Price Distortions in High-Frequency Markets*

Jakub Steiner[†]
Northwestern University

Colin Stewart[‡]
University of Toronto

May 17, 2012

Abstract

We study the effect of frequent trading opportunities and categorization on pricing of a risky asset. Frequent opportunities to trade lead to large distortions in prices if some agents forecast future prices using a simplified model of the world that fails to distinguish between some states. In the limit as the period length vanishes, these distortions take a particular form: the price must be the same in any two states that a positive mass of agents categorize together. Price distortions therefore tend to be large when different agents categorize states in different ways. We characterize the limiting prices in terms of rational expectations prices associated with a coarsened process. Similar results hold if, instead of using a simplified model of the world, some agents overestimate the likelihood of small probability events, as in prospect theory.

*We thank Willemien Kets, Filip Matejka, Stephen Morris, Andreas Park, Rani Spiegler, Muhamet Yildiz, and seminar participants for helpful comments.

[†]email: j-steiner@kellogg.northwestern.edu

[‡]email: colin.stewart@utoronto.ca

1 Introduction

Recent advances in technology have led to dramatic increases in trading speed. These changes have generated considerable debate about the effect of high frequencies on market prices, with one side arguing that, by providing liquidity, high frequency traders help to improve market efficiency, and the other side claiming that high frequencies can destabilize market prices. We present a model that lends support to the latter view and show that agents' use of simplifying models of the world can generate large distortions in prices at high trading frequencies.

We study a simple dynamic beauty contest in discrete time capturing pricing of a single risky asset. In each period, each agent chooses a reservation price. The market price aggregates agents' choices. The asset pays a flow dividend that depends on the current state, which is publicly observed and evolves according to a Markov process. In choosing prices, agents consider both the current dividend and the resale price in the next period. A key assumption of our model is that, to form price forecasts, some agents draw on data from similar but different past states. In its simplest form, these agents employ a simplified model of the world in which they fail to distinguish among some differing states. When forecasting the price in the next period, these agents average the price following all past periods that they perceive to be similar to the present one.

If enough agents use the finest possible categorization of states or form rational expectations, then prices are close to rational expectations prices when the period length is not too small. However, as trading opportunities become frequent, distortions become large and prices collapse across states. More precisely, whenever two states are categorized together by a positive but arbitrarily small mass of agents, the price in those two states becomes identical in the high-frequency limit. This result implies that prices are identical whenever two states are connected by a chain of states along which adjacent states are categorized together (possibly by different agents); prices may be identical even across states that *no* agent groups together. Thus distortions tend to be large when categorization is heterogeneous. Moreover, limiting prices admit a simple characterization as rational expectations prices associated with a coarsened process, one in which each state corresponds to a category of states in the true process, and dividends and transition probabilities are convex combinations of those in the true process.

Convergence of prices across large sets of states generates a particular pattern of price behavior over time exhibiting sudden large adjustments. Much of the time, prices do not respond to new information, but occasionally there is an overreaction to small changes in fundamentals. These relatively large price jumps occur when the state transitions between two categories.

Coarse prices arise from a combination of effects. First, learning using categories leads to outcomes as if agents form rational expectations based on incorrect beliefs about the process governing the evolution of states. These incorrect beliefs take a particular form: in addition to transitions that occur under the true process, agents behave as if they have assigned some probability to the state changing to another state in the same category. These beliefs are nonvanishing as the period length vanishes. At the same time, the current dividend is proportional to the length of time the asset is held. Hence shortening the period length reduces the influence of the dividend in price formation relative to the expected future price. Thus the "as if" belief that the state is likely to change drives the price together within each category.

Intuitively, one can think of the effect as follows. Suppose for contradiction that prices differ within a category used by some agents. Consider the state within that category at which the price is minimal (and suppose that there are no other agents categorizing that state together with one where the price is even lower). Then, because agents using this category act as if they believe the state is likely to transition to another state within the category, they act as if the price is likely to go up very soon. Consequently, those agents will demand a large quantity at the current price. This demand drives up the price, causing agents with rational expectations to view the asset as overpriced, and hence to sell. Note, however, that as the trading frequency shrinks, so do the dividend and the transition probabilities, causing the incentive of rational agents to sell the asset to decrease in proportion to the period length. This means that, when trading is frequent, agents using categorization expect a much larger gain from buying than do agents using rational expectations from selling, causing them to demand more and drive the price up further. As the period length vanishes, these two pressures on the market price will be balanced only if the differences in prices within the category also vanish.

At a technical level, our results can be understood in terms of higher order expectations about prices. Consider prices in a steady state, and suppose for simplicity that some agents categorize all states together. Because those agents act as if they overestimate the likelihood of transitions, higher order expectations converge quickly to the expectation with respect to the stationary distribution regardless of the current state. Moreover, since this convergence pertains to the near future, the current price must be close to the higher order expectation, implying that prices are roughly constant across states. A similar argument when agents use multiple categories gives rise to a characterization of prices in terms of rational expectations prices associated with a simplified process governing the evolution of states.

Our results extend naturally along two lines. First, category-based learning can be generalized to a broad class of learning rules in which agents extrapolate from similar but different past states. Defining categories in terms of states that are perceived to be similar yields the same type of results, although the precise characterization of prices is somewhat more complicated. Second, similar results hold if, instead of learning by similarity, agents behave as if they form rational expectations but some agents overweight the likelihood of small probability events, as in prospect theory. Standard weighting functions used in the prospect theory literature (see, e.g., Prelec 1998, Gonzalez and Wu 1999) have the property that, relative to the true probabilities, weighted probabilities grow large as the probability vanishes. This property has essentially the same effect as the nonvanishing transition probabilities in the "as if" beliefs that arise under learning by similarity. Hence we obtain a similar result exhibiting coarse prices.

Throughout most of the paper, we take a reduced form approach to price formation that eschews modelling demand and supply explicitly, and rather expresses current market price directly in terms of forecasts of future prices. The reduced form approach is sufficient to highlight the mechanism underlying the distortionary effect, and the main reason for taking it is tractability. The pricing equation we apply directly would follow in a standard CARA-Normal overlapping generations framework. However, a key assumption of the standard framework is that agents live for only two periods, which is seemingly incompatible with our focus on vanishingly small period length. Accordingly, we examine robustness of our results in a model with explicit supply and demand and agents with long investment horizons. Although such a model appears to be intractable in general, it is tractable when the state space is binary. In Section 6 we solve this special case of the model, and show that the results are in accordance with the reduced form analysis. The microfounded analysis has the added benefit of permitting the study of trade volumes and wealth accumulation. As the period length vanishes, we find that traders take ever larger positions, with, in any given state, agents using one form of categorization demanding increasingly large quantities that are supplied by agents using a different categorization (and both groups changing sides when the state changes).

Our main result is stark and should not be taken too literally; the main goal of the paper is simply to elucidate a mechanism by which high frequencies may amplify distortions resulting from imperfect rationality. The discussion section describes some variations on the model that may go against the constant price result (although the effect remains).

2 Related Literature

A number of earlier papers have highlighted the role of strategic complementarities in amplifying the effect of irrational agents (e.g. Haltiwanger and Waldman 1985, Haltiwanger and Waldman 1989, Fehr and Tyran 2005). Our main result is driven in part by this effect, which is compounded in our model because high frequencies strengthen strategic complementarities.

Our results can be understood in terms of higher order expectations about future prices. Among others, Allen, Morris, and Shin (2006) and Bacchetta and Van Wincoop (2008) have highlighted the role of higher order expectations in financial markets with asymmetric information. Since our model is one of complete information, the main thrust of those papers is somewhat orthogonal to the present one.

De Long, Shleifer, Summers, and Waldmann (1990b) show that irrational traders can induce rational agents to behave in a way that destabilizes prices: if irrational traders chase trends, rational traders' demands increase ahead of an upturn in anticipation of greater demand from irrational traders. Our model does not have this feature. Rational traders act as a stabilizing influence, but do not fully stabilize prices. Similarly, in De Long, Shleifer, Summers, and Waldmann (1990a) the same authors study how noise trader risk can distort prices, and show that risk aversion, by limiting the size of positions, can cause rational traders to receive lower expected returns than do noise traders. These papers are related in spirit to our point that irrationality can drive prices away from rational expectations, but the mechanisms are very different.

While the style investing of Barberis and Shleifer (2003) is similar to the categorization in our model, the focus is quite different. They consider a case with a large fraction of investors who divide assets into a common set of styles over a fixed time horizon, while we focus on the effect

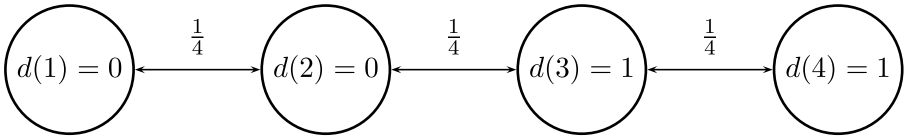

Figure 1: Markov process for the examples of Section 3. Arrows depict possible transitions labelled according to the rate at which they occur. Each $d(\omega)$ denotes the dividend in state ω.

of shortening horizons when agents use a variety of categorizations. Another key difference is that Barberis and Shleifer assume that demands are based on relative past performance, while in our model they are based on absolute past prices. Similar comments apply to Bianchi and Jehiel (2010), who show that bubbles and crashes can arise when some agents form expectations about price movements that are incorrect but consistent with the average across multiple periods.

3 Examples

We begin with a simple example to illustrate how categorization by some agents leads to large distortions in prices when trading is frequent.

3.1 Two categories

A continuum of agents trade a single asset at times $t = 0, \Delta, 2\Delta \ldots$. The asset pays a dividend of $d(\omega_k)\Delta$ per period that depends on the current state $\omega_k \in \{1, 2, 3, 4\}$ in period $k = t/\Delta$. The state is publicly observed prior to trade in each period. The state evolves according to the continuous-time Markov process depicted in Figure 1. Flow dividends $d(\omega)$ are 0 in states 1 and 2 and 1 in states 3 and 4.

We take a reduced-form approach to price formation. In each period k, each agent i chooses a price p_k^i, which can be interpreted as the price at which her demand is zero. The market price p_k is the average of all individual prices; that is,

$$p_k = \int_i p_k^i di.$$

For each agent i, the individual choice of price in period k is given by

$$p_k^i = \int_{k\Delta}^{(k+1)\Delta} d(\omega_k) e^{-(t-k\Delta)} dt + e^{-\Delta} Q_{k+1}^i = d(\omega_k)(1 - e^{-\Delta}) + e^{-\Delta} Q_{k+1}^i,$$

where Q_{k+1}^i is agent i's forecast of the price at the next trading opportunity at time $(k+1)\Delta$, formed as described below. Thus the price chosen by agent i is equal to her expected gross profit from holding one unit of the asset for one period. The discount rate is normalized to 1, which is without loss of generality. For simplicity, we assume that flow dividend is constant throughout each trading period.

