slope coefficients of the Fama regression in a heterogeneous-agent economy, in which fraction $1-\gamma$ of agents are constrained to using a single-factor model, while the remaining γ fraction can entertain models with any number of factors. Figure 5 plots the implied coefficients in an economy populated by 90% rational and 10% behavioral agents, i.e., for $\gamma = 0.9$. As the figure illustrates, the model-implied slope coefficients in the heterogeneous-agent economy look very similar to those in the representative-agent economy consisting of only behavioral agents (i.e., $\gamma = 0$). This indicates that even small fractions of behavioral agents can lead to notable deviations from the rational expectations benchmark, generating patterns that are consistent with the slope coefficients in the data.

4.2 Time-Series Momentum and Reversal in Equity Returns

One of the starkest challenges to the "random walk hypothesis" of asset prices is the existence of time-series momentum and reversal, whereby past returns predict future returns. For example, Moskowitz, Ooi, and Pedersen (2012) document that returns of a diverse set of futures and forward contracts exhibit persistence for one to 12 months, an effect that partially reverses over longer horizons.

As a second illustration of our framework, we focus on time-series momentum and reversal in equity returns. As in the previous subsection, we start by reproducing the empirical findings that show the existence of return predictability. We then apply our theoretical results from Section 3 and compare the degree of return predictability in the data to that implied by our framework.

We use MSCI price and total return indices for Australia, Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom, and the United States (from Datastream). Since volatility varies across the different country indices, we follow Moskowitz, Ooi, and Pedersen (2012) and scale excess returns by their lagged volatility.[23] From the raw return series, we also calculate dividend growth rates.[24]

As a first exercise, we follow Cutler, Poterba, and Summers (1991) and calculate average autocorrelations of returns across the different equity indices, with the results reported in Table 1. The first column illustrates the well-known pattern of positive serial correlation over horizons shorter than one year, indicating short-term time-series momentum. The average autocorrelations, however, turn negative over the horizon of 12–24 months, as is evident from the second column of Table 1, pointing towards reversals in excess returns at longer horizons.

[23] The conditional return volatilities are calculated using a $\mathrm{GARCH}(1,1)$ model. Our results remain unchanged without this normalization.

[24] Because dividends feature a strong seasonal component, we follow Ang and Bekaert (2006) and Golez (2014) and construct dividend growth series based on a twelve-month trailing sum of dividends.

Table 1. Autocorrelations for Equity Excess Returns

	Horizon (in months)			
	1–12	13–24	25–36	37–48
Australia	0.0004	−0.0189	0.0057	−0.0121
Belgium	0.0335	−0.0035	−0.0025	−0.0282
Canada	0.0086	−0.0336	−0.0009	0.0034
France	0.0220	−0.0240	−0.0065	−0.0151
Germany	0.0134	−0.0217	−0.0057	−0.0023
Italy	0.0325	−0.0174	−0.0157	−0.0241
Japan	0.0508	−0.0198	−0.0023	−0.0294
The Netherlands	0.0140	−0.0159	0.0000	−0.0038
Sweden	0.0137	−0.0336	0.0054	−0.0051
Switzerland	0.0161	−0.0068	−0.0059	−0.0161
United Kingdom	0.0189	−0.0043	−0.0128	−0.0205
United States	0.0267	−0.0131	0.0078	0.0017

Notes: This table reports average autocorrelations of MSCI country excess returns for the 12 months indicated time period. Data are monthly and run from January 1971 to December 2019.

We also test for the extent of return predictability by running a pooled panel regression, as in Moskowitz, Ooi, and Pedersen (2012), of the form

$$\mathrm{rx}^s_{t+h} = \alpha^{\mathrm{mom}}_h + \beta^{\mathrm{mom}}_h \mathrm{rx}^s_t + \epsilon^s_{t,h}, \tag{22}$$

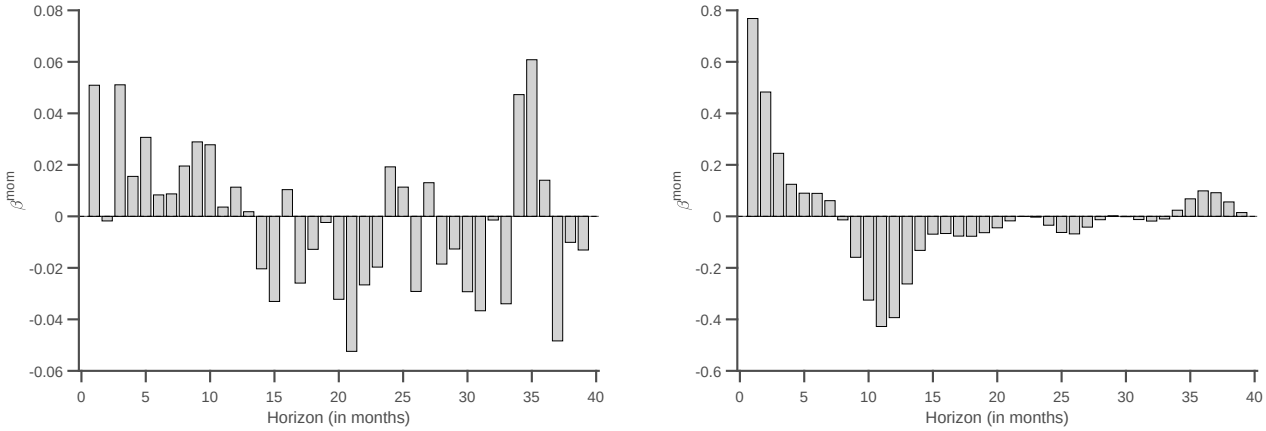
where rx^s_t is the excess return of equity index s at time t. The left panel in Figure 6 plots the slope coefficients of the pooled regression for $h = 1, 2, \ldots, 40$ months. The results echo the findings in Table 1: the estimated slope coefficients are positive up to twelve months (thus indicating short-term momentum) and turn negative at longer horizons (indicating long-term reversal).

To apply our framework to this context, we start with Campbell's (1991) return decomposition, according to which, to a first-order approximation,

$$\begin{aligned} \log R_{t+1} - \mathbb{E}_t[\log R_{t+1}] = &\sum_{\tau=0}^{\infty} \rho^\tau (\mathbb{E}_{t+1}[\Delta \log d_{t+\tau+1}] - \mathbb{E}_t[\Delta \log d_{t+\tau+1}]) \\ &- \sum_{\tau=1}^{\infty} \rho^\tau (\mathbb{E}_{t+1}[\log R_{t+\tau+1}] - \mathbb{E}_t[\log R_{t+\tau+1}]), \end{aligned} \tag{23}$$

where R_{t+1} denotes the return on equity between t and $t+1$, $\Delta \log d_{t+1}$ denotes dividend growth, $\rho < 1$ is a positive constant, and $\mathbb{E}_t[\cdot]$ denotes investors' subjective expectations. Assuming constant expected returns, the above equation reduces to equation (5) in our framework, in which the fundamental, x_t, corresponds to the dividend growth process and rx_{t+1} on the left-hand side is the equity excess return.[25] Furthermore, equation (22) becomes the empirical counterpart to the

[25]For example, if investors are risk neutral with discount factor δ, then $\mathbb{E}_t[\log R_{t+1}] = -\log \delta$, thus guaranteeing that



(a) empirically-estimated β_h^{mom}

(b) model-implied β_h^{mom}

Figure 6. Time-Series Momentum and Reversal

Notes: This figure plots the slope coefficients of the pooled regression $\text{rx}_{t+h}^s = \alpha_h^{\text{mom}} + \beta_h^{\text{mom}} \text{rx}_t^s + \epsilon_{t,h}^s$ for $h = 1, \ldots, 40$ months and rx_t^s is the excess return of index s from the data (left panel) and the model-implied regression coefficient (right panel). Data are monthly and run from January 1971 to December 2019.

momentum regression (7) in Section 2.

