The Gambler's and Hot-Hand Fallacies in a Dynamic-Inference Model

Matthew Rabin*
Department of Economics
University of California, Berkeley

Dimitri Vayanos†
London School of Economics
CEPR and NBER

December 5, 2005‡

Abstract

We develop a model of belief in the gambler's fallacy, and explore the link with the seemingly opposite hot-hand fallacy. We show that individuals who observe a sequence of signals and are subject to the gambler's fallacy can overstate the time-variation of an underlying state. This can generate belief in the hot hand when the state's persistence is known or when signals are serially correlated. In each case we determine whether the hot hand arises—overtaking the gambler's fallacy—after long or short streaks of signals. Our model provides a flexible framework for studying the gambler's fallacy in economic and finance settings.

* rabin@econ.berkeley.edu
† d.vayanos@lse.ac.uk
‡ We thank Anat Admati, Markus Brunnermeier, Peter DeMarzo, Michael Jansson, Leonid Kogan, Jon Lewellen, Francisco Penaranda, Paul Pfleiderer, Jean Tirole, and seminar participants at LSE and MIT for useful comments. Jeff Holman provided excellent research assistance. Rabin acknowledges financial support from the NSF.

1 Introduction

Many people fall under the spell of the "gambler's fallacy," expecting outcomes in random sequences to exhibit systematic reversals. When observing flips of a fair coin, for example, people believe that a streak of heads makes it more likely that the next flip will be a tail.[1] On the other hand, people also sometimes predict that random sequences will exhibit excessive persistence rather than reversals. While several studies have shown the belief to be fallacious, basketball fans believe that players have significant "hot hands," being more likely to make a shot following a successful streak.[2]

At first blush, the hot-hand fallacy appears to directly contradict the gambler's fallacy because it involves belief in excessive persistence rather than reversals. Several researchers have, however, suggested that the two fallacies might be related, with the hot-hand fallacy arising as a consequence of the gambler's fallacy.[3] Consider an investor who believes that the performance of a mutual fund is a combination of the manager's ability and luck. Convinced that luck should revert, the investor underestimates the likelihood that a manager of average ability will exhibit a streak of above- or below-average performances. Following good or bad streaks, therefore, the investor will over-infer that the current manager is above or below average, and so in turn will predict continuation of unusual performances.

In this paper we examine the relationship between the gambler's and hot-hand fallacies. We show that an individual who is subject to the gambler's fallacy but otherwise updates rationally tends to develop a fallacious belief in the hot hand, confirming the above intuition. At the same time, our model qualifies this intuition. We show that while the individual ends up believing in excessive time-variation of an underlying state—e.g., ability of a fund manager—this belief does not always lead to predictions of excessive persistence for the observable outcomes—e.g., fund returns. Under some conditions, this belief simply offsets the gambler's fallacy, and the individual predicts the outcomes correctly. Under other conditions, however, predictions are incorrect, and we characterize when they involve excessive persistence or reversals. The model and techniques we develop provide a flexible framework for applying the gambler's fallacy to economic and finance settings.

[1]The gambler's fallacy is commonly interpreted as deriving from a fallacious belief in the "law of small numbers," namely that a small sample should resemble closely the underlying population. In coin-flipping experiments, for example, subjects seem to believe that heads and tails should balance even in small samples. Evidence for the gambler's and law-of-small-numbers fallacies comes from experiments where subjects must predict, evaluate, or generate random sequences, as well as from settings outside the laboratory, such as high-stakes lottery play. See Rabin (2002) for a review of some of the evidence. In coining the term "law of small numbers," Tversky and Kahneman (1971) relate it to the broader bias of the representativeness heuristic.

[2]See, for example, Gilovich, Vallone, and Tversky (1985) and Tversky and Gilovich (1989a, 1989b). See also Camerer (1989) who shows that betting markets for basketball games exhibit a small hot-hand bias.

[3]See, for example, Camerer (1989) and Rabin (2002). Rabin also summarizes evidence from Edwards (1961) that suggests a causal link from the gambler's to the hot-hand fallacy.

In Section 2 we present the model. An individual observes a sequence of signals whose probability distribution depends on an underlying state. The signal s_t in Period $t = 1, 2, ..$ is

$$s_t = \theta_t + \epsilon_t,$$

where θ_t is the state and ϵ_t an *i.i.d.* normal shock. For example, the signal can be interpreted as the return on a mutual fund, the state as the ability of the fund manager, and the shock ϵ_t as the manager's luck in period t. We model the gambler's fallacy as the mistaken belief that the sequence $\{\epsilon_t\}_{t\geq 1}$ is not *i.i.d.* but exhibits systematic reversals: according to the individual,

$$\epsilon_t = \tilde{\epsilon}_t - \alpha \sum_{k=0}^{\infty} \delta^k \epsilon_{t-1-k},$$

where the sequence $\{\tilde{\epsilon}_t\}_{t\geq 1}$ is *i.i.d.* and normal, and $\alpha, \delta \in [0, 1)$ are exogenous parameters. The parameter α characterizes the strength of the belief in the gambler's fallacy. When $\alpha = 0$, the individual is an error-free Bayesian. When $\alpha > 0$, however, the individual believes that a high realization of $\epsilon_{t'}$ for $t' < t$ implies a low expected realization of ϵ_t. Thus, the investor believes that if a fund manager was lucky in Period $t' < t$, luck should reverse in Period t. The parameter δ characterizes the memory duration of the gambler's fallacy: the smaller is δ, the more quickly the investor believes luck should reverse. For expositional ease, from now on we follow the convention in the literature of referring to an individual subject to the gambler's fallacy ($\alpha > 0$) as "Freddy," and to a Bayesian ($\alpha = 0$) as "Tommy."