For this example, agents are divided into two groups. A fraction $1 - \pi \in [0, 1)$ of agents form rational expectations given all parameters of the model. The remaining fraction π are *coarse thinkers* who form forecasts based on a simplified model of the market that fails to distinguish between some states. For the purpose of this example, coarse thinkers fail to distinguish between states that have the same dividend (the general analysis allows for arbitrary categorization that may group together states having different dividends). That is, coarse thinkers group states 1 and 2 together, and states 3 and 4 together; accordingly, let $C_{12} = \{1, 2\}$ and $C_{34} = \{3, 4\}$ denote these categories. Coarse thinkers form price forecasts according to the average of prices p_{s+1} across all past periods s in which the state was in the same category as the current state. For example, if the current state $\omega_k = 2$ then their forecast in period k of the price p_{k+1} is

$$\frac{\sum_{s<k-1:\omega_s \in C_{12}} p_{s+1}}{\sum_{s<k-1:\omega_s \in C_{12}} 1}$$

whenever the denominator is nonzero. If the denominator is 0 the forecast is an arbitrary fixed value.

We focus on Markovian steady-state prices, which consist of a collection $P(\omega)$ of prices, one for each state, satisfying

$$P(\omega) = d(\omega)(1 - e^{-\Delta}) + e^{-\Delta} \left((1 - \pi) Q^R(\omega) + \pi Q^C(\omega) \right), \tag{1}$$

where $Q^R(\omega)$ and $Q^C(\omega)$ denote the steady-state price forecasts in state ω of agents with rational expectations and coarse thinkers, respectively. We show in Section 4 that convergence to the

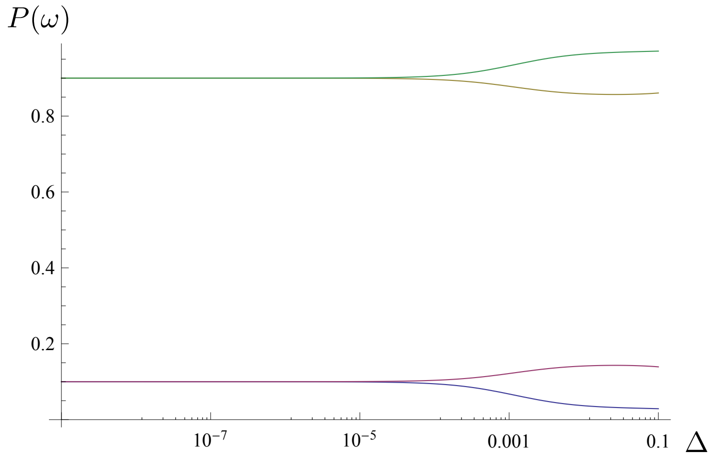

Figure 2: Steady-state prices as a function of Δ when the share of the coarse thinker is 10^{-3}. Each of the four curves depicts the price in one of the four states.

steady-state prices is guaranteed under the learning process described above.

The steady-state price forecasts $Q^R(\omega)$ and $Q^C(\omega)$ are determined by the steady-state prices together with the transition probabilities and the categorization. Letting $q_\Delta(\omega, \omega')$ denote the single-period transition probability from ω to ω', the rational expectations forecasts are

$$Q^R(\omega) = \sum_{\omega'} q_\Delta(\omega, \omega')P(\omega'). \tag{2}$$

Coarse thinkers' forecasts are given by

$$Q^C(\omega) = \frac{1}{2}\sum_{\omega'} q_\Delta(\omega, \omega')P(\omega') + \frac{1}{2}\sum_{\omega'} q_\Delta(\hat{\omega}, \omega')P(\omega'), \tag{3}$$

where $\hat{\omega} \neq \omega$ is the other state in the same category as ω. To understand the formula, first note that the stationary distribution of the process q_Δ assigns equal probability to all four states. Thus when the coarse thinker forms the price forecast by averaging over past periods, on average half of the relevant past states will be ω and the other half $\hat{\omega}$.

If all agents form rational expectations, the price in each state is equal to the expected discounted sum of dividends over the entire future. In particular, rational expectations prices differ across all four states; for two states with the same dividend, the difference arises because of the difference in transition rates to states with another dividend.

Returning to the case in which some agents are coarse thinkers, by equations (1), (2), and (3), for each Δ, steady-state prices solve a system of linear equations, one for each state. Figure 2 depicts steady-state prices as a function of the period length Δ when coarse thinkers form 0.1% of the population. Not surprisingly, when the period length is not too short, prices are close to those obtained if all agents form rational expectations. However, as the period length vanishes, prices collapse within each category used by coarse thinkers: prices in states 1 and 2 become identical, as do prices in states 3 and 4. Prices fail to distinguish between states within each category used by coarse thinkers even though coarse thinkers form only a small fraction of agents. Consequently, relative to rational expectations prices, the time path of steady-state prices involves infrequent changes in price combined with occasional overreactions to changes in the state (occurring when the state changes from 2 to 3 or vice versa).

Why do prices collapse within categories at high frequencies? Coarse thinkers behave as if they were overestimating transition rates within each category. They form forecasts exactly as if they believe that, at the beginning of each period, there is a 50% chance that the state will immediately transition to the other state in the current category (after which the state follows the process q_Δ until the next period). According to this belief, when the period length is short, transitions within categories are much more likely than they are under the true process. Consequently, prices within categories are driven together because coarse thinkers' forecasts of the future price assign significant weight to the price in the other state within the current category. To some extent, rational agents counteract this effect. However, since their expected profits per period vanish along with the period length, so does the gap between their individual prices and the market price; in the limit, the corrective pressure exerted by rational agents disappears. As a result, coarse thinkers effectively determine prices at high frequencies even if they form only a small fraction of the market.

Note that the distortion in prices at high frequencies is due to the use of coarse categories by some agents, not because of the naïve rule used to forecast prices. If all agents followed the same learning rule using the finest categorization (in which each state lies in its own category), then steady-state prices would be identical to the rational expectations prices.

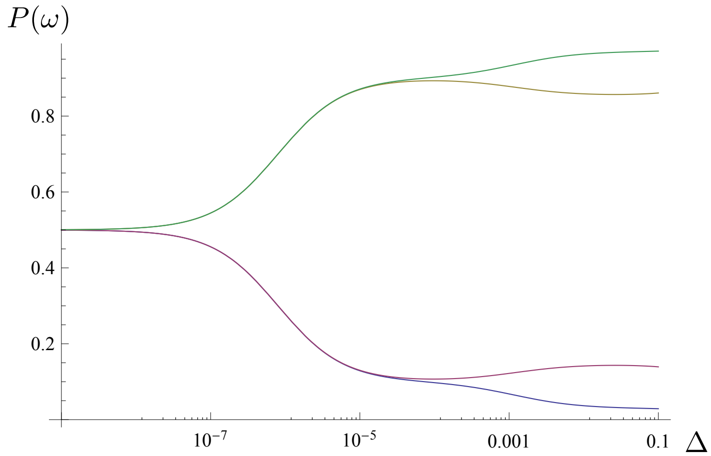

Figure 3: Steady-state prices as a function of Δ for $\pi = 10^{-3}$ and $\pi' = 10^{-6}$. Each of the four curves depicts the price in one of the four states.

3.2 Overlapping categories

Consider the modification of the previous example obtained by replacing a fraction $\pi' \in (0, 1-\pi]$ of agents forming rational expectations with agents who categorize states 2 and 3 together but place states 1 and 4 in singleton categories. As before, a fraction π of agents use categories C_{12} and C_{34}, while the remaining fraction $1 - \pi - \pi'$ form rational expectations. Otherwise, all elements of the example are identical.

Figure 3 depicts the steady-state prices as a function of Δ for $\pi = 10^{-3}$ and $\pi' = 10^{-6}$. As in the previous example, when the period length is not too short, prices are close to rational expectations prices. Since π' is small relative to π, as Δ shrinks prices initially remain similar to those of the previous example. In contrast, however, once Δ is small enough, prices collapse across all states to a single price.

Relative to the previous example, the addition of a group categorizing one state in C_{12} together with another in C_{34} is enough to drive all prices together under high frequencies. In particular, prices in states 1 and 4 are identical in the limit as Δ vanishes even though *no* agent categorizes those states together and they differ in terms of fundamentals. In general, our results indicate that if agents categorize states in different ways, price distortions tend to be severe in many states when Δ is small; overlapping categorization leads to constant prices across large sets of states.

4 Categorization

The convergence of prices across states in the examples of Section 3 occurs much more generally as a result of coarse thinking. In this section, we present a general model capturing the idea that, when learning to predict prices, some agents may employ a simplified model of the world that fails to perfectly distinguish among all relevant states. The learning process of this section is generalized in Appendix A to a broad class of similarity-based procedures. Sections 5 and 6 present related models that exhibit similar coarse pricing results.

4.1 Model

We consider a single asset whose dividend depends on a state $\omega(t)$ drawn from a finite set Ω. The state evolves according to an ergodic continuous-time stationary Markov process with transition rates $q(\omega, \omega')$ from ω to ω'. Trading occurs at discrete times $t = 0, \Delta, 2\Delta, \ldots$. We refer to time $k\Delta$ as period k, and write ω_k for the state $\omega(k\Delta)$ in period k. Sampling the continuous-time process q at times $k\Delta$ gives rise to a discrete-time Markov process (sometimes called the discrete skeleton of q at scale Δ) with transition probabilities $q_\Delta(\omega, \omega')$. The state affects the flow dividend $d(\omega_k)$ of the asset, which is paid at a constant rate from time $k\Delta$ to $(k+1)\Delta$.[1]

A continuum of agents indexed by $i \in [0, 1]$ trades the asset in each trading period k. Trading decisions are based on the current dividend and on agents' forecasts of the prices in the following period. Agents form these forecasts as follows. Each agent i categorizes states according to a partition Π^i of Ω that is fixed across all periods. Letting $\Pi_1, \ldots, \Pi_N$ denote those partitions belonging to a positive measure of agents, we write π_n for the measure of agents using partition Π_n. For each state ω, let $\Pi(\omega)$ denote the element of the partition Π containing ω. In period k, agent i forms a forecast Q^i_{k+1} of the price in period $k+1$ according to

$$Q^i_{k+1} = \frac{\sum_{s<k-1:\omega_s \in \Pi^i(\omega_k)} p_{s+1}}{\sum_{s<k-1:\omega_s \in \Pi^i(\omega_k)} 1}$$

whenever the denominator is nonzero (otherwise take the forecast to be some arbitrary fixed number), where p_s denotes the market price in period s as described below. Thus the price forecast

[1] The assumption that the dividend remains constant between periods instead of changing according to the continuous-time Markov process simplifies the notation but makes no difference for our results.