Given the mapping to our framework, we can now use our characterization results in Section 3 to calculate the term structure of the model-implied slope coefficients of the momentum regression. To this end, we first calculate the autocorrelation function, $\{\xi_\tau^*\}_{\tau \geq 1}$, of the dividend growth process for each of the country equity indices in our sample. Then, assuming that investors can only entertain single-factor models, we use equation (17) to obtain the model-implied β_h^{mom} for each index at various horizons. The right panel of Figure 6 plots the resulting term structure of model-implied slope coefficients under a pooled regression specification.[26] Comparing the two panels of Figure 6 illustrates that our framework generates return predictability patterns that are similar to what is observed in the data: slope coefficients are positive at short horizons and turn negative at longer horizons.

We conclude by emphasizing that while the empirically-estimated slope coefficients are obtained from regressing excess returns on lagged returns, the model-implied coefficients are calculated solely from the autocorrelation of dividend growth process (as prescribed by Proposition 4) and without using returns data. We also note that our framework generates momentum and reversal without relying on time-varying (fundamental) risk premia.

the second sum on the right-hand side of (23) is equal to zero. As already mentioned, we abstract from time-varying discount rates in order to study, in the most transparent manner, how return dynamics are shaped by the constraint on the complexity of agents' models.

[26] Equation (17) in Proposition 4 allows us to obtain the model-implied slope coefficients, $\beta_h^{\text{mom,s}}$, for each country equity index s separately. The corresponding slope coefficient at horizon h for the pooled regression can then be expressed in terms of country-specific coefficients as $\beta_h^{\text{mom,pooled}} = \sum_s w_s \beta_h^{\text{mom,s}}$, for weights w_s that sum up to 1 and are proportional to $\Xi_0^{*,s}(1 - (\xi_1^{*,s})^2)/(1 - \delta\xi_1^{*,s})^2$, where $\xi_h^{*,s}$ and $\Xi_0^{*,s}$ denote the autocorrelation and the unconditional variance of the fundamental process in country s, respectively.

5 Conclusions

This paper studies how limits to the complexity of statistical models that agents use for forecasting shape asset prices, where we define the complexity of a model as the dimension of its minimal representation. We develop our results in the context of a simple framework in which a sequence of exogenous fundamentals are generated by a stochastic process that may be more complex than what agents can entertain. As a result, agents form their subjective expectations by relying on lower-dimensional approximations to the underlying data-generating process. The constraint on the complexity of agents' models is our only point of departure from the textbook rational expectations framework: we impose no-arbitrage, maintain the assumption of Bayesian updating, and assume that agents exhibit no other behavioral biases.

As our main theoretical result, we characterize how the statistical properties of the true data-generating process together with the limit on the number of factors in agents' models shape return dynamics. These results provides us with a sharp characterization of the term structures of the slope coefficients of the Fama and momentum regressions. We then apply our framework to two applications in asset pricing: (i) the violations of the uncovered interest rate parity in foreign exchange and (ii) time-series momentum and reversal in equity returns. In both cases, we find that the resulting deviations from rational expectations can generate return predictability patterns that are broadly consistent with the data.

In order to obtain a tractable framework, we made a number of simplifying assumptions. First, we focused on how relying on lower-dimensional approximations to the data-generating process shapes the term structure of return predictability for a single asset. However, phenomena such as momentum and reversal are pervasive, not just in the time series, but also in the cross section. Extending our framework to multiple assets and exploring the possible implications of constraints on model complexity for cross-sectional return predictability would be a natural next step for future work. Second, we focused on a simple environment in which prices relate to fundamentals via the simple no-arbitrage condition in (2). While this allowed us to isolate the implications of agents' inability to entertain high-dimensional models for return dynamics in a transparent manner, we abstracted from other important features, such as time-varying (fundamental) risk premia. Integrating our behavioral assumption into a more structural setting with general preferences would shed further light on how distortions in subjective expectations caused by constraints on model complexity impact asset prices.

A Proofs and Derivations

Proof of Lemma 1

Recall that since agents' subjective expectations satisfy the law of iterated expectations, excess returns are equal to the discounted sum of agents' forecast revisions given in (5). Furthermore, equation (6) implies that the slope coefficient of the Fama regression at horizon h is given by $\beta_h^{\text{Fama}} = \mathbb{E}^*[x_t \, \text{rx}_{t+h}]/\mathbb{E}^*[x_t^2]$, where $\mathbb{E}^*[\cdot]$ denotes the expectation with respect to the true data-generating process. Therefore, together with (5), this implies that β_h^{Fama} satisfies (8).

To establish equation (9), observe that equation (5) implies that

$$\mathbb{E}^*[\text{rx}_t \text{rx}_{t+h}] = \sum_{\tau=1}^{\infty} \sum_{s=1}^{\infty} \delta^{\tau+s} \left(\mathbb{E}_{t+1}[x_{t+\tau}] - \mathbb{E}_t[x_{t+\tau}]\right) \left(\mathbb{E}_{t+h+1}[x_{t+h+s}] - \mathbb{E}_{t+h}[x_{t+h+s}]\right).$$

This, coupled with the fact that slope coefficient of the momentum regression is given by $\beta_h^{\text{mom}} = \mathbb{E}^*[\text{rx}_t \, \text{rx}_{t+h}]/\mathbb{E}^*[\text{rx}_t^2]$, establishes (9). □

Proof of Proposition 1

The result follows from Theorem 3 of Shalizi (2009). □

Proof of Theorem 1

By Proposition 1, agents' long-run beliefs concentrate on the set of models with minimum KL divergence to the true model, θ^*. We thus start by characterizing $\hat{\Theta}_k = \text{argmin}_{\theta \in \Theta_k} \text{KL}(\theta^* \| \theta)$, where the KL divergence between the two models is defined in (10). As a first observation, note that instead of optimizing over Θ_k, we can optimize over $\underline{\Theta}_1 \cup \underline{\Theta}_2 \cup \cdots \cup \underline{\Theta}_k$, where $\underline{\Theta}_r$ is the set of models whose minimal realization consists of r factors. Therefore, in what follows, and without loss of generality, we assume that model $\theta = (\mathbf{A}, \mathbf{B}, c)$ is a minimal realization consisting of $r \leq k$ factors.