We study Freddy's inference in environments where the state evolves according to the autoregressive process

$$\theta_t = \mu + \rho(\theta_{t-1} - \mu) + \eta_t,$$

where μ is the long-run mean, ρ the persistence parameter, and η_t an *i.i.d.* normal shock. Suppose, for example, that a fund is run by a team of managers. Then, θ_t can be interpreted as the average ability within the team, and could change over time as managers leave or join the fund. The parameter $1 - \rho$ can be interpreted as the rate of managerial turnover, and the variance of η_t as the extent of heterogeneity in managerial ability.

In Section 3 we examine how Freddy uses the sequence of past signals to make inferences about the underlying parameters and to predict future signals. We assume that Freddy infers as a fully rational Bayesian and fully understands the structure of his environment, except for a mistaken and dogmatic belief that $\alpha > 0$. From observing the signals, Freddy infers both the underlying state θ_t and the values of parameters of his model about which he is uncertain. For example, he can learn about a fund-manager's ability (θ_t), and about the extent to which ability changes over time

($\sigma_\eta^2 \equiv Var(\eta_t)$) and is persistent ($\rho$). In fact, the case of parameter uncertainty is central to our theory. For example, if ability is constant—i.e., $\sigma_\eta^2 = 0$—and Freddy knows this, then he cannot develop a belief in the hot hand. But if Freddy initially assigns non-zero probability to $\sigma_\eta^2 > 0$, we show that after a long sequence of signals he always ends up believing falsely that $\sigma_\eta^2 > 0$. Moreover, this belief can lead him to predict excessive persistence of fund returns.

When Freddy is certain about all model parameters, his inference about unobservable variables can be treated using standard tools of recursive (Kalman) filtering, where the gambler's fallacy essentially expands the state vector to include not only the state θ_t but also a statistic of past luck realizations. When Freddy is initially uncertain about parameters, recursive filtering can be used to evaluate the likelihood of signals conditional on parameters. An appropriate version of the law of large numbers then implies that after observing many signals, Freddy converges with probability one to parameter values that maximize a limit likelihood. While the maximum likelihood when $\alpha = 0$ leads Tommy to limit posteriors corresponding to the true parameter values, Freddy's abiding belief that $\alpha > 0$ leads him generally to false limit posteriors. Identifying when and how these limit beliefs are wrong is the crux of our analysis.

In Section 4 we consider the case where the state is constant over time ($\sigma_\eta^2 = 0$), and hence signals are *i.i.d.* If Freddy is initially uncertain about the values of all parameters, he converges to the false belief that $\sigma_\eta^2 > 0$. Freddy ends up believing, however, that the state's time-variation tends to increase streaks of signals in a way that exactly offsets the reversals induced by the gambler's fallacy. As a result, he converges to the correct belief that signals are *i.i.d.* Consider, for example, Freddy's predictions after a high signal. Because he believes that the state has increased, he expects the future signals to be high and to converge back to the mean according to the persistence parameter ρ. At the same time, because he attributes the high signal partly to luck, he expects the future signals to be low and to converge to the mean according to the memory parameter δ. We show that Freddy ends up believing that the persistence is $\tilde{\rho} = \delta - \alpha$, under which the two convergence rates are equal and the effects exactly offset.

Ironically, Freddy cannot develop a false model that offsets his belief in the gambler's fallacy when he knows with certainty the correct values of some parameters. For example, when he knows that $\sigma_\eta^2 = 0$, he clearly cannot develop a belief in the time-varying state, so he always predicts signals under the spell of the gambler's fallacy. This case corresponds to the experiments where subjects know that they are dealing with fair coins.

A more interesting but less straightforward case is when Freddy knows the true value of ρ but is uncertain about σ_η^2. For example, he might observe the turnover of fund managers but be uncertain about whether managers differ in ability. We show that if $\rho > \delta - \alpha$, then the gambler's fallacy dominates for short streaks of signals, but the hot-hand fallacy dominates for longer streaks. That is, Freddy predicts a low signal following a short streak of high signals, but a high signal following a longer streak. Intuitively, a high signal has a long-lasting impact on Freddy's estimate of the state because when ρ is large, Freddy knows the state to be very persistent. Therefore, a long streak of high signals has a large cumulative impact—and this belief in the increased state eventually overtakes the gambler's fallacy, leading Freddy to predict a high signal. This result seems to reflect the intuition of previous researchers that the hot-hand fallacy arises when believers in the gambler's fallacy attempt to rationalize long streaks. Note, however, that the opposite result is possible: if $\rho < \delta - \alpha$, then a long streak of signals leads Freddy to expect a large cumulative reversal of luck. Thus, the intuition is correct in our model only under the (perhaps plausible) assumption that individuals believe the state to be persistent, but expect luck to reverse quickly.

In Section 5 we consider the case where signals are serially correlated. We show that unlike the *i.i.d.* case, Freddy cannot predict the signals correctly even when he is initially uncertain about the values of all parameters. The intuition is that in the *i.i.d.* case, Freddy gets around the gambler's fallacy by taking the state's persistence to be $\tilde{\rho} = \delta - \alpha$ rather than the true value ρ. When the state is time-varying, the persistence parameter influences the signal process, and a discrepancy between $\tilde{\rho}$ and ρ leads Freddy to incorrect predictions.