Q^i_{k+1} is formed by averaging all prices that occurred in periods immediately following those in which the state was in the same category as the current one (according to Π^i).

As in the examples, in each period, each agent chooses a price, which can be interpreted as the current price at which their demand for the asset would be zero. Agent i's choice in period k is given by

$$p^i_k = d(\omega_k)\left(1 - e^{-\Delta}\right) + e^{-\Delta} Q^i_{k+1}. \tag{4}$$

The first term captures the discounted value of the flow dividend paid from time $k\Delta$ to $(k+1)\Delta$. The second term captures the effect of the anticipated resale price. The market price p_k is defined to be the average of p^i_k across all agents i; that is, letting p^n_k denote the price of each agent using partition Π_n,

$$p_k = \sum_{n=1}^{N} \pi_n p^n_k. \tag{5}$$

4.2 Steady-state prices

Proposition 1 below shows that this learning process converges to steady-state prices $P : \Omega \longrightarrow \mathbb{R}$ that depend only on the current state. Steady-state prices turn out to be identical to rational expectations prices, not with respect to the true process, but with respect to a different process that reflects both the true process q_Δ and the categorizations used by agents.

Definition 1. Given any Δ, prices $P(\omega)$ are (steady-state) *rational expectations prices* with respect to a Markov process m on Ω and a dividend function d if

$$P(\omega) = d(\omega)(1 - e^{-\Delta}) + e^{-\Delta} E_{m(\omega,\omega')}[P(\omega')]$$

for every $\omega \in \Omega$.[2]

Note that the rational expectations price $P(\omega)$ is the discounted stream of expected dividends starting from state ω under the Markov process m.

Rational expectations prices with respect to q_Δ depend on Δ only because of the simplifying assumption that flow dividends do not change between trading periods. We sometimes refer to rational expectations without reference to Δ to mean the limiting prices as Δ vanishes. Accordingly,

[2]Note that for any m and d, rational expectations prices exist and are unique.

rational expectations prices with respect to q are given by the unique solution to the system of equations

$$P(\omega) = \frac{d(\omega) + \sum_{\omega' \neq \omega} q(\omega, \omega') P(\omega')}{1 + \sum_{\omega' \neq \omega} q(\omega, \omega')}.$$

Let ϕ denote the stationary distribution of states with respect to the true process q, which satisfies the global balance equations $\sum_{\omega' \neq \omega} \phi(\omega) q(\omega, \omega') = \sum_{\omega' \neq \omega} \phi(\omega') q(\omega', \omega)$. For given initial prices and a given realization of the sequence of states $(\omega_s)_{s=0}^{k-1}$, let $p_k(\omega)$ denote the price in period k that would obtain if $\omega_k = \omega$. Define the *modified process* by

$$m_\Delta(\omega, \omega') = \sum_{n=1}^{N} \pi_n \sum_{\omega'' \in \Pi_n(\omega)} \phi\left(\omega'' \left| \Pi_n(\omega)\right.\right) q_\Delta(\omega'', \omega'). \tag{6}$$

Proposition 1. *For each Δ, the sequence $p_k(\omega)$ almost surely converges to the vector $P_\Delta(\omega)$ of rational expectations with respect to the modified process m_Δ and the dividend function d.*

All proofs are in the appendix. In particular, Proposition 1 is a corollary of Proposition 5 in Appendix A.

To understand the modified process m_Δ, first consider the case in which all agents distinguish all states, i.e. $\Pi^i(\omega) = \{\omega\}$ for every ω and i. In this this case, $m_\Delta = q_\Delta$, and hence the long-run prices are precisely the rational expectations prices with respect to the true process. To see why, consider the forecasting procedure. In period k, each agent uses data from previous periods $s < k - 1$ in which the state was indistinguishable from the current state (according to her own categorization). For the finest categorization, these relevant periods are those s such that $\omega_s = \omega_k$. In the steady state, the agent's forecast is just the average of $P(\omega_{s+1})$ across the relevant periods s. In the long run, the forecast is equal to $\sum_{\omega'} q_\Delta(\omega_k, \omega') P(\omega')$, coinciding with the rational expectation of the price in the next period.

For general categorizations, a given agent's forecast is based on all previous periods s in which the state ω_s belonged to the current category $\Pi^i(\omega_k)$. In the long run, the average of $P(\omega_{s+1})$ for those values of s is equal to

$$\sum_{\omega'' \in \Pi^i(\omega)} \phi\left(\omega'' \left| \Pi^i(\omega)\right.\right) q_\Delta(\omega'', \omega') P(\omega'),$$

where the term $\phi\left(\omega'' \left|\Pi^i(\omega)\right.\right)$ captures the long-run frequency of state ω'' in the sample of relevant periods s. Taking the average across agents, the population-wide forecast is the expectation with respect to the modified process m_Δ in (6).

Appendix A extends Proposition 1 in two directions. First, we extend the price forecasting rule to a general class of similarity-based rules in which agents forecast using data from similar past states. Unlike the categorization considered here, the weights assigned to different states may vary according to the perceived degree of similarity. Second, we allow for an arbitrary fraction of agents to form rational expectations knowing all parameters of the model, including other agents' forecasting procedures. In the long run, such agents have the same effect on prices as agents who categorize every state separately.

4.3 High frequency

While Proposition 1 provides a general characterization of long-run prices for any Δ, focusing on high frequencies (i.e. vanishing Δ) leads to a striking result: prices generally fail to distinguish among states that may differ substantially in terms of fundamentals.

Let Π denote the meet of $\Pi_1, \ldots, \Pi_N$.[3] We refer to the elements of Π as *aggregate categories*. Two states ω and ω' lie in the same aggregate category if and only if there exists a sequence $\omega_1, \ldots, \omega_r$ of states such that $\omega = \omega_1$, $\omega' = \omega_r$, and for each $\ell = 1, \ldots, r-1$, $\omega_{\ell+1} \in \Pi_n(\omega_\ell)$ for some $n \in \{1, \ldots, N\}$. In particular, given two states in different aggregate categories, almost every agent distinguishes between those two states, but the converse is not true in general; two states in the same aggregate category may be distinguished by all agents.

The main result of this paper shows that, in the limit as Δ vanishes, prices are constant on aggregate categories. Moreover, prices approach rational expectations prices with respect to a process that is coarser than the true process, with each aggregate category playing the role of an individual state.

Define the *coarse process* $\overline{q}$ to be the continuous-time Markov process on the state space Π with transition rates

$$\overline{q}(C, C') = \sum_{\omega \in C} \phi(\omega|C) \sum_{\omega' \in C'} q(\omega, \omega').$$

[3]The meet of a collection of partitions is defined to be their finest common coarsening.

That is, the coarse process $\overline{q}$ is obtained by averaging the true process q across each category with respect to the stationary distribution ϕ. Similarly, define the *coarse dividend function* $\overline{d}$ by averaging dividends on each category $C \in \Pi$; that is,

$$\overline{d}(C) = \sum_{\omega \in C} \phi(\omega|C) d(\omega)$$

for each $C \in \Pi$.

Theorem 2. *As Δ vanishes, prices become constant on each aggregate category; that is, for any ω, ω' such that $\Pi(\omega) = \Pi(\omega')$,*

$$\lim_{\Delta \to 0} \left(P_\Delta(\omega) - P_\Delta(\omega') \right) = 0.$$

Moreover, for each ω, $\lim_{\Delta \to 0} P_\Delta(\omega)$ is equal to the rational expectations price with respect to $\overline{q}$ and $\overline{d}$ in state $\Pi(\omega)$.

The theorem is a special case of Theorem 7 in Appendix B.

For high trading frequencies, this result indicates that whenever a positive mass of agents fail to distinguish between two states the market price will be the same in those states. However, that is not all: prices may often be the same in two states even if *no* agent categorizes them together. This is the case if there is an overlapping chain of categories connecting these states; for example, if, as in the example of Section 3.2, one group categorizes ω and ω'' together while another group categorizes ω'' and ω' together, then ω and ω' lie in the same aggregate category and prices are the same across all three states. Consequently, if agents do not use the same categories, aggregate categories will typically be large, potentially leading to large distortions in prices. Put differently, market prices represent a coarser view of the world than that held by individual market participants.

4.4 Example: Pricing of irrelevant information

To illustrate Theorem 2, consider the underlying process depicted in Figure 4 with state space $\{1, 2, 3, 4\}$. States 1, 2, and 3 are identical in terms of underlying fundamentals (i.e. in terms of the distribution of future dividends) since they share the same dividend and transition rate to state 4, the only state at which the dividend differs from zero. If all agents form rational expectations, it

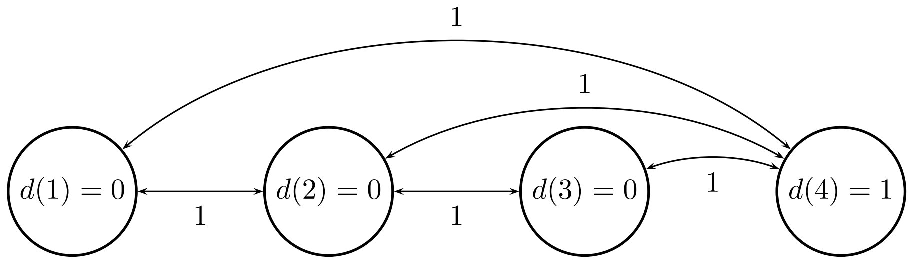

Figure 4: Markov process for the example of Section 4.4.

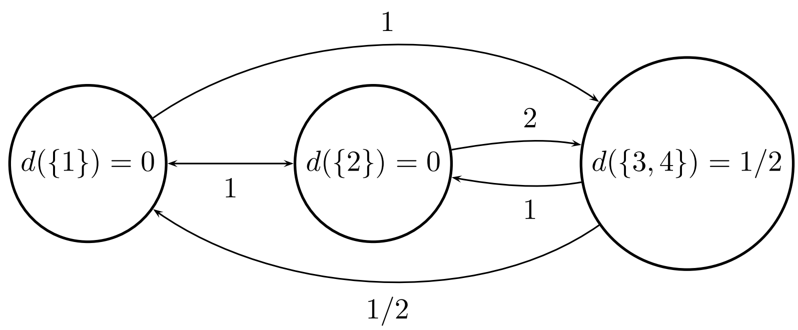

Figure 5: Coarsened Markov process for Example 4.4.

is straightforward to show that, in the limit as Δ vanishes, prices approach

$$P(1) = P(2) = P(3) = \frac{1}{6} \qquad \text{and} \qquad P(4) = \frac{1}{3}.$$

Now suppose that a positive mass of agents categorize states 3 and 4 together, that is, they use the partition $\Pi = \{\{1\}, \{2\}, \{3, 4\}\}$, while all other agents use the finest partition. Hence aggregate categories are given by Π. According to Theorem 2, limiting prices are equal to rational expectations prices associated with a coarsened process with state space Π, as depicted in Figure 5. The coarsened process is obtained by averaging the true process with weights determined by the stationary distribution ϕ. Since the stationary distribution is uniform, transition rates out of state $\{3, 4\}$ in the coarsened process are obtained by taking a simple average of the corresponding rates in the true process, while transition rates to state $\{3, 4\}$ are obtained by adding the corresponding rates in the true process. Similarly, the dividend in state $\{3, 4\}$ is simply the average of the original dividends.