Next, note that under model θ, agents believe that the fundamental is described by the process in (4), where $\omega_t \in \mathbb{R}^r$ is the vector of r hidden factors. As a result, conditional on $\{x_{t-\tau}\}_{\tau=0}^{\infty}$, agents believe that ω_{t+1} is normally distributed with mean $\hat{\omega}_t = \mathbb{E}_t[\omega_{t+1}]$ and variance $\hat{\mathbf{\Sigma}}$, where $\hat{\mathbf{\Sigma}}$ is the unique positive definite matrix that satisfies the algebraic Riccati equation

$$\hat{\mathbf{\Sigma}} = \mathbf{A} \left(\hat{\mathbf{\Sigma}} - \frac{1}{c' \hat{\mathbf{\Sigma}} c} \hat{\mathbf{\Sigma}} c c' \hat{\mathbf{\Sigma}} \right) \mathbf{A}' + \mathbf{B}\mathbf{B}', \tag{A.1}$$

$\hat{\omega}_t$ is defined recursively as $\hat{\omega}_t = (\mathbf{A} - gc')\hat{\omega}_{t-1} + gx_t$, and $g \in \mathbb{R}^r$ is the Kalman gain given by

$$g = \mathbf{A}\hat{\mathbf{\Sigma}} c (c' \hat{\mathbf{\Sigma}} c)^{-1}. \tag{A.2}$$

Conditional on $\{x_{t-\tau}\}_{\tau=0}^{\infty}$, agents believe that the fundamental x_{t+1} is normally distributed with mean $\mathbb{E}_t[x_{t+1}] = c'\hat{\omega}_t$ and variance $\hat{\sigma}_x^2 = c'\hat{\mathbf{\Sigma}} c$. Furthermore, their s-step-ahead forecasts of the future realization of the fundamental is given by

$$\mathbb{E}_t[x_{t+s}] = c' \mathbf{A}^{s-1} \sum_{\tau=0}^{\infty} (\mathbf{A} - gc')^{\tau} g x_{t-\tau} \tag{A.3}$$

for all $s \geq 1$, where g is the Kalman gain in (A.2). The above expression implies that the KL divergence (10) of agents' model θ from the true data-generating process θ^* is given by

$$\begin{aligned}\mathrm{KL}(\theta^*\|\theta) = &-\frac{1}{2}\log(\hat{\sigma}_x^{-2}) + \frac{1}{2}\log(2\pi) + \frac{1}{2}\hat{\sigma}_x^{-2}\Xi_0^* - \sum_{s=1}^{\infty}\hat{\sigma}_x^{-2}\Xi_s^* c'(\mathbf{A}-gc')^{s-1}g \\ &+ \frac{1}{2}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\hat{\sigma}_x^{-2}c'(\mathbf{A}-gc')^{s-1}g\Xi_{\tau-s}^* c'(\mathbf{A}-gc')^{\tau-1}g + \mathbb{E}^*[\log f^*(x_{t+1}|x_t,\dots)],\end{aligned} \tag{A.4}$$

where $\Xi_s^* = \mathbb{E}^*[x_t x_{t+s}]$ denotes the auto-covariance of the fundamental at lag s under the true process. Now, to minimize the KL divergence between agents' model and the true data-generating process over $\underline{\Theta}_r$, we normalize the model parameters via a change of variables. In particular, since $\hat{\mathbf{\Sigma}}$ is positive definite, let

$$\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{A}\hat{\mathbf{\Sigma}}^{1/2} \qquad \text{and} \qquad u = \frac{\hat{\mathbf{\Sigma}}^{1/2}c}{\sqrt{c'\hat{\mathbf{\Sigma}}c}}. \tag{A.5}$$

Note that u is a r-dimensional vector of unit length and $\mathbf{M}$ is a $r \times r$ stable matrix. Given this change of variables, the agents' s-step ahead forecasts in (A.3) can be written as

$$\mathbb{E}_t[x_{t+s}] = u'\mathbf{M}^{s-1}\sum_{\tau=0}^{\infty}[\mathbf{M}(\mathbf{I}-uu')]^{\tau}\mathbf{M}ux_{t-\tau} \tag{A.6}$$

for all $s \geq 1$. Similarly, substituting for $\mathbf{A}$, c, and g in terms of $\mathbf{M}$ and u in (A.4) implies that the KL divergence between agents' model and the true data-generating process is given by

$$\begin{aligned}\mathrm{KL}(\theta^*\|\theta) = &\mathbb{E}^*[\log f^*(x_{t+1}|x_t,\dots)] - \frac{1}{2}\log(\hat{\sigma}_x^{-2}) + \frac{1}{2}\log(2\pi) \\ &+ \frac{1}{2}\hat{\sigma}_x^{-2}\Xi_0^* - \sum_{s=1}^{\infty}\hat{\sigma}_x^{-2}\Xi_s^* u'[\mathbf{M}(\mathbf{I}-uu')]^{s-1}\mathbf{M}u \\ &+ \frac{1}{2}\sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\hat{\sigma}_x^{-2}u'[\mathbf{M}(\mathbf{I}-uu')]^{s-1}\mathbf{M}u\Xi_{\tau-s}^* u'[\mathbf{M}(\mathbf{I}-uu')]^{\tau-1}\mathbf{M}u.\end{aligned} \tag{A.7}$$

Given the one-to-one correspondence between $(\mathbf{A},\mathbf{B},c)$ and $(\mathbf{M},u,\hat{\sigma}_x^{-2})$, minimizing the above over $(\mathbf{M},u,\hat{\sigma}_x^{-2})$ is equivalent to minimizing (A.4) over $(\mathbf{A},\mathbf{B},c)$. We thus first minimize (A.7) with respect to $\hat{\sigma}_x^{-2}$. Taking the corresponding first-order conditions and plugging back the result into (A.7) implies that minimizing the KL divergence between agents' model and the true underlying model is equivalent to minimizing

$$H(\mathbf{M},u) = 1 - 2\sum_{s=1}^{\infty}\phi_s\xi_s^* + \sum_{s=1}^{\infty}\sum_{\tau=1}^{\infty}\xi_{\tau-s}^*\phi_s\phi_\tau,$$

with respect to $\mathbf{M}$ and u, where $\xi_s^* = \Xi_s^*/\Xi_0^*$ denotes the true autocorrelation of the fundamental at lag s and $\phi_s = u'[\mathbf{M}(\mathbf{I}-uu')]^{s-1}\mathbf{M}u$. Therefore, minimizing the KL divergence between a k-factor model $\theta \in \Theta_k$ and the true underlying model θ^* is equivalent to minimizing (15) over $\mathbf{M}$, and u.

To establish (13) and (14), recall that excess returns satisfy the recursive equation (5). As a result,

$$\mathrm{rx}_{t+h} = \delta x_{t+h} + \delta u'(\mathbf{I}-\delta\mathbf{M})^{-1}\sum_{\tau=0}^{\infty}[\mathbf{M}(\mathbf{I}-uu')]^{\tau}\mathbf{M}u(\delta x_{t+h-\tau} - x_{t+h-1-\tau}),$$

where we are using the fact that agents' forecasts of future realizations of fundamentals are given by (A.6). Rearranging terms, we obtain

$$\mathrm{rx}_{t+h} = \delta u'(\mathbf{I} - \delta \mathbf{M})^{-1} u \left(x_{t+h} - \sum_{\tau=1}^{\infty} \phi_\tau x_{t+h-\tau} \right). \tag{A.8}$$

The above equation, coupled with the fact that $\beta_h^{\text{Fama}} = \mathbb{E}^*[x_t \, \mathrm{rx}_{t+h}]/\mathbb{E}^*[x_t^2]$, implies that the slope coefficient of the Fama regression (6) satisfies (13). Similarly, noting that $\beta_h^{\text{mom}} = \mathbb{E}^*[\mathrm{rx}_t \, \mathrm{rx}_{t+h}]/\mathbb{E}^*[\mathrm{rx}_t^2]$ and using (A.8) establishes that the slope coefficient of the momentum regression is given by (14). □

Proof of Proposition 2

Proof of part (a) When $k \geq n$, the true model is within the set of models considered by the agents, i.e., $\theta^* \in \Theta_n \subseteq \Theta_k$. As a result, agents' subjective expectations coincide with rational expectations. Consequently, equation (5) implies that

$$\mathbb{E}_t^*[\mathrm{rx}_{t+h}] = \sum_{\tau=1}^{\infty} \delta^\tau \left(\mathbb{E}_t^* \mathbb{E}_{t+1}^*[x_{t+\tau}] - \mathbb{E}_t^*[x_{t+\tau}] \right).$$