Freddy's prediction errors turn out not to depend on the comparison between ρ and $\delta - \alpha$: he always under-predicts the next signal after a short streak of high signals, over-predicts after a longer streak, and under-predicts again after a very long streak. The intuition is easier to see when $\rho > \delta - \alpha$, i.e., Freddy expects luck to reverse quickly and yet sees highly persistent signals. To explain the absence of quick reversals, Freddy believes that shocks to the state are large but short-lived. Thus, he underestimates the persistence parameter, converging to a value $\tilde{\rho}$ between $\delta - \alpha$ and ρ. Because $\tilde{\rho} > \delta - \alpha$, Freddy's estimate of the state after a long streak overtakes the gambler's fallacy, which is why the initial under-prediction is followed by over-prediction. The subsequent under-prediction is because Freddy underestimates the state's persistence, and does not update as much as Tommy after a very long streak.[4] In generating over-prediction, our model shows that the hot-hand fallacy can arise even when individuals infer the persistence parameter from the data. In that case, however, there must be true serial correlation, and under-prediction after very long streaks.

Our model derives a number of implications from the single psychological bias of the gambler's

[4] The intuition when $\rho < \delta - \alpha$ is somewhat different. In that case, Freddy expects luck to reverse slowly, and the positive correlation in the signals is mainly in the short run. To explain the absence of slow reversals, Freddy overestimates the persistence parameter, converging to a value $\tilde{\rho}$ between ρ and $\delta - \alpha$. Because $\delta - \alpha$ exceeds $\tilde{\rho}$ and ρ, the gambler's fallacy overtakes any updating on the state after a very long streak, thus generating under-prediction. The effects of the gambler's fallacy are weaker after shorter streaks, and Freddy over-predicts relative to Tommy because he overestimates the state's persistence. Finally, the initial under-prediction is because Tommy updates heavily after short streaks due to the high short-run correlation in the signals.

fallacy and provides a flexible framework for studying further implications.[5] In fact, we conclude this paper in Section 6 by sketching some potential applications of the model. An intriguing broad implication is that people might end up believing in predictability even in *i.i.d.* environments. Thus, even when asset returns are *i.i.d.*, investors might be willing to pay for information about past market movements, or hire financial experts who are assumed to observe these movements. This could help explain why people invest in actively-managed funds in spite of the evidence that these funds do not outperform the market. Our model could also speak to other finance puzzles such as the equity-premium puzzle or the momentum/reversals in stock returns.

2 The Model

We assume that an individual observes a sequence of signals whose probability distribution depends on an underlying state. The signal s_t in Period $t = 1, 2, ..$ is

$$s_t = \theta_t + \epsilon_t, \tag{1}$$

where θ_t is the state and ϵ_t is an *i.i.d.* normal shock with mean zero and variance $\sigma_\epsilon^2 > 0$. The state evolves according to the auto-regressive process

$$\theta_t = \mu + \rho(\theta_{t-1} - \mu) + \eta_t, \tag{2}$$

where $\rho \in [0, 1)$ is the reversion rate to the long-run mean μ, and η_t is an *i.i.d.* normal shock with mean zero, variance σ_η^2, and independent of ϵ_t. The signal can be interpreted, for example, as the return on a mutual fund, the state as the fund manager's ability, the shock ϵ_t as the manager's luck, and the parameter ρ as the extent to which ability is persistent. Alternatively, if the fund consists of multiple managers with heterogeneous abilities, $1-\rho$ can be interpreted as the managerial turnover.

We model the gambler's fallacy as the mistaken belief that the sequence $\{\epsilon_t\}_{t\geq 1}$ is not *i.i.d.*,

[5]The tractability and applicability of our model is an important advantage over Rabin's (2002) related model of the law of small numbers. In Rabin, Freddy draws from an urn with replacement but believes falsely that the urn is replenished only every odd period. Thus, the law of small numbers applies every even period. Our model eliminates the artificial distinction between odd and even periods, allows all variables to be continuous rather than two-valued, and characterizes beliefs explicitly for general sequences of signals. More importantly, by allowing for a time-varying state our model is suitable for analyzing the hot-hand fallacy, which involves precisely a belief in the state's time-variation. Our work is also related to the theory of momentum and reversals in Barberis, Shleifer and Vishny (1998). In BSV, individuals observe a random walk but believe incorrectly that innovations are drawn either from a regime with excess reversals or from one with excess streaks. If the reversal regime is the more common, individuals under-react to short streaks because they expect a reversal. They over-react, however, to longer streaks because they take them as sign of a switch to the streak regime. To provide a psychological foundation for their assumptions, BSV appeal to a combination of biases: the conservatism bias for the reversal regime and the representativeness bias for the streak regime. Our model, by contrast, not only derives such biases from the single underlying bias of the gambler's fallacy, but in doing so provides predictions as to which biases are the more relevant in different informational settings.

but exhibits systematic reversals. More specifically, we assume that according to the individual,

$$\epsilon_t = \tilde{\epsilon}_t - \alpha \sum_{k=0}^{\infty} \delta^k \epsilon_{t-1-k}, \tag{3}$$

where the shocks $\tilde{\epsilon}_t$ are *i.i.d.* and normal with mean zero and variance $\tilde{\sigma}_\epsilon^2$, and $\alpha, \delta \in [0,1)$ are exogenous parameters.[6] The parameter α characterizes the strength of the belief in the gambler's fallacy. When $\alpha = 0$, the individual is an error-free Bayesian (Tommy) treating the sequence $\{\epsilon_t\}_{t\geq 1}$ correctly as *i.i.d.*. When $\alpha > 0$, however, the individual (Freddy) has the wrong model of how signals are generated. To see how this corresponds to the gambler's fallacy, we take the expectation of both sides of Equation (3) conditional on Freddy's information as of Period $t-1$:

$$E_{t-1}(\epsilon_t) = -\alpha \sum_{k=0}^{\infty} \delta^k E_{t-1}(\epsilon_{t-1-k}). \tag{4}$$

If Freddy believes that all shocks up to $t-1$ have been positive, then he expects the shock ϵ_t to be negative because $\alpha > 0$ and $\delta \geq 0$. Thus, consistent with the gambler's fallacy, Freddy believes that if a fund manager was lucky up to Period $t-1$, luck should reverse in Period t.