Calculating the rational expectations prices associated with the coarsened process yields

$$P(\{1\}) = \frac{6}{35}, \quad P(\{2\}) = \frac{1}{5}, \quad \text{and} \quad P(\{3, 4\}) = \frac{11}{35}.$$

While our results may suggest that, with high frequencies, categorization leads to an unambiguous coarsening of prices, the present example shows that not to be true in general. Not only do prices in states 3 and 4 become equal, but prices also differ across states 1 and 2 since these states differ in terms of the rate of transitions to the aggregate category $\{3, 4\}$. In particular, even though all agents correctly distinguish states 1 and 2 from all other states, and these two states feature identical fundamentals, the price is higher in state 2 than in state 1.

5 Weighted probabilities

The last section indicates that coarsening of steady-state prices arises because agents who categorize two given states together effectively overestimate the probability of transitions between them. In this section we analyze a simple setting in which a similar result arises not from coarse thinking but from the use of weighted probabilities, as in prospect theory. In particular, we assume here that some agents overweight the likelihood of small probability events.

The sets of agents and actions, the discount rate, and the Markov process are the same as in Section 4. The key difference is in agents' forecasts of future prices. Each agent i forms forecasts using a belief $m^i_\Delta(\omega, \omega')$ capturing the probability she assigns to the state at time $t + \Delta$ being ω' if the state at time t is ω. Let

$$m_\Delta(\omega, \omega') = \int_i m^i_\Delta(\omega, \omega') di \tag{7}$$

denote the average belief. We focus on rational expectations prices with respect to m_Δ and d.

For some $\pi \in (0, 1]$, a fraction of $1 - \pi$ of all agents i have correct beliefs $m^i_\Delta(\omega, \omega') \equiv q_\Delta(\omega, \omega')$. The remaining fraction π form beliefs using a probability weighting function $\lambda : [0, 1] \longrightarrow \mathbb{R}_+$; that is, each such agent i has beliefs given by

$$m^i_\Delta(\omega, \omega') = \frac{\lambda(q_\Delta(\omega, \omega'))}{\sum_{\omega''} \lambda(q_\Delta(\omega, \omega''))}. \tag{8}$$

We assume that λ overweights the likelihood of small probability events in the sense that $\lim_{p \to 0^+} \frac{\lambda(p)}{p} = +\infty$. This assumption holds for most weighting functions commonly used in prospect theory (e.g., Prelec 1998, Gonzalez and Wu 1999).

For simplicity, assume in addition that $q(\omega, \omega') > 0$ for all distinct $\omega, \omega' \in \Omega$. Without this assumption, the same result holds under more restrictive conditions on λ. For example, Proposition

3 holds without this assumption for weighting functions of the form $\lambda(p) = \exp\left(-\zeta(-\ln p)^{\xi}\right)$ with $\zeta > 0$ and $\xi \in (0,1)$, as axiomatized by Prelec (1998).

For each Δ, let $P_\Delta \in \mathbb{R}^\Omega$ be rational expectations prices with respect to m_Δ and d.

Proposition 3. *As Δ vanishes, prices become constant across all states; that is, for any $\omega, \omega' \in \Omega$, we have*

$$\lim_{\Delta \to 0} P_\Delta(\omega) - P_\Delta(\omega') = 0.$$

Moreover, in the limit, the price in each state is equal to the expected dividend with respect to the limit of the stationary distribution of m_Δ as $\Delta \to 0$ (provided that the limit exists).

Proposition 3 is a special case of Theorem 7 in Appendix B.

The intuition behind Proposition 3 is as follows. According to the average belief m_Δ, when Δ is small, the state is likely to transition away from ω within a short time (although it may take many periods). In fact, with respect to absolute time, the distribution of future states quickly converges to the stationary distribution of m_Δ. Since current prices ultimately depend on higher-order expectations of prices in future periods, when Δ is small, prices can be approximated by prices obtained when dividends and transition probabilities are replaced with their averages according to the stationary distribution.

5.1 Example

Assume that all agents form forecasts using probabilities weighted according to $\lambda(p) = p^\beta$, where $\beta \in (0,1)$. That is, for any agent i,

$$m^i_\Delta(\omega, \omega') = \frac{(q_\Delta(\omega, \omega'))^\beta}{\sum_{\omega''} (q_\Delta(\omega, \omega''))^\beta}$$

for every ω and ω'. Consider the two-state process depicted in Figure 6 with $q_1, q_2 > 0$.

By Proposition 3, prices are constant across states in the limit as $\Delta \to 0$. Moreover, prices are equal to a weighted average of the dividends in the two states. However, unlike the case of categorization, the average is not weighted according to the stationary distribution ϕ, but rather according to the stationary distribution of m_Δ. As $\Delta \to 0$, the stationary distribution of m_Δ tends to the distribution assigning probability $q_2^\beta / \left(q_1^\beta + q_2^\beta\right)$ to state 1 (whereas the stationary

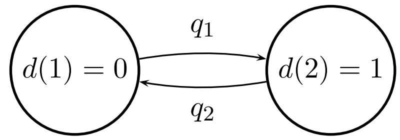

Figure 6: Markov process for the example of Section 5.1.

distribution of the true process is identical except with 1 in place of each β). Therefore, by Proposition 3, the price in each state approaches $q_1^\beta / \left(q_1^\beta + q_2^\beta\right)$ as Δ vanishes. The less frequent state receives disproportionate weight in price formation.

6 Long horizons

For the sake of tractability, we have taken a reduced form approach to price formation. The pricing equations (4) and (5) that we use arise naturally in CARA-Normal overlapping generations models where agents live for only two periods (see, e.g., Allen, Morris, and Shin (2006)). Solving such a model with longer-lived agents (and uncertainty that is not normally distributed) is intractable in general; we can, however, solve the case in which there are only two states, $\Omega = \{0, 1\}$. Our basic insight extends to this setting. If a fraction of the population categorizes the two states together or overestimates transition probabilities then prices collapse across the two states.

Consider the following overlapping generations model. Each agent i participates in the market for a random number of periods $n^i \geq 1$, where $n^i = n$ with probability $\Delta(1-\Delta)^{n-1}$. Agent i does not observe n^i until she exits the market. Exiting agents are replaced by entrants. Agents trade an asset of 0 supply whose flow dividend $d(\omega)$ depends on the current state, which evolves according to a Markov chain q_Δ on $\{0, 1\}$. The analysis of this section focuses on the steady state; agents form their demands facing steady-state prices $P(\omega)$. That is, an agent i entering the market in period k_0 chooses in each period $k \in \{k_0, \ldots, k_0 + n^i - 1\}$ a demand α_k, and accumulates wealth

$$w = \sum_{k \in \{k_0, \ldots, k_0 + n^i - 1\}} e^{-(k-k_0)\Delta} \alpha_k \left(-P(\omega_k) + \left(1 - e^{-\Delta}\right) d(\omega_k) + e^{-\Delta} P(\omega_{k+1})\right),$$

where the term $e^{-(k-k_0)\Delta}$ captures time discounting. Each agent consumes all of her wealth when she exits (after period $k_0 + n^i - 1$), and has constant absolute risk aversion utility $u(w) = -e^{-w}$.

The population of agents consists of a group R of rational agents and a group C of coarse thinkers, with measures π_R and π_C respectively, where $\pi_R + \pi_C = 1$. Agents from group $G \in \{R, C\}$ believe that the asset evolves according to a Markov chain $q^G_\Delta(\omega, \omega')$, where q^R_Δ is the true process q_Δ. Where there is no risk of confusion, we omit the subscript Δ.

A *stationary equilibrium* consists of prices $P(\omega)$ and demands $\alpha^G(\omega)$ such that

1. Agents optimize:

$$\alpha^G(\omega_k) \in \arg\max_\alpha E_G \left[-\exp\left(-w' - \alpha \left(-P(\omega_k) + \left(1 - e^{-\Delta}\right) d(\omega_k) + e^{-\Delta} P(\omega_{k+1}) \right) \right) \middle| \, \omega_k \right]$$

   where E_G denotes the expectation with respect to the process q^G_Δ and

$$w' = \sum_{s \in \{k_0, \dots, k-1, k+1, \dots, k_0+n^i-1\}} e^{-(s-k_0)\Delta} \alpha^G(\omega_s) \left(-P(\omega_s) + \left(1 - e^{-\Delta}\right) d(\omega_s) + e^{-\Delta} P(\omega_{s+1}) \right)$$

   is the wealth accumulated in periods other than k. Note that, as implied by the notation, individual demands are independent of wealth and time spent in the market. The former follows from constant absolute risk aversion and the latter from stationarity.

2. The market clears:

$$\sum \pi_G \alpha^G(\omega) = 0 \text{ for all } \omega.$$

We assume that coarse thinkers overweight transition probabilities across the two states in such a way that $\lim_{\Delta \to 0} \frac{q^C_\Delta(\omega, \omega')}{q_\Delta(\omega, \omega')} = \infty$ for $\omega \neq \omega'$.

Proposition 4. *As the period length vanishes, prices collapse across the two states; that is,*

$$\lim_{\Delta \to 0} \left(P_\Delta(1) - P_\Delta(0) \right) = 0.$$

7 Discussion

In order to highlight the effect of high frequencies, we have focused on a simple tractable model in which the resulting prices take a stark form. A number of natural modifications of the model may moderate the effect while retaining significant price distortions with high frequencies. In this section, we speculate about the consequences of various extensions and modifications to the main model.

Traders in our model live forever and have no limits on losses. Since traders using coarse models tend to lose money against traders using refinements of those models, forcing traders to exit once reaching a given loss threshold could drive all agents out of the market except those who form rational expectations (or use the finest categorization), thereby eliminating price distortions. However, since our results hold independent of the fractions of agents using various partitions, we conjecture that our results hold as long as there is continual entry of a nonvanishing mass of new traders using coarse categories.

In the variant of our model that incorporates individual demands in Section 6, risk aversion limits the size of the position taken by each trader. Since agents who form rational expectations perceive the risk of state transitions over a short horizon to be very low, risk aversion tends to limit their positions less than those of agents who use coarse categorization. Thus at a fixed trading frequency we expect that increasing risk aversion should reduce price distortions. However, as Section 6 indicates, risk aversion alone does not overturn the constant price result in the high frequency limit.