Hence, by the law of iterated expectations, $\mathbb{E}_t^*[\mathrm{rx}_{t+h}] = 0$, which guarantees that $\beta_h^{\text{Fama}} = \beta_h^{\text{mom}} = 0$ for all $h \geq 1$. □

Proof of part (b) We first show that if $\beta_h^{\text{Fama}} = 0$ for all h, then $k \geq n$. Suppose to the contrary that $k < n$ and define ϕ_s and ϕ_s^* to denote the output from the minimization in (15) in Theorem 1 for the k- and n-factor models, respectively. By part (a) of the theorem, the slope coefficients of the return predictability regression arising from the optimal n-factor model are equal to zero at all horizons. Therefore, by equation (13),

$$\xi_h^* = \sum_{\tau=1}^{\infty} \phi_\tau^* \xi_{h-\tau}^*. \tag{A.9}$$

for all h. Furthermore, by assumption, $\beta_h^{\text{Fama}} = 0$ for all h under the k-factor model. Therefore, (13) implies that

$$\xi_h^* = \sum_{\tau=1}^{\infty} \phi_\tau \xi_{h-\tau}^* \tag{A.10}$$

for all h. Multiplying both sides of the first equation by ϕ_h and the second by ϕ_h^* and summing over all h, we get

$$\sum_{h=1}^{\infty} \phi_h^* \xi_h^* = \sum_{h=1}^{\infty} \phi_h \xi_h^*. \tag{A.11}$$

Next, note equations (A.9) and (A.10) also imply that the objective function (15) evaluated at the optimal solution within the set of all n and k-factor models is, respectively, equal to

$$H^* = 1 - \sum_{s=1}^{\infty} \phi_s^* \xi_s^*$$
$$H = 1 - \sum_{s=1}^{\infty} \phi_s \xi_s^*.$$

Comparing the above two equations with (A.11) implies that $H = H^*$. Consequently, the k-factor model results in the same KL divergence to the data generating process as the n-factor model, whose KL divergence to the data-generating process is equal to zero by assumption. This means that the data-generating process has a representation with $k < n$ factors, which contradicts the assumption that n is the number of factors in the minimal representation of the data-generating process.

As the final step of the proof, we show that if the slope coefficient of the momentum regression (7) is zero at all horizons, then $k \geq n$. It is sufficient to show that $\beta_h^{\text{mom}} = 0$ for all h implies that $\beta_h^{\text{Fama}} = 0$ for all h, as we can then use the result for the slope coefficients of the Fama regression proved earlier to conclude that $k \geq n$. To this end, note that if $\beta_h^{\text{mom}} = 0$ for all h, then equation (14) in Theorem 1 implies that

$$\xi_h^* - \sum_{s=1}^{\infty} \phi_s (\xi_{h-s}^* + \xi_{h+s}^*) + \sum_{s,\tau=1}^{\infty} \phi_s \phi_\tau \xi_{h+s-\tau}^* = 0$$

for all $h \geq 1$. Multiplying both sides of the above equation by $\delta u'(\mathbf{I} - \delta \mathbf{M})^{-1} u$ and using (13) leads to

$$\beta_h^{\text{Fama}} - \sum_{s=1}^{\infty} \phi_s \beta_{h+s}^{\text{Fama}} = 0. \tag{A.12}$$

Define the sequence $(\kappa_1, \kappa_2, \dots)$ recursively as $\kappa_h = \phi_h + \sum_{\tau=1}^{h-1} \phi_{h-\tau} \kappa_\tau$. By (A.12),

$$\sum_{h=1}^{\infty} \kappa_h \beta_{h+r}^{\text{Fama}} - \sum_{h=1}^{\infty} \sum_{s=1}^{\infty} \kappa_h \phi_s \beta_{h+r+s}^{\text{Fama}} = 0$$

for all $r \geq 1$. Using the recursive definition of κ_h, we obtain

$$\sum_{h=1}^{\infty} \phi_h \beta_{h+r}^{\text{Fama}} + \sum_{h=1}^{\infty} \sum_{\tau=1}^{h-1} \phi_{h-\tau} \kappa_\tau \beta_{h+r}^{\text{Fama}} - \sum_{h=1}^{\infty} \sum_{s=1}^{\infty} \kappa_h \phi_s \beta_{h+r+s}^{\text{Fama}} = 0.$$

It is straightforward to verify that the second and the third terms on the left-hand side above add up to zero, whereas equation (A.12) implies that the first term is equal to β_r^{Fama}. This therefore establishes that $\beta_r^{\text{Fama}} = 0$ for all $r \geq 1$, which completes the proof. □

Proof of Proposition 3

As a first observation, note that since the underlying process that generates the fundamental is stationary, its autocorrelation function $\xi_s^* = \mathbb{E}^*[x_t x_{t+s}]/\mathbb{E}^*[x_t^2]$ decays at an exponential rate as

$s \to \infty$. Next, we show that $\phi_s = u'[\mathbf{M}(\mathbf{I} - uu')]^{s-1}\mathbf{M}u$ also decays at an exponential rate, where $\mathbf{M}$ and u are given by (A.5). To this end, first note that, $\mathbf{M}(\mathbf{I} - uu') = \hat{\mathbf{\Sigma}}^{-1/2}(\mathbf{A} - gc')\hat{\mathbf{\Sigma}}^{1/2}$. Therefore, it is sufficient to show that all eigenvalues of $\mathbf{A} - gc'$ are inside the unit circle. Rewriting the algebraic Riccati equation in (A.1), we obtain

$$(\mathbf{A} - gc')\hat{\mathbf{\Sigma}}(\mathbf{A} - gc')' - \hat{\mathbf{\Sigma}} + \mathbf{B}\mathbf{B}' = 0,$$

which is a discrete Lyapunov equation in $\mathbf{A} - gc'$. Since $(\mathbf{A}, \mathbf{B}, c)$ is the minimal representation of the state-space model, Kalman's decomposition theorem implies that (i) $\hat{\mathbf{\Sigma}}$ is positive definite and (ii) the pair $(\mathbf{B}, \mathbf{A})$ is controllable. Therefore, by Lyapunov's theorem, all eigenvalues of $\mathbf{A} - gc'$ are inside the unit circle, thus guaranteeing that ϕ_s decays at an exponential rate as $s \to \infty$.

With the above in hand, we next show that $\lim_{h\to\infty} \beta_h^{\text{Fama}} = 0$. Recall from Theorem 1 that the slope coefficient of Fama regression satisfies (13). Therefore, by triangle inequality,

$$|\beta_h^{\text{Fama}}| \leq \delta |u'(\mathbf{I} - \delta\mathbf{M})^{-1}u| \left(|\xi_h^*| + \sum_{\tau=1}^{h} |\phi_\tau||\xi_{h-\tau}^*| + \sum_{\tau=1}^{\infty} |\phi_{\tau+h}||\xi_\tau^*| \right).$$

Since ξ_s^* and ϕ_s converges to zero at exponential rates, there are constants $c_1, c_2 > 0$ and $\rho_1, \rho_2 < 1$ such that $|\xi_s^*| \leq c_1\rho_1^s$ and $|\phi_s| \leq c_2\rho_2^s$ for all s. Consequently,

$$|\beta_h^{\text{Fama}}| \leq \delta |u'(\mathbf{I} - \delta\mathbf{M})^{-1}u| \left(c_1\rho_1^h + c_1c_2\rho_1^h \sum_{\tau=1}^{h} (\rho_2/\rho_1)^\tau + c_1c_2\rho_2^h \sum_{\tau=1}^{\infty} (\rho_1\rho_2)^\tau \right),$$

and as a result, $|\beta_h^{\text{Fama}}| \leq c_3 h \rho_3^h$ for all h for some constant $c_3 > 0$ and $\rho_3 = \max\{\rho_1, \rho_2\}$. This inequality then guarantees that $\lim_{h\to\infty} \beta_h^{\text{Fama}} = 0$.