The parameter δ characterizes the memory duration of the gambler's fallacy. When $\delta = 0$, Freddy believes that the shock ϵ_t ought to counteract only the shock in Period $t-1$. By contrast, when δ is close to one, Freddy believes that ϵ_t ought to counteract the average of all past shocks. The size of δ relative to α affects the precise manifestation of the gambler's fallacy, as shown in the following lemma:

Lemma 1 *Freddy's expectation as of Period $t-1$ of a future shock is*

$$E_{t-1}(\epsilon_{t'}) = -\alpha(\delta - \alpha)^{t'-t} \sum_{k=0}^{\infty} \delta^k E_{t-1}(\epsilon_{t-1-k}).$$

Lemma 1 generalizes Equation (4) to Freddy's prediction of all future shocks rather than only ϵ_t. Suppose that Freddy believes that all shocks up to $t-1$ have been positive. When $\delta > \alpha$, he expects all future shocks to be negative, in line with the gambler's fallacy. By contrast, when $\delta < \alpha$, he expects future shocks to follow an oscillating pattern: the shock in Period t will be negative to compensate for the positive shocks up to $t-1$, the shock in Period $t+1$ will be positive to compensate for the negative shock in Period t, and so on. In the intermediate case $\delta = \alpha$, the gambler's fallacy lasts exactly one period: Freddy expects the shock in Period t to be negative and all future shocks to be zero on average.[7]

[6] We set $\epsilon_t = 0$ for $t \leq 0$, so that all terms in the infinite sum are well defined.

[7] Intuitively, sufficient memory rules out oscillation by preventing the predicted counter-shocks from looming too

To our knowledge, the issue of whether the gambler's fallacy generates oscillatory patterns has not been raised in the psychology literature. Because such patterns seem somewhat counter-intuitive, however, we focus on the case $\delta \geq \alpha$ in the remainder of this paper. Appendix A contains a brief discussion of the case $\delta < \alpha$.

Notice that we are allowing Freddy to perceive correctly the sequence $\{\eta_t\}_{t\geq 1}$ as *i.i.d.* Thus, we are assuming that the gambler's fallacy applies to the relationship between the signals and the underlying states (i.e., the sequence $\{\epsilon_t\}_{t\geq 1}$), but not to the relationship between the states and their long-run average. Since signals are observable but states are not, the former relationship represents an aspect of randomness that is more salient to Freddy, and perhaps more conducive to biases. But although we find the gambler's fallacy more compelling for $\{\epsilon_t\}_{t\geq 1}$, our formalism can easily cover the case where the fallacy applies also to $\{\eta_t\}_{t\geq 1}$. We sketch this extension in Appendix A and show that most of our results carry through.

A final note concerning our model is in order. While to our knowledge, the gambler's fallacy is discussed in the psychology literature solely in *i.i.d.* environments (e.g., coin flips), we are applying it to environments that are potentially not *i.i.d.* We believe that doing so in a psychologically compelling way is important for understanding the fallacy's implications. Indeed, our derivation of the hot-hand fallacy from the gambler's fallacy involves agents coming to believe (perhaps wrongly) that the environment is not *i.i.d.*. This requires a theory of how the gambler's fallacy would manifest itself in such environments. Moreover, in many economic and finance applications of our model, it is reasonable to assume that the true environment is not *i.i.d.*

3 Freddy Infers

In this section we formulate Freddy's inference problem, and establish some general results that serve as the basis for the more specific results of Sections 4 and 5. The inference problem consists in using the signals to learn about the underlying state θ_t, and possibly about the parameters of the model. Freddy's model is characterized by the variance σ_η^2 of the shocks to the state, the persistence ρ, the variance $\tilde{\sigma}_\epsilon^2$ of the noise in the signal (where the noise refers to the shocks $\tilde{\epsilon}_t$ rather than ϵ_t), the long-run mean μ, and the gambler's fallacy parameters (α, δ). We assume that Freddy does not question his belief in the gambler's fallacy, i.e., has a dogmatic point prior on (α, δ). He can, however, learn about the other parameters. From now on, we reserve the notation $(\sigma_\eta^2, \rho, \mu)$ for the true parameter values, and denote generic values by $(\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\mu})$. Thus, Freddy can learn about the parameter vector $\tilde{p} \equiv (\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu})$.

large relative to the initial shock as to force predictions back in the other direction.