For the sake of parsimony, we have assumed that agents employ categories that are fixed across time. Alternatively, one might expect agents to adjust their categories as they learn the correct model. If learning leads to successive refinements in categorization toward the finest categorization, then our results may not hold in the long-run. As with bankruptcies, however, we expect that continual entry of agents using coarse categories would suffice to generate persistent price distortions.

Another simplifying assumption of our model is that all agents have identical short horizons. If instead some agents maintain their positions for some fixed time that is independent of the period length then these agents may drive prices back toward fundamentals. On the one hand, under our reduced form approach in Sections 4 and 5, the addition of such traders would reduce the impact of coarse categorization, leading to prices lying between rational expectations and constant prices on categories. On the other hand, explicitly considering individual demands may mitigate the influence of these agents since agents with shorter horizons face less risk and therefore tend to take larger positions. Again, as Section 6 indicates, coarse pricing may prevail even when agents face long trading horizons.

A Convergence of Prices

This appendix proves convergence of the learning process described in Section 4, extends the result to a more general class of processes in which agents learn from similar past states, and shows that our results remain unchanged if we allow for some agents to form rational expectations. We start by describing learning by similarity, which includes categorization as a special case. We then consider an even more general class of processes that is sufficiently broad to allow for the inclusion of agents who form rational expectations about future states and other agents' behavior.

A.1 Learning by similarity

The categorization framework of Section 4 is a special case of a model in which agents learn prices based on past prices in states similar to the current one, but do not necessarily apply equal weight to all similar states. The convergence and characterization result of Proposition 1 extends to this more general case.

Each agent i is endowed with a symmetric similarity function $g_i : \Omega \times \Omega \longrightarrow \mathbb{R}_+$ determining the weight assigned to various states in forming forecasts of future prices. We assume that for each i and ω, there exists some ω' such that $g_i(\omega, \omega') \neq 0$. Given a history of states and prices up to period $k-1$, agent i's forecast in period k of the price in period $k+1$ is

$$Q^i_{k+1} = \frac{\sum_{s<k-1} g_i(\omega_k, \omega_s) p_{s+1}}{\sum_{s<k-1} g_i(\omega_k, \omega_s)}$$

whenever the denominator is nonzero, and some fixed constant otherwise. Thus the forecast is formed by averaging the one-period-ahead prices in all past states, weighted according to the degree of similarity to the current state. The categorization of Section 4 is a special case of this framework in which, for each i, g_i takes only the values 0 and 1.

For simplicity, we assume that only a finite number of different similarity functions are used by the agents. That is, there exists a finite partition of the population into groups of measures

$\pi_1, \ldots, \pi_N$, and similarity functions $g_1, \ldots, g_N$ such that, for each n, every agent in n's group uses similarity function g_n. As before, each agent's action in each period is given by (4) and the market price in period k is the population-wide average action given by (5).

We show in the next subsection that Proposition 1 carries over directly to this setting except that the modified process m_Δ is defined more generally by

$$m_\Delta(\omega, \omega') = \sum_{n=1}^{N} \pi_n \frac{\sum_{\omega''} g_n(\omega, \omega'')\phi(\omega'')q_\Delta(\omega'', \omega')}{\sum_{\omega''} g_n(\omega, \omega'')\phi(\omega'')}. \tag{9}$$

Within each group, in the steady-state price forecasts, the weight given to each possible state ω' one period ahead is based on the likelihood of transitions to ω' from each state ω'' similar to the current state ω. The weight given to the transition from ω'' to ω' depends on the similarity between ω and ω'' together with the frequency $\phi(\omega'')$ with which state ω'' occurs. The aggregate distribution m_Δ is obtained by averaging the individual distributions across all agents. Note that, as before, agents should be interpreted as behaving, in the long-run, *as if* they believe (on average) that the state evolves according to m_Δ; agents do not literally hold these beliefs.

A.2 Proof of Proposition 1

The general learning process is as follows. The state space Ω and the true process q are as in the main text. Without loss of generality, we focus on the case $\Delta = 1$. Let $\boldsymbol{\omega}^k = (\omega_s)_{s=0}^k$ denote the finite history of states up to period k, and $\mathbf{p}^k = (p_s)_{s=0}^{k-1}$ be the history of prices up to period $k-1$. We assume that all prices lie in a bounded interval $[\underline{p}, \overline{p}]$. The price p_k in period k is determined according to

$$p_k = (1-\rho)d(\omega_k) + \rho Q\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right), \tag{10}$$

where $Q : \bigcup_k \left(\Omega^k \times [\underline{p}, \overline{p}]^{k-1}\right) \longrightarrow [\underline{p}, \overline{p}]$ can be interpreted as the average forecast of the price in period $k+1$ and $\rho = e^{-1}$ is the discount factor.

We assume that Q satisfies the following condition.

A1. There exists a continuous monotone function

$$\mathcal{E} : [\underline{p}, \overline{p}]^\Omega \longrightarrow [\underline{p}, \overline{p}]^\Omega$$

such that, for any $\underline{P}, \overline{P} \in [\underline{p}, \overline{p}]^{\Omega}$, any K, and any $\varepsilon > 0$, if

$$\Pr\left(p_k \in [\underline{P}, \overline{P}]\; \forall k > K\right) > 1 - \varepsilon, \tag{11}$$

then for any $\delta > 0$ there exists K' such that, for each ω,

$$\Pr\left(Q\left(\left(\boldsymbol{\omega}^k, \omega\right), \mathbf{p}^k\right) \in \left(\mathcal{E}(\underline{P})(\omega) - \delta, \mathcal{E}(\overline{P})(\omega) + \delta\right)\; \forall k > K'\right) > 1 - \varepsilon - \delta. \tag{12}$$

In the case of a homogeneous population using similarity function g, the learning process from Section A.1 (and hence also the categorization-based learning from Section 4) is captured by

$$Q^{\text{sim}}\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) = \begin{cases} \frac{\sum_{s<k-1} g(\omega_k, \omega_s) p_{s+1}}{\sum_{s<k-1} g(\omega_k, \omega_s)} & \text{if } \sum_{s<k-1} g(\omega_k, \omega_s) > 0, \\ p_0 & \text{otherwise,} \end{cases}$$

where p_0 is arbitrary. For a heterogeneous population, Q is obtained by aggregating the values of Q^{sim} across groups (see Lemma 2).

Lemma 1. *For any similarity function g, Q^{sim} satisfies A1 with*

$$\mathcal{E}(P)(\omega) = \sum_{\omega'} m_{\Delta}(\omega, \omega') P(\omega'),$$

where m_{Δ} is the modified process in (9).

Proof. We prove only the upper bound; the proof for the lower bound is similar. Suppose that for some K, $\varepsilon > 0$, and $\overline{P}$, $\Pr\left(p_k \leq \overline{P}\; \forall k > K\right) > 1 - \varepsilon$. Given any $\delta > 0$ and $\gamma > 0$, by the Law of Large Numbers, there exists some $K' > K$ such that, with probability greater than $1 - \delta$, for every pair (ω', ω'') and every $k > K'$, the fraction of periods $s < k$ such that $(\omega_s, \omega_{s+1}) = (\omega', \omega'')$ lies in $(\phi(\omega') q_{\Delta}(\omega', \omega'') - \gamma, \phi(\omega') q_{\Delta}(\omega', \omega'') + \gamma)$. Since the process q_{Δ} is ergodic, we can choose K' such that this property holds regardless of the history $\boldsymbol{\omega}^K$. Furthermore, for $K' > K/\gamma$, $p_s \leq \overline{P}(\omega_s)$ for a fraction of at least $1 - \gamma$ periods $s \leq k$ with probability greater than $1 - \varepsilon$, in which case the average of the prices p_{s+1} across those periods s such that $(\omega_s, \omega_{s+1}) = (\omega', \omega'')$ is at most

$(1-\gamma)\overline{P}(\omega'') + \gamma\overline{p}$. Hence for $k > K'$, we have

$$Q^{\text{sim}}\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) \leq \frac{\sum_{\omega',\omega''} g(\omega_k, \omega'')(\phi(\omega'')q_\Delta(\omega'', \omega') + \gamma)((1-\gamma)\overline{P}(\omega') + \gamma\overline{p})}{\sum_{\omega''} g(\omega_k, \omega'')(\phi(\omega'') - \gamma)}$$

with probability greater than $1-\varepsilon-\delta$. Given $\delta > 0$, we can choose $\gamma > 0$ sufficiently small so that the right-hand side of the preceding inequality is less than $\mathcal{E}(\overline{P})(\omega_k) + \delta$, as needed. □

In the main text, agents differ in their forecasting procedures and the price is determined by the average of agents' forecasts. The following lemma indicates that A1 aggregates across heterogeneous groups.

Lemma 2. *Suppose that a fraction π_n of the population use prediction rule Q^n, with $\sum_{n=1}^N \pi_n = 1$. Suppose moreover that all rules Q^n satisfy A1 with functions $\mathcal{E}^n(P)$ respectively. Finally assume that price evolution is governed by* (10) *with prediction rule $Q = \sum \pi_n Q^n$. Then Q satisfies A1 with $\mathcal{E}(P) = \sum_n \pi_n \mathcal{E}^n(P)$.*

Proof. Using the property of A1 with $\pi_n\delta$ for each subpopulation and taking the maximum of the K' needed for each process gives the result. □

Proposition 1 is a special case of the following convergence result.

Proposition 5. *If Q satisfies A1, prices are determined according to* (10)*, and the mapping $(1-\rho)d + \rho\mathcal{E}$ is a contraction (with respect to some metric) then prices almost surely converge to the unique fixed point of $d + \rho\mathcal{E}$.*

In the case of learning by similarity, $(1-\rho)d+\rho\mathcal{E}$ is a contraction with respect to the sup norm, and we therefore obtain convergence to a unique price profile, proving Proposition 1.

Proof of Proposition 5. The mapping $(1-\rho)d+\rho\mathcal{E}$ has extreme fixed points $\underline{P}^*$, $\overline{P}^*$: for every fixed point P^*, we have $\underline{P}^* \leq P^* \leq \overline{P}^*$. This follows immediately from Tarski's Fixed Point Theorem since $[\underline{p}, \overline{p}]^\Omega$ is a complete lattice and $(1-\rho)d + \rho\mathcal{E}$ is continuous and monotone.

We will prove that for each ω, the set of cluster points of $(p_k(\omega))_k$ is almost surely contained in $[\underline{P}^*(\omega), \overline{P}^*(\omega)]$. The proposition follows immediately since the fixed point is unique when $(1-\rho)d + \rho\mathcal{E}$ is a contraction.

We prove only that the cluster points are almost surely at most $\overline{P}^*(\omega)$. The proof of the lower bound is similar.