To establish that the slope coefficients of the momentum regression also converge to zero in long horizons, note that the characterization result in equations (13) and (14) implies that

$$\beta_h^{\text{mom}} = \frac{\beta_h^{\text{Fama}} - \sum_{s=1}^{\infty} \phi_s \beta_{h+s}^{\text{Fama}}}{\beta_0^{\text{Fama}} - \sum_{s=1}^{\infty} \phi_s \beta_s^{\text{Fama}}},$$

with the convention that $\beta_0^{\text{Fama}} = 1 - \sum_{s=1}^{\infty} \phi_s \xi_s^*$. Therefore, $|\beta_h^{\text{Fama}}| \leq c_3 h \rho_3^h$ implies that

$$|\beta_h^{\text{mom}}| \leq \frac{\rho_3^h}{\beta_0^{\text{Fama}} - \sum_{s=1}^{\infty} \phi_s \beta_s^{\text{Fama}}} \left(c_3 h + c_1 c_3 \sum_{s=1}^{\infty} (h+s)(\rho_1\rho_3)^s \right).$$

As a result, $\lim_{h\to\infty} \beta_h^{\text{mom}} = 0$. □

Proof of Proposition 4

Recall from Theorem 1 that the slope coefficients of the Fama and momentum regressions are given by (13) and (14), where the sequence $(\phi_1, \phi_2, \dots)$ is obtained by minimizing (15) over the $k \times k$ stable matrix $\mathbf{M}$ and the unit vector $u \in \mathbb{R}^k$. Therefore, when $k = 1$, it must be the case that $\mathbf{M}$ is a scalar, denoted by m, satisfying $|m| < 1$ and $u \in \{-1, 1\}$. This immediately implies that $\phi_1 = m$ and $\phi_s = 0$

for all $s \geq 2$. As a result, the expressions for the Fama and momentum regressions reduce to

$$\beta_h^{\text{Fama}} = \frac{\delta}{1-\delta m}\left(\xi_h^* - m\xi_{h-1}^*\right) \tag{A.13}$$

$$\beta_h^{\text{mom}} = \frac{\xi_h^* - m(\xi_{h-1}^* + \xi_{h+1}^*) + m^2\xi_h^*}{1 - 2m\xi_1^* + m^2}, \tag{A.14}$$

respectively, while the objective function in (15) is given by

$$H(m,u) = 1 - 2m\xi_1^* + m^2. \tag{A.15}$$

Optimizing the above over $m \in (-1,1)$ implies that $m = \xi_1^*$. Plugging in the results into (A.13) and (A.14) then completes the proof. □

Proof of Proposition 5

In the proof of Proposition 4 we established that when $k = 1$, the optimal solution of the optimization in (15) is given by $m = \xi_1^*$. Therefore, equation (A.5) implies that the persistence parameter of the agents' model is equal to $a = m = \xi_1^*$. Consequently, the autocorrelation function implied by the agents' model is given by $\xi_h = \xi_1^{*h}$ for all horizons h. This observation together with the characterization in (16) implies that $\beta_h^{\text{Fama}} > 0$ if and only if $\xi_h^* > \xi_{h-1}^*\xi_h/\xi_{h-1}$. Taking logarithms from both sides of this inequality establishes the result. □

Proof of Proposition 6

Let $p(t) = t^n - \sum_{i=1}^n \gamma_i t^{n-i}$ denote the characteristic polynomial of $\mathbf{A}^*$, which is the n-dimensional square matrix that governs the transitional dynamics of the true data-generating process. By the Cayley-Hamilton theorem (Horn and Johnson, 2013, p. 109), $\mathbf{A}^*$ satisfies its characteristic polynomial, i.e, $\mathbf{A}^{*n} = \sum_{i=1}^n \gamma_i \mathbf{A}^{*n-i}$. Multiplying both sides of this equation by $\mathbf{A}^{*h}$ implies that

$$\mathbf{A}^{*h} = \gamma_1 \mathbf{A}^{*h-1} + \cdots + \gamma_n \mathbf{A}^{*h-n} \tag{A.16}$$

for all $h \geq n$. It is easy to verify that the auto-covariance of the true data-generating process at lag h is given by $\Xi_h^* = c^{*\prime}\mathbf{A}^{*h}\mathbf{Q}c^*$, where $\mathbf{Q} = \sum_{j=0}^\infty \mathbf{A}^{*j}\mathbf{B}^*(\mathbf{A}^{*j}\mathbf{B}^*)'$. Multiplying both sides of (A.16) by $c^{*\prime}$ from the left and $\mathbf{Q}c^*$ from the right implies that

$$\Xi_h^* = \gamma_1 \Xi_{h-1}^* + \cdots + \gamma_n \Xi_{h-n}^* \tag{A.17}$$

for all $h \geq n$. On the other hand, recall from (16) that when $k = 1$, the slope coefficient of the Fama regression at horizon h is given by $\beta_h^{\text{Fama}} = d(\Xi_h^*\Xi_0^* - \Xi_1^*\Xi_{h-1}^*)$, where $d \neq 0$ is some constant. This observation, combined with (A.17) then implies that

$$\beta_h^{\text{Fama}} = \gamma_1 \beta_{h-1}^{\text{Fama}} + \cdots + \gamma_n \beta_{h-n}^{\text{Fama}} \tag{A.18}$$

for all $h \geq n+1$. In other words, the slope coefficients of the Fama regression satisfy a recursive equation with coefficients given by the coefficients of the characteristic polynomial of matrix $\mathbf{A}^*$.

Since the roots of the characteristic polynomial of $\mathbf{A}^*$ is equal to its eigenvalues, the solution to the recursive equation in (A.18) when all eigenvalues of $\mathbf{A}^*$ is given by

$$\beta_h^{\text{Fama}} = \sum_{i=1}^{n} \zeta_i \lambda_i^{h-1}$$

for all h, where $(\lambda_1, \ldots, \lambda_n)$ denote the eigenvalues of $\mathbf{A}^*$ and $(\zeta_1, \ldots, \zeta_n)$ are a collection of constants that do not depend on horizon h. Furthermore, note that, by (16), $\beta_1^{\text{Fama}} = 0$, which implies that $\sum_{i=1}^{n} \zeta_i = 0$. Therefore, in the special case that $n = 2$, the slope coefficients of the Fama regression are given by $\beta_h^{\text{Fama}} = \zeta_1(\lambda_1^{h-1} - \lambda_2^{h-1})$ for all $h \geq 1$, where ζ_1 is some constant and λ_1 and λ_2 are the eigenvalues of $\mathbf{A}^*$. Consequently,

$$\beta_h^{\text{Fama}} = \beta_2^{\text{Fama}} \left(\frac{\lambda_1^{h-1} - \lambda_2^{h-1}}{\lambda_1 - \lambda_2} \right). \tag{A.19}$$

With the above in hand, we can now prove the various statements of the proposition. In what follows, we assume that λ_1 is the eigenvalue with the largest modulus, i.e., $|\lambda_1| \geq |\lambda_2|$.