3.1 No Parameter Uncertainty

We start our analysis with the case where Freddy is certain about all model parameters. This case is relatively simple and serves as an input for the parameter-uncertainty case. Freddy's inference problem can be formulated as one of recursive (Kalman) filtering. Recursive filtering is a technique for solving inference problems where (i) inference concerns a "state vector" evolving according to a stochastic process, (ii) a noisy signal of the state vector is observed in each period, (iii) the stochastic structure is linear and normal.[8]

To formulate the recursive-filtering problem, we must define the state vector, the equation according to which the state vector evolves, and the equation linking the state vector to the signal. The state vector must include not only the state θ_t, but also some measure of the past realizations of luck since according to Freddy luck reverses predictably. It turns out that all past luck realizations can be condensed into an one-dimensional statistic. This statistic can be appended to the state θ_t, and therefore, recursive filtering can be used even in the presence of the gambler's fallacy. We define the state vector as

$$x_t \equiv \left[\theta_t - \tilde{\mu}, \epsilon_t^\delta\right]',$$

where the statistic of past luck realizations is

$$\epsilon_t^\delta \equiv \sum_{k=0}^{\infty} \delta^k \epsilon_{t-k},$$

and v' denotes the transpose of the vector v. Equations (2) and (3) imply that the state vector evolves according to

$$x_t = Ax_{t-1} + w_t, \tag{5}$$

where

$$A \equiv \begin{bmatrix} \tilde{\rho} & 0 \\ 0 & \delta - \alpha \end{bmatrix}$$

and

$$w_t \equiv [\eta_t, \tilde{\epsilon}_t]'.$$

Equations (1)-(3) imply that the signal is related to the state vector through

$$s_t = \tilde{\mu} + Cx_{t-1} + v_t, \tag{6}$$

[8] For textbooks on recursive filtering see, for example, Anderson and Moore (1979) and Balakrishnan (1987). We are using the somewhat cumbersome term "state vector" because we are reserving the term "state" for θ_t, and the two concepts differ in our model.

where

$$C \equiv [\tilde{\rho}, -\alpha]$$

and $v_t \equiv \eta_t + \tilde{\epsilon}_t$. To start the recursion, we must specify Freddy's prior beliefs for the initial state x_0. We denote the mean and variance of θ_0 by $\overline{\theta}_0$ and $\sigma^2_{\theta,0}$, respectively. Since $\epsilon_t = 0$ for $t \leq 0$, the mean and variance of ϵ^δ_0 are both zero. Proposition 1 determines Freddy's beliefs about the state in Period t, conditional on the history of signals $\mathcal{H}_t \equiv \{s_{t'}\}_{t'=1,..,t}$ up to that period.

Proposition 1 *Conditional on $\mathcal{H}_t$, x_t is normal with mean $\overline{x}_t$ given recursively by*

$$\overline{x}_t = A\overline{x}_{t-1} + G_t \left[s_t - \tilde{\mu} - C\overline{x}_{t-1}\right], \qquad \overline{x}_0 = [\overline{\theta}_0 - \tilde{\mu}, 0]', \tag{7}$$

and covariance matrix Σ_t given recursively by

$$\Sigma_t = A\Sigma_{t-1}A' - \left[C\Sigma_{t-1}C' + V\right] G_t G_t' + W, \qquad \Sigma_0 = \left[\begin{array}{cc} \sigma^2_{\theta,0} & 0 \\ 0 & 0 \end{array} \right], \tag{8}$$

where

$$G_t \equiv \frac{1}{C\Sigma_{t-1}C' + V} \left[A\Sigma_{t-1}C' + U\right], \tag{9}$$

$V \equiv E(v_t^2)$, $W \equiv E(w_t w_t')$, and $U \equiv E(v_t w_t)$.

Freddy's conditional expectation evolves according to Equation (7). This is simply a regression equation: the state vector in Period t is regressed on that period's signal, conditional on the history up to Period $t-1$. The regression coefficient G_t depends on Freddy's conditional variance of the state Σ_{t-1}. Proposition 2 shows that when t goes to ∞, this variance converges to a limit that is independent of the initial value Σ_0.

Proposition 2 *$Lim_{t\to\infty}\Sigma_t = \Sigma$, where Σ is the unique solution in the set of positive matrices of*

$$\Sigma = A\Sigma A' - \frac{1}{C\Sigma C' + V} \left[A\Sigma C' + U\right] \left[A\Sigma C' + U\right]' + W. \tag{10}$$

Proposition 2 implies that there is convergence to a steady state where the conditional variance Σ_t is equal to the constant Σ, the regression coefficient G_t is equal to the constant

$$G \equiv \frac{1}{C\Sigma C' + V} \left[A\Sigma C' + U\right], \tag{11}$$

and the conditional expectation of the state vector x_t evolves according to a linear equation with constant coefficients. The steady state plays an important role in our analysis: it is also the limit in the case of parameter uncertainty because Freddy eventually becomes certain about the parameter values. In Sections 4 and 5 we rely on the steady state when, for example, considering the effect of a signal on Freddy's future predictions. The linearity result of Proposition 1 implies that the effect is deterministic and independent of the history of past signals. The convergence result of Proposition 2 implies that the effect is time-independent.

3.2 Parameter Uncertainty

We next allow Freddy to be uncertain about the parameters of his model. Parameter uncertainty is a natural assumption in many settings. For example, Freddy might be uncertain about the extent to which the ability of fund managers varies over time (σ_η^2) or is persistent (ρ). Alternatively, under the managerial-turnover interpretation, Freddy might be unsure about whether managers differ in ability.

Because parameter uncertainty eliminates the normality that is necessary for recursive filtering, Freddy's inference problem threatens to be less tractable. Recursive filtering can, however, be used as part of a two-stage procedure. In a first stage, we fix each model parameter to a given value, and compute the likelihood of a history of signals conditional on these values. Because the conditional probability distribution is normal, the likelihood can be computed using the recursive-filtering formulas of Section 3.1. In a second stage, we combine the likelihood with Freddy's prior beliefs, through Bayes' law, and determine Freddy's posteriors on the parameters. We show, in particular, that Freddy's limit posteriors when t goes to ∞ can be derived by maximizing a limit likelihood over all possible parameter values.