Let $\overline{P}_0 = \overline{p}\mathbf{1}$, where $\mathbf{1}$ denotes the vector with a 1 in each component, and for $l \in \mathbb{N}_+$, let $\overline{P}_l = (1-\rho)d + \rho\mathcal{E}(\overline{P}_{l-1})$. Since $\overline{P}_l$ is nonincreasing in l, $\lim_l \overline{P}_l$ exists and is a fixed point of $(1-\rho)d + \rho\mathcal{E}$ (by continuity of $\mathcal{E}$).

Note that $p_k \leq \overline{P}_0(\omega_k)$ for each $k > 0$. Suppose for induction that, given any $\varepsilon > 0$, there exists K_l such that

$$\Pr\left(p_k < \overline{P}_l(\omega_k) + \varepsilon \text{ for all } k > K_l\right) > 1 - \varepsilon.$$

We will show that the same condition holds when each l is replaced with $l+1$.

For any $\delta > 0$, combining A1 with the inductive hypothesis, there exists some K_{l+1} such that

$$\Pr\left(Q\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) < \mathcal{E}(\overline{P}_l + \varepsilon\mathbf{1})(\omega_k) + \delta \; \forall k > K_{l+1}\right) > 1 - \varepsilon - \delta.$$

Substituting for $Q\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right)$ using (10), we have

$$\Pr\left(p_k < (1-\rho)d(\omega_k) + \rho\mathcal{E}(\overline{P}_l + \varepsilon\mathbf{1})(\omega_k) + \delta \; \forall k > K_{l+1}\right) > 1 - \varepsilon - \delta.$$

Given any $\gamma > 0$, since $\mathcal{E}$ is continuous, there exist some $\varepsilon, \delta \in (0, \gamma)$ such that, for each ω, $\rho\mathcal{E}(\overline{P}_l + \varepsilon\mathbf{1})(\omega) + \delta < \rho\mathcal{E}(\overline{P}_l)(\omega) + \gamma$. Since ε and δ are arbitrary, we have that, for some K_{l+1},

$$\Pr\left(p_k < (1-\rho)d(\omega_k) + \rho\mathcal{E}(\overline{P}_l)(\omega_k) + \gamma \; \forall k > K_{l+1}\right) > 1 - \gamma.$$

Since $\overline{P}_{l+1} = (1-\rho)d + \rho\mathcal{E}(\overline{P}_l)$, this completes the proof of the inductive step. □

A.3 Presence of rational agents

We now consider a setting in which some agents form rational expectations. For simplicity, we assume that the population consists of two parts. A fraction π of agents are rational while the remaining $1-\pi$ are coarse thinkers who use a prediction rule Q^C satisfying A1. Rational agents know Q^C and the underlying Markov process, and form rational expectations of the forecasts formed by coarse thinkers in the next period. The rational agents' prediction rule Q^R satisfies

$$Q^R\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) = E\left[(1-\rho)d(\omega_{k+1}) + \rho\left((1-\pi)Q^C\left(\boldsymbol{\omega}^{k+1}, \mathbf{p}^{k+1}\right) + \pi Q^R\left(\boldsymbol{\omega}^{k+1}, \mathbf{p}^{k+1}\right)\right)\Big|\,\boldsymbol{\omega}^k\right]. \tag{13}$$

This equation implies that rational agents correctly predict prices given the history to date and the prediction rules used by other agents.

While the model in the main text does not include rational agents, it does allow for some agents to perfectly distinguish among states. These agents are not rational insofar as their price forecasts are based only on past data and do not explicitly account for other agents' forecasts. We show here that, in the long-run, the difference between these agents and rational agents is immaterial. Long-run prices are identical if we replace any share of agents using the finest categorization with agents who form rational expectations.

A2. For each $\omega \in \Omega$, $K \in \mathbb{N}$, and almost every $\boldsymbol{\omega} \in \Omega^{\mathbb{N}}$,

$$\lim_{k\to\infty}\left(Q^C\left(\left(\boldsymbol{\omega}^k, \omega\right), \mathbf{p}^k\right) - Q^C\left(\left(\boldsymbol{\omega}^{k+K}, \omega\right), \mathbf{p}^{k+K}\right)\right) = 0,$$

where, for each κ, $\boldsymbol{\omega}^\kappa$ denotes the projection of $\boldsymbol{\omega}$ onto its first κ components.

Roughly speaking, A2 says that data from a fixed finite number of recent periods eventually has little impact on forecasts once the total quantity of data is large. Note that A2 is satisfied by the similarity-based learning procedure of Section A.1.

Proposition 6. *Suppose that a fraction π of the population form rational expectations, and the remaining $1-\pi$ use a prediction procedure Q^C satisfying A1 with bound $\mathcal{E}^C(P)$ and A2. Suppose further that the mapping $(1-\rho)d + \rho\mathcal{E}^C(P)$ is a contraction. Then the price vector $P(\omega)$ almost surely converges to the unique solution of*

$$P(\omega_k) = (1-\rho)d(\omega_k) + \rho\left(\pi E\left[P\left(\omega_{k+1}\right) \mid \omega_k\right] + (1-\pi)\mathcal{E}^C(P)(\omega_k)\right).$$

Lemma 3. *If Q^C satisfies A1 with bound $\mathcal{E}^C$ and A2 then Q^R satisfies A1 with bound*

$$\mathcal{E}^R(P)(\omega_k) = E\left[\sum_{l=1}^{\infty}(\pi\rho)^{l-1}(1-\rho)d(\omega_{k+l}) + (1-\pi)\rho\sum_{l=1}^{\infty}(\pi\rho)^{l-1}\mathcal{E}^C(P)(\omega_{k+l})\Bigg|\,\omega_k\right]. \tag{14}$$

Proof of Lemma 3. Iterating (13) gives

$$Q^R\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) = E\left[\sum_{l=1}^{\infty}(\pi\rho)^{l-1}(1-\rho)d(\omega_{k+l}) + (1-\pi)\rho\sum_{l=1}^{\infty}(\pi\rho)^{l-1}Q^C\left(\boldsymbol{\omega}^{k+l}, \mathbf{p}^{k+l}\right)\middle|\,\boldsymbol{\omega}^k\right].$$

We need to show that for any $\underline{P}, \overline{P} \in [\underline{p}, \overline{p}]^{\Omega}$, any K, and any $\varepsilon > 0$, if condition (11) holds, then for any $\delta > 0$ there exists K' such that (12) holds for Q^R. We prove only the upper bound; the proof of the lower bound is similar.

Accordingly, suppose that (11) holds for some $\varepsilon > 0$ and K. Fix $\delta > 0$. Since Q^C and $\mathcal{E}^C$ are bounded, there exists M such that, for every $\boldsymbol{\omega}^k$ and $\mathbf{p}^k$

$$Q^R\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) < E\left[\sum_{l=1}^{\infty}(\pi\rho)^{l-1}(1-\rho)d(\omega_{k+l}) + (1-\pi)\rho\left(\sum_{l=1}^{M}(\pi\rho)^{l-1}Q^C\left(\boldsymbol{\omega}^{k+l}, \mathbf{p}^{k+l}\right) + \sum_{l=M+1}^{\infty}(\pi\rho)^{l-1}\mathcal{E}^C(\overline{P})(\omega_{k+l})\right)\middle|\,\boldsymbol{\omega}^k\right] + \delta/3. \quad (15)$$

Since Q^C satisfies A1, there exists some K' such that, for each ω,

$$\Pr\left(Q^C\left(\left(\boldsymbol{\omega}^{k-1}, \omega\right), \mathbf{p}^k\right) < \mathcal{E}^C(\overline{P})(\omega) + \delta/3M \;\forall k > K'\right) > 1 - \varepsilon - \delta/2. \quad (16)$$

By A2, there exists some K'' such that, for each $l = 1, \ldots, M$,

$$\Pr\left(Q^C\left(\boldsymbol{\omega}^{k+l}, \mathbf{p}^{k+l}\right) < Q^C\left(\left(\boldsymbol{\omega}^{k-1}, \omega_{k+l}\right), \mathbf{p}^k\right) + \delta/3M \;\forall k > K'' \mid \boldsymbol{\omega}^k\right) > 1 - \delta/2M. \quad (17)$$

Combining (16) and (17) gives

$$\Pr\left(Q^C\left(\boldsymbol{\omega}^{k+l}, \mathbf{p}^{k+l}\right) < \mathcal{E}^C(\overline{P})(\omega_{k+l}) + 2\delta/3M \;\forall k > \max\{K', K''\}, \forall l = 1, \ldots, M \mid \boldsymbol{\omega}^k\right) > 1-\varepsilon-\delta.$$

Combining the last inequality with (15) gives

$$\Pr\left(Q^R\left(\boldsymbol{\omega}^k, \mathbf{p}^k\right) < \mathcal{E}^R(\overline{P})(\omega_k) + \delta \;\forall k > \max\{K', K''\}\right) > 1 - \varepsilon - \delta,$$

as needed. □

Proof of Proposition 6. Combining Lemma 2, Lemma 3, and Proposition 5, the cluster points lie between the extremal solutions to

$$P = (1-\rho)d + \rho\left(\pi\mathcal{E}^R(P) + (1-\pi)\mathcal{E}^C(P)\right).$$

Substituting for $\mathcal{E}^R(P)$ using (14) leads to

$$\begin{aligned}
P(\omega_k) &= (1-\rho)d(\omega_k) + \rho\pi E\left[\sum_{l=1}^{\infty}(\pi\rho)^{l-1}(1-\rho)d(\omega_{k+l}) + (1-\pi)\rho\sum_{l=1}^{\infty}(\pi\rho)^{l-1}\mathcal{E}^C(P)(\omega_{k+l})\middle|\,\omega_k\right] \\
&\quad + \rho(1-\pi)\mathcal{E}^C(P)(\omega_k) \\
&= (1-\rho)d(\omega_k) + \rho\pi E\left[E\left[\sum_{l=1}^{\infty}(\pi\rho)^{l-1}(1-\rho)d(\omega_{k+l}) + (1-\pi)\rho\sum_{l=1}^{\infty}(\pi\rho)^{l-1}\mathcal{E}^C(P)(\omega_{k+l})\middle|\,\omega_{k+1}\right]\middle|\,\omega_k\right] \\
&\quad + \rho(1-\pi)\mathcal{E}^C(P)(\omega_k) \\
&= (1-\rho)d(\omega_k) + \rho\left(\pi E[P(\omega_{k+1})|\omega_k] + (1-\pi)\mathcal{E}^C(P)(\omega_k)\right),
\end{aligned}$$

where the second last equality follows from the Law of Iterated Expectations, and the final equality uses the first equality with ω_{k+1} in place of ω_k. □

B Proof of Theorem 2 and Proposition 3

We provide a unified proof of Theorem 2 and Proposition 3 that also covers the setting from Appendix A.1. In each case, we let $m_\Delta(\omega,\omega')$ denote the population-average belief about transition probabilities. The functional form of $m_\Delta(\omega,\omega')$ for the setting from Appendix A.1 (that covers the categorization from Section 4) is given by (9), while for the setting from Section 5, m_Δ is given by (7) and (8).