Proof of part (a) Suppose both λ_1 and λ_2 are real and that $\lambda_1 > 0$. It is immediate that the right-hand side of (A.19) never changes sign. Therefore, β_h^{Fama} has the same sign as β_2^{Fama} for all h. □

Proof of part (b) Suppose both λ_1 and λ_2 are real and that $\lambda_1 < 0$. In this case, the right-hand side of (A.19) has the same sign as β_2^{Fama} when h is even, whereas it has the opposite sign of β_2^{Fama} when h is odd. This means that β_h^{Fama} changes sign at every horizon. □

Proof of part (c) Suppose both λ_1 and λ_2 are complex. Since they are eigenvalues of real matrix $\mathbf{A}^*$, they form a complex conjugate pair. We can therefore represent them in polar coordinates as $\lambda_1 = re^{i\varphi}$ and $\lambda_2 = re^{-i\varphi}$, where without loss of generality we assume that $0 < \varphi < \pi$. Plugging these values into the right-hand side of (A.19) implies that

$$\beta_{h+1}^{\text{Fama}} = \beta_2^{\text{Fama}} r^{h-1} \left(\frac{e^{ih\varphi} - e^{-ih\varphi}}{e^{i\varphi} - e^{-i\varphi}} \right) = \beta_2^{\text{Fama}} r^{h-1} \frac{\sin(h\varphi)}{\sin \varphi}$$

for all $h \geq 1$. Since $0 < \varphi < \pi$, the denominator of the ratio on the right-hand side of the above equation is always positive. Therefore, $\beta_{h+1}^{\text{Fama}}$ has the same sign as β_2^{Fama} if $\sin(h\varphi) > 0$ and has the opposite sign if β_2^{Fama} if $\sin(h\varphi) < 0$. Therefore, $\beta_{h+1}^{\text{Fama}}$ changes sign at frequency π/φ. □

Proof of Proposition 7

Recall from Proposition 4 that when agents are restricted to single-factor models, the slope coefficient of the momentum regression at horizon h is given by (17). As a result,

$$\sum_{h=1}^{\infty} (\xi_1^*)^h \beta_h^{\text{mom}} = \frac{1}{1 - \xi_1^{*2}} \sum_{h=1}^{\infty} \left(\xi_1^{*h} (\xi_h^* - \xi_1^* \xi_{h-1}^*) - \xi_1^{*h+1} (\xi_{h+1}^* - \xi_1^* \xi_h^*) \right).$$

Since $|\xi_1^*| < 1$, we have

$$\sum_{h=1}^{\infty} \xi_1^{*h} \beta_h^{\text{mom}} = \frac{\xi_1^*(\xi_1^* - \xi_1^*\xi_0^*)}{1 - \xi_1^{*2}}.$$

The fact that $\xi_0^* = 1$ then guarantees that the right-hand side of the above equation is equal to zero. Therefore, $\sum_{h=1}^{\infty} \xi_1^{*h} \beta_h^{\text{mom}} = 0$. It is then immediate that when $\xi_1^* > 0$, then either (i) $\beta_h^{\text{mom}} = 0$ for all h or (ii) there exist h and h' such that $\beta_h^{\text{mom}} > 0 > \beta_{h'}^{\text{mom}}$. But the first possibility is ruled out by Proposition 2(b), which guarantees that there exists at least one h such that $\beta_h^{\text{mom}} \neq 0$ whenever $n > k$. Therefore, there are at least two horizons h and h' such that $\beta_h^{\text{mom}} > 0 > \beta_{h'}^{\text{mom}}$. □

Proof of Proposition 8

Recall from Subsection 3.5 that excess returns in the heterogenous-agent economy are given by (20). Since rational agents can fully construct the model used by behavioral agents, the regress of expectations in (20) is given by

$$\bar{\mathbb{E}}_t \bar{\mathbb{E}}_{t+1} \ldots \bar{\mathbb{E}}_{t+\tau}[x_{t+\tau}] = \gamma^{\tau+1} \mathbb{E}_t^*[x_{t+\tau}] + (1-\gamma) \sum_{s=0}^{\tau} \gamma^s \mathbb{E}_t^* \mathbb{E}_{t+s}[x_{t+\tau}]$$

for all $\tau \geq 0$, where $\mathbb{E}[\cdot]$ and $\mathbb{E}^*[\cdot]$ are the subjective expectations of the behavioral and rational agents, respectively, and γ denotes the fraction of rational agents. Consequently,

$$\begin{aligned} \text{rx}_t(\gamma) = (1-\gamma) \sum_{s=0}^{\infty} (\delta\gamma)^s \sum_{\tau=1}^{\infty} \delta^\tau \Big(\mathbb{E}_t^* \mathbb{E}_{t+s}[x_{t+\tau+s-1}] - \mathbb{E}_{t-1}^* \mathbb{E}_{t+s-1}[x_{t+\tau+s-1}] \Big) \\ + \sum_{\tau=1}^{\infty} (\delta\gamma)^\tau \Big(\mathbb{E}_t^*[x_{t+\tau-1}] - \mathbb{E}_{t-1}^*[x_{t+\tau-1}] \Big). \end{aligned}$$

Taking expectations from both sides of the above equation and using the expression in (5) for excess returns in the representative-agent economy therefore implies that

$$\mathbb{E}_{t-1}^*[\text{rx}_t(\gamma)] = (1-\gamma) \sum_{s=0}^{\infty} (\delta\gamma)^s \mathbb{E}_{t-1}^*[\text{rx}_{t+s}(0)].$$

Therefore, expected excess return in the heterogenous-agent economy is the discounted sum of all future excess returns of a representative-agent economy populated by the behavioral agents only. Multiplying both sides of the above equation by x_{t-h} and taking expectations $\mathbb{E}^*[\cdot]$ then establishes the result. □

Derivations for Example 1

By Proposition 1, agents' long-run beliefs concentrate on the set of models with the minimum KL convergence to the true data-generating process. Furthermore, in the proof of Theorem 1, we established that minimizing the KL divergence between the k-dimensional model $\theta_k = (\mathbf{A}, \mathbf{B}, c)$ and the true model θ^* is equivalent to minimizing (A.7) over $(\mathbf{M}, u, \hat{\sigma}_x^{-2})$, where $\mathbf{M}$ and u are given by (A.5) and $\hat{\sigma}_x^2 = c'\hat{\mathbf{\Sigma}}c$, in which $\hat{\mathbf{\Sigma}}$ is the solution to the algebraic Riccati equation (A.1).

We now set $k = 1$ and determine the parameters of agents' model $\theta = (a, b, c)$. When $k = 1$, the objective function in (A.7) reduces to

$$\mathrm{KL}(\theta^* \| \theta) = \mathbb{E}^*[\log f^*(x_{t+1}|x_t, \dots)] - \frac{1}{2}\log(\hat{\sigma}_x^{-2}) + \frac{1}{2}\log(2\pi) + \frac{1}{2}\hat{\sigma}_x^{-2}(1+m^2)\Xi_0^* - \hat{\sigma}_x^{-2} m \Xi_1^*,$$

where m is a scalar such that $|m| < 1$. Minimizing the above over m and $\hat{\sigma}_x^{-2}$ implies that

$$m = \xi_1^* \qquad \text{and} \qquad \hat{\sigma}_x^2 = \Xi_0^*(1 - \xi_1^{*2}),$$

where ξ_h^* denotes the autocorrelation of the fundamental at lag h. The first equation above, coupled with (A.5), implies that $a = \xi_1^*$, whereas the second equation together with the fact that $\hat{\sigma}_x^2 = c'\hat{\mathbf{\Sigma}}c$ implies that $b^2c^2 = \Xi_0^*(1 - \xi_1^{*2})$, where we are using the fact that when $k = 1$ the solution to the algebraic Riccati equation (A.1) is equal to b^2. Replacing for ξ_1^* and Ξ_0^* for the model in (11) establishes (12). It additionally implies that constants b and c in the agents' model satisfy

$$\frac{(bc)^2}{1-a^2} = \sum_{i=1}^{n} \frac{(b_i^* c_i^*)^2}{1 - a_i^{*2}}.$$

In juxtaposition with (12), the above equation implies that when the true underlying model is one-dimensional (i.e., $n = 1$), the agents' beliefs concentrate over the set of single-factor models $\theta = (a, b, c)$, where $a = a^*$ and $|bc| = |b^*c^*|$. Note that even though the agents cannot learn the exact values of b^* and c^*, all models with the same $|b^*c^*|$ are observationally equivalent. As a result, when $k = n = 1$, the agents learn the true data-generating process, consistent with the prediction of Proposition 1. □