We assume that Freddy's prior beliefs over parameter vectors $\tilde{p} \equiv (\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu})$ have support P, and to avoid technicalities, we restrict P to be finite. We denote by $\pi_0(\tilde{p})$ Freddy's prior probability of $\tilde{p}$. As we show below, $\pi_0(\tilde{p})$ affects Freddy's limit posteriors only through its support.

The likelihood function $L_t(\mathcal{H}_t|\tilde{p})$ associated to a parameter vector $\tilde{p}$ and history $\mathcal{H}_t = \{s_{t'}\}_{t'=1,..,t}$ is the probability density of observing the signals conditional on $\tilde{p}$. From Bayes' law, this density is

$$L_t(\mathcal{H}_t|\tilde{p}) = L_t(s_1 \cdots s_t|\tilde{p}) = \prod_{t'=1}^{t} \ell_{t'}(s_{t'}|s_1 \cdots s_{t'-1}, \tilde{p}) = \prod_{t'=1}^{t} \ell_{t'}(s_{t'}|\mathcal{H}_{t'-1}, \tilde{p}),$$

where $\ell_t(s_t|\mathcal{H}_{t-1}, \tilde{p})$ denotes the density of s_t conditional on $\tilde{p}$ and $\mathcal{H}_{t-1}$. The latter density can be computed using the recursive-filtering formulas of Section 3.1. Indeed, Proposition 1 shows that conditional on $\tilde{p}$ and $\mathcal{H}_{t-1}$, x_{t-1} is normal. Since s_t is a linear function of x_{t-1}, it is also normal with a mean and variance that we denote by $\overline{s}_t(\tilde{p})$ and $\sigma^2_{s,t}(\tilde{p})$, respectively. Thus:

$$\ell_t(s_t|\mathcal{H}_{t-1},\tilde{p}) = \frac{1}{\sqrt{2\pi\sigma^2_{s,t}(\tilde{p})}} \exp\left[-\frac{[s_t-\overline{s}_t(\tilde{p})]^2}{2\sigma^2_{s,t}(\tilde{p})}\right],$$

and

$$L_t(\mathcal{H}_t|\tilde{p}) = \frac{1}{\sqrt{(2\pi)^t\prod_{t'=1}^t \sigma^2_{s,t'}(\tilde{p})}} \exp\left[-\sum_{t'=1}^t \frac{[s_{t'}-\overline{s}_{t'}(\tilde{p})]^2}{2\sigma^2_{s,t'}(\tilde{p})}\right]. \tag{12}$$

Freddy's posterior beliefs over parameter vectors can be derived from his prior beliefs and the likelihood through Bayes' law. Denoting the posterior probability of $\tilde{p}$ in Period t by $\pi_t(\tilde{p})$, we have

$$\pi_t(\tilde{p}) = \frac{\pi_0(\tilde{p})L_t(\mathcal{H}_t|\tilde{p})}{\sum_{\tilde{p}'\in P}\pi_0(\tilde{p}')L_t(\mathcal{H}_t|\tilde{p}')}. \tag{13}$$

To determine Freddy's posterior beliefs in the limit when t goes to ∞, we need to determine the asymptotic behavior of the likelihood function $L_t(\mathcal{H}_t|\tilde{p})$. Intuitively, this behavior depends on how well Freddy can fit the data (i.e., the history of signals) using the model corresponding to $\tilde{p}$. To evaluate the fit of a model, we consider the true model according to which the data are generated. The true model is characterized by $\alpha = 0$ and the true parameters $p \equiv (\mu, \rho, \sigma^2_\eta, \sigma^2_\epsilon)$. We denote by $\overline{s}^*_t$ and $\sigma^{*2}_{s,t}$, respectively, the true mean and variance of s_t conditional on $\mathcal{H}_{t-1}$, and by P^* and E^*, respectively, the true probability measure and expectation operator.

Theorem 1

$$\lim_{t\to\infty}\frac{\log L_t(\mathcal{H}_t|\tilde{p})}{t} = -\frac{1}{2}\left[\log\left[2\pi\sigma^2_s(\tilde{p})\right] + \frac{\sigma^{*2}_s + e(\tilde{p})}{\sigma^2_s(\tilde{p})}\right] \equiv F(\tilde{p}) \tag{14}$$

almost surely, where

$$\begin{aligned}
\sigma^2_s(\tilde{p}) &\equiv \lim_{t\to\infty}\sigma^2_{s,t}(\tilde{p}),\\
\sigma^{*2}_s &\equiv \lim_{t\to\infty}\sigma^{*2}_{s,t},\\
e(\tilde{p}) &\equiv \lim_{t\to\infty}E^*\left[\overline{s}^*_t - \overline{s}_t(\tilde{p})\right]^2.
\end{aligned}$$

Theorem 1 implies that the likelihood function is asymptotically equal to

$$L_t(\mathcal{H}_t|\tilde{p}) \sim \exp\left[tF(\tilde{p})\right],$$

thus growing exponentially at the rate $F(\tilde{p})$. Note that $F(\tilde{p})$ does not depend on the specific history $\mathcal{H}_t$ of signals, and is thus deterministic. That the likelihood function becomes deterministic for large t follows from the law of large numbers, which is the main result that we need to prove the theorem. The appropriate large-numbers law in our setting is one applying to non-independent and non-identically distributed random variables. Non-independence is because the expected values $\overline{s}_t^*$ and $\overline{s}_t(\tilde{p})$ involve the entire history of past signals, and non-identical distributions are because at any finite time we are not at the steady state.

The growth rate $F(\tilde{p})$ can be interpreted as the fit of the model corresponding to $\tilde{p}$. A straightforward corollary of Theorem 1 is that when t goes to ∞, Freddy gives positive probability only to values of $\tilde{p}$ that maximize $F(\tilde{p})$ over the set P. We denote the set of these values by $m(P)$.