Define a binary relation R on Ω by $\omega R\omega'$ if $\lim_{\Delta\to 0} m_\Delta(\omega,\omega')/\Delta = \infty$. Let $\sim$ denote the transitive closure of R, and note that $\sim$ is reflexive and symmetric.[4] Then $\sim$ is an equivalence relation and hence corresponds to a partition of Ω into *aggregate categories*. Let $\mathcal{C}$ denote this partition, with a typical element $C \subset \Omega$. We write $C(\omega)$ to denote the aggregate category in $\mathcal{C}$ containing ω.

[4]For the setting from Appendix A.1, symmetry of $\sim$ follows from the symmetry of the similarity functions. Symmetry of $\sim$ is trivial for the setting from Section 5 since $\omega \sim \omega'$ for all $\omega,\omega' \in \Omega$.

For the setting from Appendix A.1, two states ω and ω' lie in the same aggregate category if and only if there exists a sequence $\omega_1, \ldots, \omega_r$ of states such that $\omega = \omega_1$, $\omega' = \omega_r$, and for each $\ell = 1, \ldots, r-1$, $g_n(\omega_{\ell+1}, \omega_\ell) > 0$ for some group $n \in \{1, \ldots, N\}$. Aggregate categorization is trivial for the setting from Section 5; all states $\omega \in \Omega$ lie in the same aggregate category.

Given an aggregate category C, for each Δ let $\tilde{m}_\Delta$ denote the transition probabilities of the restriction of m_Δ to C, that is, the probabilities defined on $C \times C$ obtained by conditioning $m_\Delta(\omega, \omega')$ on $\omega' \in C$. Let ϕ_C^Δ denote the stationary distribution of $\tilde{m}_\Delta$, and let $\phi_C = \lim_{\Delta \to 0} \phi_C^\Delta$. Note that for the setting from Section 5, $\tilde{m}_\Delta = m_\Delta$ since there is just one aggregate category. Let $\overline{d}(C) = E_{\phi_C}[d(\omega)]$, and let $\overline{m}$ be the Markov process on $\mathcal{C}$ with transition probabilities

$$\overline{m}_\Delta(C, C') = \sum_{\omega \in C, \omega' \in C'} \phi_C(\omega) m_\Delta(\omega, \omega').$$

In the special case of coarse categorization from Section 4, ϕ_C is simply the stationary distribution of the true process q restricted to C.[5] Thus the process $\overline{m}$ is is identical to the coarse process $\overline{q}$ defined in Section 4.

The following Theorem unifies Proposition 3 and Theorem 2, and extends the latter to the setting from Appendix A.1. For each Δ, let $P_\Delta \in \mathbb{R}^\Omega$ be rational expectations prices with respect to m_Δ and d.

Theorem 7. *For any $\omega, \omega' \in \Omega$ such that $C(\omega) = C(\omega')$, we have*

$$\lim_{\Delta \to 0} P_\Delta(\omega) - P_\Delta(\omega') = 0.$$

Moreover, prices are given by the rational expectations prices associated with the process $\overline{m}_\Delta(C, C')$ and dividends $\overline{d}(C)$ in the limit as $\Delta \to 0$.

Lemma 4. *There exists $K(\eta, \Delta)$ such that, for each $\eta > 0$,*

[5] To see this, note that for each individual i categorizing ω and ω' together, the limiting transition probability $m_\Delta^i(\omega, \omega')$ from ω to ω' is proportional to the stationary distribution mass assigned to ω'. Hence the stationary distribution of q restricted to C is also stationary with respect to each individual belief $m_\Delta^i(\omega, \omega')$. Aggregating across individuals gives the claim.

1. for each $\Delta > 0$ and $\omega \in C$,

$$\frac{1}{K(\eta,\Delta)} \sum_{k=0}^{K(\eta,\Delta)-1} \|\tilde{m}_\Delta^k(\omega,\cdot) - \phi_\Delta^C\| < \eta,$$

where $\|\cdot\|$ is the 1-norm and $\tilde{m}_\Delta^k$ are the transition probabilities for k steps of $\tilde{m}_\Delta$; and

2. $K(\eta,\Delta)\Delta \to 0$ as $\Delta \to 0$.

Proof. We claim that, for each $\eta > 0$ and $\omega \in C$, there exists K_0 such that

$$\|\tilde{m}_\Delta^k(\omega,\cdot) - \phi_\Delta^C\| < \eta/2$$

for every $k \geq K_0$, and $K_0\Delta \to 0$ as $\Delta \to 0$. Since $\|\tilde{m}_\Delta^k(\omega,\cdot) - \phi_\Delta^C\| \leq 2$ for every k, taking $K(\eta,\Delta) = 4K_0/\eta$ proves the result.

We will show that there exists ε_Δ such that (i)

$$\|\tilde{m}_\Delta^k(\omega,\cdot) - \phi_\Delta^C\| \leq 2(1-\varepsilon_\Delta)^{k-1} \tag{18}$$

for every k and ω, and (ii) $\lim_{\Delta\to 0} \varepsilon_\Delta/\Delta = \infty$. Then, letting

$$K_0(\eta,\Delta) = 2 + \frac{\log(\eta/4)}{\log(1-\varepsilon_\Delta)},$$

straightforward algebraic manipulation shows that $2(1-\varepsilon_\Delta)^{k-1} < \eta/2$ for every $k \geq K_0$, as needed. Moreover, $K_0(\eta,\Delta)\Delta \to 0$ as $\Delta \to 0$ since $\lim_{\Delta\to 0} \Delta/\log(1-\varepsilon_\Delta) \to 0$ by (ii).

Existence of ε_Δ satisfying (i) and (ii) follows from Corollary 1.2 of Hartfiel (1998). The corollary implies that if there exist $\delta_\Delta \geq 0$ and L such that $\tilde{m}_\Delta^L(\omega,\omega') \geq \delta_\Delta$ for all ω and ω', then

$$\|\tilde{m}_\Delta^k(\omega,\cdot) - \phi_\Delta^C\| \leq 2(1-\delta_\Delta)^{\frac{k}{L}-1}$$

for every $k > 0$.

For the setting from Section 5, (18) follows by taking $L = 1$, $\delta_\Delta = \min_{\omega,\omega'} \tilde{m}_\Delta(\omega,\omega')$, and $\varepsilon_\Delta = \delta_\Delta$.

For the setting from Section 4, (18) follows by taking $L = |C|$. Notice that $\tilde{m}_\Delta^{|C|}(\omega,\omega')$ is bounded from below by a constant δ independent of Δ. Thus we can choose ε_Δ to be $1-(1-\delta)^{\frac{1}{L}}$.

$\square$

Proof of Theorem 7. Let $\omega \in C$ and rewrite the equation

$$P(\omega) = d(\omega)(1-e^{-\Delta}) + e^{-\Delta} E_{m_\Delta(\omega,\omega')}[P(\omega')]$$

defining rational expectations prices as

$$P(\omega) = d(\omega)(1-e^{-\Delta}) + e^{-\Delta} \underbrace{E_{m_\Delta(\omega,\omega')}[P(\omega')|\omega' \notin C]}_{f(\omega,\Delta)} \underbrace{\Pr_{m_\Delta(\omega,\omega')}[\omega' \notin C]}_{\varepsilon_\omega \Delta + O(\Delta^2)}$$
$$+ e^{-\Delta} \underbrace{(1-\Pr_{m_\Delta(\omega,\omega')}[\omega' \notin C])}_{1-\varepsilon_\omega\Delta+O(\Delta^2)} E_{\tilde{m}_\Delta(\omega,\omega')}[P(\omega')]$$

where $\varepsilon_\omega = \lim_{\Delta\to 0} \Pr_{m_\Delta(\omega,\omega')}[\omega' \notin C]/\Delta$.[6]

Using the approximation $d(\omega)(1-e^{-\Delta}) = d(\omega)\Delta + O(\Delta^2)$, the last equation can be rewritten as

$$P(\omega) = d(\omega)\Delta + f(\omega,\Delta)\varepsilon_\omega\Delta + e^{-\Delta}(1-\varepsilon_\omega\Delta)E_{\tilde{m}_\Delta(\omega,\omega')}[P(\omega')] + O(\Delta^2).$$

This can be interpreted as the pricing equation of a process in which the asset pays a dividend $d(\omega)\Delta$, with some probability $\varepsilon_\omega\Delta$ the process terminates giving a final payoff $f(\omega,\Delta)$, and with the remaining probability $1-\varepsilon_\omega\Delta$ the process continues to the next trading period, in which the state will be $\omega' \in C$.

Iterating the last equation for K periods gives

$$P(\omega) = \sum_{k=0}^{K-1} e^{-k\Delta} E\left[\left(\prod_{k'=0}^{k-1}(1-\varepsilon_{\omega_{k'}}\Delta)\right)(d(\omega_k)\Delta + f(\omega_k,\Delta)\varepsilon_{\omega_k}\Delta)\right]$$
$$+ e^{-K\Delta} E\left[\left(\prod_{k'=0}^{K-1}(1-\varepsilon_{\omega_{k'}}\Delta)\right) P(\omega_K)\right] + O(K\Delta^2),$$

where ω_k is the state in period k of the Markov process $\tilde{m}_\Delta$ on C starting from $\omega_0 = \omega$. For any

[6] The fact that $\Pr_{m_\Delta(\omega,\omega')}[\omega' \notin C] = \varepsilon_\omega\Delta + O(\Delta^2)$ holds because $m_\Delta(\omega,\omega') = q_\Delta(\omega,\omega')$ whenever ω and ω' lie in different aggregate categories.