B Technical Appendix

B.1 Model-Implied Autocorrelation Function

In this appendix, we derive the expression for the model-implied autocorrelation function when agents are restricted to models consisting of at most k factors. Let $\theta = (\mathbf{A}, \mathbf{B}, c)$ denote the collection of parameters that represent agents' model in (4). To compute the perceived autocorrelation function as a function of the parameters of agents' models, first note that, for $s \geq 0$,

$$\Xi_s = \mathbb{E}[x_t x_{t-s}] = c'\mathbb{E}[\omega_t \omega_{t-s}']c = c'\mathbf{A}^s\,\mathbb{E}[\omega_t\omega_t']c.$$

Using the change of variables (A.5) from the proof of Theorem 1 implies that

$$\Xi_s = \hat{\sigma}_x^2 u'\mathbf{M}^s\hat{\mathbf{\Sigma}}^{-1/2}\mathbb{E}[\omega_t\omega_t']\hat{\mathbf{\Sigma}}^{-1/2}u.$$

Therefore, to represent Ξ_s in terms of $\mathbf{M}$ and u, we need to find $\mathbf{Q} = \hat{\mathbf{\Sigma}}^{-1/2}\mathbb{E}[\omega_t\omega_t']\hat{\mathbf{\Sigma}}^{-1/2}$. Equation (4) implies that $\mathbb{E}[\omega_t\omega_t'] = \mathbf{A}\mathbb{E}[\omega_t\omega_t']\mathbf{A}' + \mathbf{B}\mathbf{B}'$. Multiplying both sides of this equation from left and right by $\hat{\mathbf{\Sigma}}^{-1/2}$ implies that $\mathbf{Q} = \mathbf{M}\mathbf{Q}\mathbf{M}' + \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{B}\mathbf{B}'\hat{\mathbf{\Sigma}}^{-1/2}$. On the other hand, the algebraic Riccati equation in (A.1) can be written in terms of $\mathbf{M}$ and u as $\mathbf{M}(\mathbf{I} - uu')\mathbf{M}' + \hat{\mathbf{\Sigma}}^{-1/2}\mathbf{B}\mathbf{B}'\hat{\mathbf{\Sigma}}^{-1/2} = \mathbf{I}$. Combining the last two equations implies that $\mathbf{Q}$ is the solution to the discrete Lyapunov equation:

$$\mathbf{Q} = \mathbf{M}\mathbf{Q}\mathbf{M}' + \mathbf{I} - \mathbf{M}(\mathbf{I} - uu')\mathbf{M}'. \tag{B.1}$$

Therefore, the model-implied autocorrelation at lag $s \geq 0$ is given by

$$\xi_s = \Xi_s / \Xi_0 = \frac{u' \mathbf{M}^s \mathbf{Q} u}{u' \mathbf{Q} u},$$

where $\mathbf{M}$ and u minimize the expression in (15) and $\mathbf{Q}$ is the solution to equation (B.1).

B.2 Over- and Under-Reaction to Information

Recall from Proposition 2 that when the set of models entertained by agents is not rich enough to contain the true data-generating process (i.e., when $k < n$), they end up with subjective expectations that generate return predictability. This departure from rational expectations is a consequence of the fact that, due to their misspecified model of the data-generating process, agents do not incorporate new pieces of information into their forecasts as they would have had they known the true process.

In this appendix, we show that whether agents over- or under-react to new information cannot be decoupled from the environment they live in. In particular, we argue that agents' forecasts may exhibit systematic over- or under-reaction to news depending on (i) the statistical properties of the underlying data-generating process and (ii) the horizon of interest. This means that, in our framework, over- and under-reaction of expectations are endogenous and are not baked into agents' expectations formation process.

To measure the extent of over- and under-reaction to new information at different horizons, we follow Coibion and Gorodnichenko (2015) and consider the family of regressions

$$x_{t+h} - \mathbb{E}_t[x_{t+h}] = \alpha_h^{\text{CG}} + \beta_h^{\text{CG}} (\mathbb{E}_t[x_{t+h}] - \mathbb{E}_{t-1}[x_{t+h}]), \tag{B.2}$$

where, as before, x_t is the realization of the fundamental at time t and $\mathbb{E}[\cdot]$ denotes agents' subjective expectations. The regressand on left-hand side of (B.2) is agents' forecast error for realization of the fundamental h periods in the future, while the regressor on the right-hand side is their latest forecast revision. Thus, $\beta_h^{\text{CG}} > 0$ means that when agents revise their h-step-ahead forecast of the fundamental upward at time t, they tend to undershoot its eventual realization at time $t + h$. In other words, agents systematically under-react to new information at time t. A similar logic implies that $\beta_h^{\text{CG}} < 0$ means that agents tend to systematically over-react to new information when forming expectations about the realization of the fundamental h periods in the future. As is well known, under rational expectations (i.e., when $\mathbb{E}[\cdot] = \mathbb{E}^*[\cdot]$), forecast revisions are orthogonal to future forecast errors, which means that $\beta_h^{\text{CG}} = 0$ for all horizons $h \geq 1$.

Turning to our framework, suppose that agents are constrained to using single-factor models (i.e., $k = 1$) to make forecasts about the future realization of the fundamental, even though the true data-generating process is driven by $n > 1$ factors. As we show below, under such a specification, the slope coefficient of regression (B.2) at horizon h is given by

$$\beta_h^{\text{CG}} = \frac{\xi_h^* - \xi_1^* \xi_{h+1}^*}{\xi_1^{*h} (1 - \xi_1^{*2})} - 1, \tag{B.3}$$

where ξ_h^* denotes the autocorrelation of the fundamental at lag h.

Two observations are immediate. First, whether agents over- or under-react to new information depends on the statistical properties of the underlying data-generating process. For example, it is straightforward to verify that $\beta_1^{\text{CG}} > 0$ if $\xi_1^{*2} > \xi_2^*$, whereas $\beta_1^{\text{CG}} < 0$ if $\xi_1^{*2} < \xi_2^*$. Thus, the same agent with the same level of sophistication may end up with forecasts that under- or overshoot the eventual realization of the fundamental depending on the environment. Second, holding the underlying data-generating process fixed, the sign and the magnitude of the right-hand side of (B.3) varies with h. This means that agents' expectations may over-react to new information at some horizon, while simultaneously under-reacting at others. As already discussed, the exact pattern of over- and under-reaction depends on the process that governs the fundamental.

Derivation of Equation (B.3) Since $k = 1$, the agents' model can be represented by the tuple of scalars $\theta = (a, b, c)$, where $a \in (-1, 1)$ denotes the persistence parameter of the underlying factor. This means that $\mathbb{E}_t[x_{t+h}] = a^h x_t$ for all $h \geq 0$. As a result, the slope coefficient of the regression in (B.2) is given by

$$\beta_h^{\text{CG}} = \frac{\mathbb{E}^*[(x_{t+h} - a^h x_t)(x_t - a x_{t-1})]}{a^h \mathbb{E}^*[(x_t - a x_{t-1})^2]} = \frac{\xi_h^* - a\xi_{h+1}^* - a^h(1 - a\xi_1^*)}{a^h(1 + a^2 - 2a\xi_1^*)}, \tag{B.4}$$

where ξ_h^* denotes the autocorrelation of the fundamental at lag h. Next, recall from the proof of Proposition 4 that when agents are restricted to single-factor models, the objective function in (15) reduces to (A.15), in which $|m| < 1$ is a scalar. Optimizing (A.15) over m implies that $m = \xi_1^*$. Therefore, by equation (A.5), the persistence parameter is given by $a = \xi_1^*$. Plugging this expression into (B.4) establishes (B.3).