Corollary 1 *If $\tilde{p} \notin m(P) \equiv argmax_{\tilde{p} \in P} F(\tilde{p})$, then $\lim_{t\to\infty} \pi_t(\tilde{p}) = 0$ almost surely.*

To solve the fit-maximization problem, we ignore discreteness issues from now on and allow the set P to be continuous. Proposition 3 characterizes the solution to the problem.

Proposition 3 *Suppose that P satisfies the cone property*

$$(\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu}) \in P \Rightarrow (\lambda\tilde{\sigma}_\eta^2, \tilde{\rho}, \lambda\tilde{\sigma}_\epsilon^2, \tilde{\mu}) \in P, \quad \forall \lambda > 0.$$

Then, $\tilde{p} \in m(P)$ if and only if

1. $e(\tilde{p}) = \min_{\tilde{p}' \in P} e(\tilde{p}') \equiv e(P)$
2. $\sigma_s^2(\tilde{p}) = \sigma_s^{*2} + e(\tilde{p})$.

The characterization of Proposition 3 is very intuitive. The function $e(\tilde{p})$ is the expected squared difference between the true conditional mean of s_t, and the mean that Freddy computes under the model corresponding to $\tilde{p}$. Thus, $e(\tilde{p})$ measures the error in Freddy's predictions relative to those of the true model, and a model maximizing the fit must minimize this error.

A model maximizing the fit must also generate the right measure of uncertainty about the future signals. Freddy's uncertainty under the model corresponding to $\tilde{p}$ is measured by $\sigma_s^2(\tilde{p})$, the conditional variance of s_t. This must equal to the true error in Freddy's predictions, which is the sum of two orthogonal components: the error $e(\tilde{p})$ relative to the predictions of the true model, and the error in the true model's predictions, i.e., the true conditional variance σ_s^{*2}.

The cone property in Proposition 3 ensures that in maximizing the fit, there is no conflict between minimizing $e(\tilde{p})$ and setting $\sigma_s^2(\tilde{p}) = \sigma_s^{*2} + e(\tilde{p})$. Indeed, $e(\tilde{p})$ depends on $\tilde{\sigma}_\eta^2$ and $\tilde{\sigma}_\epsilon^2$ only through their ratio $\tilde{s}_\eta^2 \equiv \tilde{\sigma}_\eta^2/\tilde{\sigma}_\epsilon^2$ because only $\tilde{s}_\eta^2$ affects the vector G of regression coefficients. The cone property ensures that given any feasible ratio $\tilde{s}_\eta^2$, we can scale $\tilde{\sigma}_\eta^2$ and $\tilde{\sigma}_\epsilon^2$ to make $\sigma_s^2(\tilde{p})$ equal to $\sigma_s^{*2} + e(\tilde{p})$. The cone property is satisfied, in particular, when the set P includes all parameter values:

$$P = P_0 \equiv \left\{(\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu}) : \tilde{\sigma}_\eta^2 \in \mathbb{R}^+, \tilde{\rho} \in [0,1), \tilde{\sigma}_\epsilon^2 \in \mathbb{R}^+, \tilde{\mu} \in \mathbb{R}\right\}.$$

Maximizing the fit is potentially complicated. Indeed, the function $e(\tilde{p})$ depends on the vector G of regression coefficients, which in turn depends on $\tilde{p}$ in a complicated fashion through the recursive-filtering formulas of Section 3.1. In the Appendix (Lemma 4), however, we derive an expression for $e(\tilde{p})$ that we can easily minimize numerically. Moreover, in Sections 4 and 5 we derive closed-form characterizations of the solution to the minimization problem when Freddy is close to rational (α small).

We conclude this section by determining Tommy's limit posteriors. We examine, in particular, whether Tommy converges to the true parameter values when he initially entertains all values, i.e., $P = P_0$. Since Tommy is a Bayesian, his limit posteriors solve the fit-maximization problem for $\alpha = 0$.

Proposition 4 *Suppose that $\alpha = 0$.*

- *If $\sigma_\eta^2 > 0$ and $\rho > 0$, then $m(P_0) = \{(\sigma_\eta^2, \rho, \sigma_\epsilon^2, \mu)\}$.*
- *If $\sigma_\eta^2 = 0$ or $\rho = 0$, then* $m(P_0) = \left\{ \begin{array}{r} (\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \mu) : [\tilde{\sigma}_\eta^2 = 0, \tilde{\rho} \in [0,1), \tilde{\sigma}_\epsilon^2 = \sigma_\eta^2 + \sigma_\epsilon^2] \\ or\ [\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2 = \sigma_\eta^2 + \sigma_\epsilon^2, \tilde{\rho} = 0] \end{array} \right\}.$

Proposition 4 shows that Tommy converges to the true parameter values if $\sigma_\eta^2 > 0$ and $\rho > 0$. If $\sigma_\eta^2 = 0$ or $\rho = 0$, however, then he remains undecided between the true model and a set of other models. The intuition is that when the state is constant over time ($\sigma_\eta^2 = 0$) or not persistent ($\rho = 0$), signals are *i.i.d.*, and Tommy cannot identify which of the two parameters is zero. Of course, Tommy's failure to converge to the true model is inconsequential because all models that he converges to predict correctly that signals are *i.i.d.*

4 Independent Signals

In this section we consider Freddy's inference problem when the signals are *i.i.d.* As pointed out in the previous section, *i.i.d.* signals can be generated when the state is constant over time ($\sigma_\eta^2 = 0$) or not persistent ($\rho = 0$). We first study Freddy's "free-form" inference when he initially entertains all parameter values ($P = P_0$). We next allow for prior knowledge, i.e., assume that Freddy knows with certainty the true values of some parameters.