K, we may rewrite the last equation as

$$\frac{1}{K}P(\omega) = \frac{1}{K}\sum_{k=0}^{K-1} E\left[\left(d(\omega_k) + f(\omega_k,\Delta)\varepsilon_{\omega_k}\right)\Delta\right] + \frac{1}{K}(1-K\Delta)E\left[\left(1-\sum_{k'=0}^{K-1}\varepsilon_{\omega_{k'}}\Delta\right)P(\omega_K)\right] + O(K\Delta^2). \quad (19)$$

Taking $K = K(\eta,\Delta)$ from Lemma 4 and K_0 as in the proof of the lemma, we have

$$\begin{aligned} E\left[\left(1-\sum_{k'=0}^{K-1}\varepsilon_{\omega_{k'}}\Delta\right)P(\omega_K)\right] &= E\left[\left(1-\sum_{k'=0}^{K-K_0-1}\varepsilon_{\omega_{k'}}\Delta - \sum_{k'=K-K_0}^{K-1}\varepsilon_{\omega_{k'}}\Delta\right)P(\omega_K)\right] \\ &= E\left[\left(1-\sum_{k'=0}^{K-K_0-1}\varepsilon_{\omega_{k'}}\Delta\right)E\left[P(\omega_K)\mid\omega_{K-K_0}\right]\right] + O(\eta K\Delta) \\ &= E\left[\left(1-\sum_{k'=0}^{K-1}\varepsilon_{\omega_{k'}}\Delta\right)E\left[P(\omega_K)\mid\omega_{K-K_0}\right]\right] + O(\eta K\Delta). \end{aligned}$$

Substituting into (19) and applying Lemma 4 gives

$$\frac{1}{K(\eta,\Delta)}P(\omega) = \left(d^C + [f\varepsilon]^C\right)\Delta + (1-K(\eta,\Delta)\Delta)\left(1-K(\eta,\Delta)\varepsilon^C\Delta\right)\frac{1}{K(\eta,\Delta)}P^C + O(\eta/K(\eta,\Delta) + \eta\Delta + K(\eta,\Delta)\Delta^2), \quad (20)$$

where x^C denotes the average of $x(\omega)$ with respect to the stationary distribution of the process $\tilde{m}_\Delta$.[7] Rearranging yields

$$P(\omega) - P^C = \left(d^C + [f\varepsilon]^C\right)K(\eta,\Delta)\Delta + \left(-K(\eta,\Delta) - K(\eta,\Delta)\varepsilon^C + \varepsilon^C K(\eta,\Delta)^2\Delta\right)\Delta P^C + O(\eta + K(\eta,\Delta)\eta\Delta + K(\eta,\Delta)^2\Delta^2).$$

Since $K(\eta,\Delta)\Delta \to 0$ as $\Delta \to 0$, this last equation implies that $\lim_{\Delta\to 0}\left(P(\omega) - P^C\right) = 0$. More

[7] Note that, although it is omitted from the notation, each of these averages depends on Δ.

precisely,

$$\begin{aligned} P(\omega) - P^C &= O(K(\eta,\Delta)\Delta + \eta + K(\eta,\Delta)\eta\Delta + K(\eta,\Delta)^2\Delta^2) \\ &= O(K(\eta,\Delta)\Delta + \eta). \end{aligned}$$

Taking the average of (20) with respect to the stationary distribution of $\tilde{m}_\Delta$ (which amounts to replacing $P(\omega)$ with P^C), dropping a term of order $K(\eta,\Delta)\Delta^2$, and simplifying leads to

$$\frac{1}{K(\eta,\Delta)}\left(K(\eta,\Delta)\Delta + K(\eta,\Delta)\varepsilon^C\Delta\right)P^C = \left(d^C + [f\varepsilon]^C\right)\Delta + O(\eta/K(\eta,\Delta) + \eta\Delta + K(\eta,\Delta)\Delta^2),$$

and hence

$$P^C = \frac{d^C + [f\varepsilon]^C}{1+\varepsilon^C} + O\left(\frac{\eta}{\Delta K(\eta,\Delta)} + \eta + K(\eta,\Delta)\Delta\right).$$

Recall that, as $\Delta \to 0$, the stationary distribution of $\tilde{m}_\Delta$ approaches ϕ_C. It suffices to show that there exists $\Delta(\eta)$ such that $\eta/\left(\Delta(\eta)K\left(\eta,\Delta(\eta)\right)\right)$ and $K\left(\eta,\Delta(\eta)\right)\Delta(\eta)$ vanish as $\eta \to 0$. Given $a, b \in (0,1)$ such that $a < b$, it suffices to take $\Delta(\eta)$ such that $\eta^b < \Delta(\eta)K\left(\eta,\Delta(\eta)\right) < \eta^a$. By Lemma 4, the upper bound is satisfied for sufficiently small Δ. If the lower bound is not satisfied for any $\Delta > 0$ then we can simply replace $K(\eta,\Delta)$ with a larger value for a particular Δ in order to satisfy both bounds. □

C Proof of Proposition 4

Proof. The first-order conditions for the demands $\alpha^G(\omega)$ of each group G are

$$\sum_{\omega'} q^G(\omega,\omega')e^{-\alpha^G(\omega)e^{-\Delta}P(\omega')}E_G\left[e^{-w''}|\omega'\right]\left(-P(\omega) + \left(1-e^{-\Delta}\right)d(\omega) + e^{-\Delta}P(\omega')\right) = 0,$$

where $w'' = \sum_{s\in\{k+1,\dots,k_0+n-1\}} e^{-(s-k_0)\Delta}\alpha^G(\omega_s)\left(-P(\omega_s) + \left(1-e^{-\Delta}\right)d(\omega_s) + e^{-\Delta}P(\omega_{s+1})\right)$ is the agent's future gain.

Let $\hat{q}^G(\omega,\omega') = q^G(\omega,\omega')e^{-\alpha^G(\omega)e^{-\Delta}P(\omega')}E_G\left[e^{-w''}|\omega'\right]$, and notice that, for both values of ω, the vectors $\left(\hat{q}^C(\omega,\omega')\right)_{\omega'}$ and $\left(\hat{q}^R(\omega,\omega')\right)_{\omega'}$ are collinear. This follows from the first-order conditions, which indicate that these 2-dimensional vectors are both perpendicular to the vector $\left(-P(\omega)+\left(1-e^{-\Delta}\right)d(\omega)+e^{-\Delta}P(\omega')\right)_{\omega'}$. Therefore, there exist $\gamma_\omega^G(\Delta)$ and a virtual process $\hat{q}_\Delta(\omega,\omega')$ such that

$$\gamma_\omega^G \hat{q}^G(\omega,\omega') = \hat{q}(\omega,\omega') \text{ for each } G,$$

and $P(\omega)$ are rational expectation prices with respect to $\hat{q}(\omega,\omega')$, that is,

$$\sum_{\omega'} \hat{q}(\omega,\omega')\left(-P(\omega)+\left(1-e^{-\Delta}\right)d(\omega)+e^{-\Delta}P(\omega')\right)=0.$$

The virtual process $\hat{q}(\omega,\omega')$ can be partially characterized as follows (recall that the group populations π_G sum up to 1):

$$\begin{aligned}
\hat{q}(\omega,\omega') &= \prod_G \left[\hat{q}(\omega,\omega')\right]^{\pi_G} \\
&= \prod_G \left[\gamma_\omega^G \hat{q}^G(\omega,\omega')\right]^{\pi_G} \\
&= \prod_G \left[\gamma_\omega^G\right]^{\pi_G} \prod_G \left[q^G(\omega,\omega')\right]^{\pi_G} \prod_G \left[e^{-\alpha^G(\omega)e^{-\Delta}P(\omega')}\right]^{\pi_G} \prod_G \left(E_G\left[e^{-w''}|\omega'\right]\right)^{\pi_G}.
\end{aligned}$$

The market clearing condition $\sum_G \pi_G \alpha^G = 0$ implies that $\prod_G \left[e^{-\alpha^G(\omega)e^{-\Delta}P(\omega')}\right]^{\pi_G} = 1$. To simplify notation, let $x_{\omega'} = \prod_G \left(E_G\left[e^{-w''}|\omega'\right]\right)^{\pi_G}$ and $\gamma_\omega = \prod_G \left[\gamma_\omega^G\right]^{\pi_G}$. We have shown that

$$\hat{q}(\omega,\omega') = \gamma_\omega x_{\omega'} \prod_G \left[q^G(\omega,\omega')\right]^{\pi_G}. \tag{21}$$

From the equations for rational expectations prices with respect to $\hat{q}$, we obtain

$$P(1)-P(0) = \frac{d(1)-d(0)}{1+\frac{e^{-\Delta}}{1-e^{-\Delta}}(\hat{q}(1,0)+\hat{q}(0,1))}.$$

To show the convergence of prices, it suffices to show that $\frac{\hat{q}(1,0)+\hat{q}(0,1)}{\Delta}$ diverges as Δ vanishes (recall that $\hat{q}$ depends on Δ).

Let

$$\hat{\varphi}(\omega,\omega') = \frac{\hat{q}(\omega,\omega')}{\hat{q}(\omega,\omega)}$$

for $\omega \neq \omega'$. We need to show that

$$\frac{\frac{\hat{\varphi}(1,0)}{1+\hat{\varphi}(1,0)} + \frac{\hat{\varphi}(0,1)}{1+\hat{\varphi}(0,1)}}{\Delta} \tag{22}$$

diverges.

By (21), $\hat{\varphi}(1,0) = \varphi(1,0)y$ and $\hat{\varphi}(0,1) = \frac{\varphi(1,0)}{y}$ where

$$\varphi(\omega, \omega') = \frac{\prod_G \left[q^G(\omega, \omega')\right]^{\pi_G}}{\prod_G \left[q^G_{\omega,\omega}\right]^{\pi_G}},$$

and $y = x_0/x_1$. Note that both $\frac{\frac{\varphi(1,0)}{1+\varphi(1,0)}}{\Delta}$ and $\frac{\frac{\varphi(0,1)}{1+\varphi(0,1)}}{\Delta}$ diverge.

Let $\varphi = \min\{\varphi(1,0), \varphi(0,1)\}$. Then (22) is at least $\frac{\frac{\varphi}{1+\varphi}}{\Delta}$, which diverges. □

References

Allen, F., S. Morris, and H. Shin (2006). Beauty contests and iterated expectations in asset markets. *Review of Financial Studies 19*(3), 719–752.

Bacchetta, P. and E. Van Wincoop (2008). Higher order expectations in asset pricing. *Journal of Money, Credit and Banking 40*(5), 837–866.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*(2), 161–199.

Bianchi, M. and P. Jehiel (2010). Bubbles and crashes with partially sophisticated investors. *PSE Working Papers*.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann (1990a). Noise trader risk in financial markets. *Journal of Political Economy 98*(4), 703–738.

De Long, J., A. Shleifer, L. Summers, and R. Waldmann (1990b). Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance 45*(2), 379–395.

Fehr, E. and J. Tyran (2005). Individual irrationality and aggregate outcomes. *Journal of Economic Perspectives 19*(4), 43–66.

Gonzalez, R. and G. Wu (1999). On the shape of the probability weighting function. *Cognitive Psychology 38*(1), 129–166.

Haltiwanger, J. and M. Waldman (1985). Rational expectations and the limits of rationality: An analysis of heterogeneity. *American Economic Review 75*(3), 326–340.

Haltiwanger, J. and M. Waldman (1989). Limited rationality and strategic complements: the implications for macroeconomics. *Quarterly Journal of Economics 104*(3), 463–483.

Hartfiel, D. (1998). *Markov set-chains*. Springer.

Prelec, D. (1998). The probability weighting function. *Econometrica 66*, 497–527.