References

Allen, Franklin, Stephen Morris, and Hyun Song Shin (2006), "Beauty contests and iterated expectations in asset markets." *Review of Financial Studies*, 19(3), 719–752.

Ang, Andrew and Geert Bekaert (2006), "Stock return predictability: Is it there?" *Review of Financial Studies*, 20(3), 651–707.

Angeletos, George-Marios and Zhen Huo (2020), "Myopia and anchoring." *American Economic Review* (forthcoming).

Angeletos, George-Marios and Chen Lian (2018), "Forward guidance without common knowledge." *American Economic Review*, 108(9), 2477–2512.

Angeletos, George-Marios and Karthik A. Sastry (2020), "Managing expectations: Instruments vs. targets." *Quarterly Journal of Economics* (forthcoming).

Antoulas, Athanasios C. (2005), *Approximation of Large-Scale Dynamical Systems*. Society for Industrial and Applied Mathematics.

Aoki, Masanao (1983), "Equivalence of ARMA and state space models." In *Notes on Economic Time Series Analysis: System Theoretic Perspectives*, 22–32, Springer-Verlag, Berlin.

Bacchetta, Philippe and Eric van Wincoop (2010), "Infrequent portfolio decisions: A solution to the forward discount puzzle." *American Economic Review*, 100(3), 870–904.

Bacchetta, Philippe and Eric van Wincoop (2020), "Puzzling exchange rate dynamics and delayed portfolio adjustment." NBER Working Paper, No. 26259.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer (2015), "X-CAPM: An extrapolative capital asset pricing model." *Journal of Financial Economics*, 115(1), 1–24.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998), "A model of investors sentiment." *Journal of Financial Economics*, 49(3), 307–343.

Barillas, Francisco and Kristoffer P. Nimark (2017), "Speculation and the term structure of interest rates." *Review of Financial Studies*, 30(11), 4003–4037.

Berk, Robert H. (1966), "Limiting behavior of posterior distributions when the model is incorrect." *The Annals of Mathematical Statistics*, 37, 51–58.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2019), "Diagnostic expectations and stock returns." *Journal of Finance*, 74(6), 2839–2874.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018), "Diagnostic expectations and credit cycles." *Journal of Finance*, 73(1), 199–227.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2020), "Memory, attention, and choice." *Quarterly Journal of Economics*, 135(3), 1399–1442.

Campbell, John Y. (1991), "A variance decomposition for stock returns." *Economic Journal*, 101(405), 157–179.

Chinn, Menzie D. and Jeffrey A. Frankel (2020), "A third of a century of currency expectations data: The carry trade and the risk premium." Working Paper, Harvard University.

Coibion, Olivier and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review*, 105(8), 2644–2678.

Cutler, David M., James M. Poterba, and Lawrence H. Summers (1991), "Speculative dynamics." *Review of Economic Studies*, 58(3), 529–546.

da Silveira, Rava Azeredo, Yeji Sung, and Michael Woodford (2020), "Optimally imprecise memory and biased forecasts." NBER Working paper No. 28075.

Engel, Charles (2016), "Exchange rates, interest rates, and the risk premium." *American Economic Review*, 106(2), 436–474.

Esponda, Ignacio and Demian Pouzo (2016), "Berk–Nash equilibrium: A framework for modeling agents with misspecified models." *Econometrica*, 84(3), 1093–1130.

Esponda, Ignacio and Demian Pouzo (2020), "Equilibrium in misspecified Markov decision processes." *Theoretical Economics* (forthcoming).

Fama, Eugene F. (1984), "Forward and spot exchange rates." *Journal of Monetary Economics*, 14(3), 319–338.

Farhi, Emmanuel and Iván Werning (2019), "Monetary policy, bounded rationality, and incomplete markets." *American Economic Review*, 109(11), 3887–3928.

Froot, Kenneth A. and Jeffrey A. Frankel (1989), "Forward discount bias: Is it an exchange risk premium?" *Quarterly Journal of Economics*, 104(1), 139–161.

Fuster, Andreas, Benjamin Hébert, and David Laibson (2012), "Natural expectations, macroeconomic dynamics, and asset pricing." *NBER Macroeconomics Annual*, 26, 1–48.

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.

Gabaix, Xavier (2017), "Behavioral macroeconomics via sparse dynamic programming." NBER Working Paper No. 21848.

Golez, Benjamin (2014), "Expected returns and dividend growth rates implied by derivative markets." *Review of Financial Studies*, 27(3), 790–822.

Gourinchas, Pierre-Olivier and Aaron Tornell (2004), "Exchange rate puzzles and distorted beliefs." *Journal of International Economics*, 64(2), 303–333.

Guo, Hongye and Jessica A. Wachter (2019), "Superstitious investors." Working Paper, University of Pennsylvania.

Hirshleifer, David, Jun Li, and Jianfeng Yu (2015), "Asset pricing in production economies with extrapolative expectations." *Journal of Monetary Economics*, 76, 87–106.

Horn, Roger A. and Charles R. Johnson (2013), *Matrix Analysis*, second edition. Cambridge University Press, New York, NY.

Kailath, Thomas (1980), *Linear Systems*. Prentice-Hall, Englewood Cliffs, NJ.

Martin, Ian and Stefan Nagel (2020), "Market efficiency in the age of big data." NBER Working Paper No. 26586.

Molavi, Pooya (2019), "Macroeconomics with learning and misspecification: A general theory and applications." Working Paper, Northwestern University.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen (2012), "Time series momentum." *Journal of Financial Economics*, 104(2), 228–250.

Nagel, Stefan and Zhengyang Xu (2019), "Asset pricing with fading memory." NBER Working Paper No. 26255.

Newey, Whitney K. and Kenneth D. West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, 55(3), 703–708.

Panageas, Stavros (2020), "The implications of heterogeneity and inequality for asset pricing." In *Foundations and Trends in Finance*, volume 12, 199–275.

Rabin, Matthew and Dimitri Vayanos (2010), "The gambler's and hot-hand fallacies: Theory and applications." *Review of Economic Studies*, 77(2), 730–778.

Sandberg, Henrik (2019), "Introduction to model order reduction." Lecture notes, http://www.diva-portal.org/smash/get/diva2:1338145/fulltext01.pdf.

Shalizi, Cosma Rohilla (2009), "Dynamics of Bayesian updating with dependent data and misspecified models." *Electronic Journal of Statistics*, 3, 1039–1074.

Valchev, Rosen (2020), "Bond convenience yields and exchange rate dynamics." *American Economic Journal: Macroeconomics*, 12(2), 124–166.

Van Nieuwerburgh, Stijn and Laura Veldkamp (2009), "Information immobility and the home bias puzzle." *Journal of Finance*, 64(3), 1187–1215.

Van Nieuwerburgh, Stijn and Laura Veldkamp (2010), "Information acquisition and under-diversification." *Review of Economic Studies*, 77(2), 779–805.

Wachter, Jessica A. and Michael Jacob Kahana (2021), "A retrieved-context theory of financial decisions." NBER Working Paper No. 26200.