4.1 No Prior Knowledge

Proposition 5 characterizes Freddy's convergent beliefs.

Proposition 5 *Suppose that $\alpha > 0$, and $\sigma_\eta^2 = 0$ or $\rho = 0$. Then $m(P_0)$ consists of the two elements*

$$\tilde{p}_1 \equiv \left(\frac{\alpha(1 - \delta(\delta - \alpha))}{\delta} \left(\sigma_\eta^2 + \sigma_\epsilon^2\right), \delta - \alpha, \frac{\delta - \alpha}{\delta} \left(\sigma_\eta^2 + \sigma_\epsilon^2\right), \mu \right)$$

and

$$\tilde{p}_2 \equiv (\sigma_\eta^2 + \sigma_\epsilon^2, 0, 0, \mu).$$

Moreover, $e(P_0) = 0$.

Since $e(P_0) = 0$, Freddy ends up predicting the signals correctly despite being subject to the gambler's fallacy. To explain the intuition for this surprising result, consider the models that Freddy converges to. Under the first model, he believes that the state θ_t exhibits time-variation (since $\tilde{\sigma}_\eta^2 = \frac{\alpha(1-\delta(\delta-\alpha))}{\delta} \left(\sigma_\eta^2 + \sigma_\epsilon^2\right) > 0$) and persistence (since $\tilde{\rho} = \delta - \alpha \geq 0$). While this belief is obviously erroneous, it exactly offsets the erroneous belief in the gambler's fallacy. Indeed, consider the impact of a high signal in Period t on Freddy's forecast of the subsequent signals. Because of the linearity of our model, we can characterize the impact simply by considering a marginal increase in the Period t signal.

Lemma 2 *In steady state,*

$$\frac{dE_t(s_{t'})}{ds_t} = CA^{t'-t-1}G = \tilde{\rho}^{t'-t}G_1 - \alpha(\delta - \alpha)^{t'-t-1}G_2,$$

where G_1 and G_2 are the components of the regression-coefficient vector G, and $t' > t$.

Following a unit increase in the Period t signal, Freddy believes that the state has increased by the regression coefficient G_1. He then expects the signal in Period $t' > t$ to be higher by $\tilde{\rho}^{t'-t}G_1$ because the state reverts to its long-run mean at the rate $\tilde{\rho}^k$. This belief in the time-varying state is counteracted by the gambler's fallacy which corresponds to the term $\alpha(\delta - \alpha)^{t'-t-1}G_2$. Freddy attributes the high signal in Period t partly to luck through the regression coefficient G_2. He then expects the future signals to be lower because luck reverses. As shown in Lemma 1, this effect decays over time at the rate $(\delta - \alpha)^k$.

Since Freddy converges to the persistence parameter $\tilde{\rho} = \delta - \alpha$, the two erroneous beliefs decay at the same rate. Therefore, they cancel each other if they have the same amplitude. This occurs when the ratio $\tilde{s}_\eta^2 \equiv \tilde{\sigma}_\eta^2/\tilde{\sigma}_\epsilon^2$, which controls the relative size of the regression coefficients G_1 and G_2, takes the value in Proposition 5.

Freddy's erroneous belief in the time-varying state resembles to a hot-hand fallacy: Freddy believes that the state varies in a serially correlated manner while in fact it is constant. This belief, however, does not constitute a hot-hand fallacy in the conventional sense because it does not manifest itself in the predictions of the signals, being offset by the gambler's fallacy. Freddy can be thought of, for example, as a basketball fan convinced that a player is going through hot and cold periods that are exactly offset by quick reversals of luck. If the player makes a shot, Freddy thinks, then probably his skill level is temporarily heightened - but having made a shot also means (even as a hot player) that he is "due" for a miss.

The second model that Freddy converges to is simpler than the first. Under this model, the state θ_t exhibits no persistence (since $\tilde{\rho} = 0$) and is equal to the signal (since $\tilde{\sigma}_\epsilon^2 = 0$). Freddy then treats the state as *i.i.d.*, and predicts correctly *i.i.d.* signals. Note, however, that the success of this model relies on the assumption that the gambler's fallacy does not apply to the relationship between the states and their long-run mean (i.e., Freddy treats the sequence $\{\eta_t\}_{t\geq 1}$ correctly as *i.i.d.*). One motivation for this assumption, given in Section 2, is that this relationship is not observable. When the signal is equal to the state, however, the relationship becomes observable and the plausibility of the assumption is stretched. In Appendix A we show that when the gambler's fallacy applies to both $\{\epsilon_t\}_{t\geq 1}$ and $\{\eta_t\}_{t\geq 1}$, Freddy can only converge to a model very similar to the first model of Proposition 5.

Note that under both models of Proposition 5, Freddy converges to the true long-run mean μ. This result is general, holding also for serially correlated signals as shown in Section 5. The intuition is that the long-run mean determines the average value of the signal, and inferring this average can be separated from inferring properties of the fluctuations around the average.

4.2 Prior Knowledge

The possibility that Freddy can develop a belief in the time-varying state that offsets the gambler's fallacy relies crucially on the unrestricted nature of his priors. For example, under the first model of Proposition 5, both $\tilde{\sigma}_\eta^2$ and $\tilde{\rho}$ differ from their true values σ_η^2 and ρ (except in the knife-edge case where $\rho = \delta - \alpha$). Therefore, if Freddy knows with certainty what the true values are, then he cannot converge to that model. In other words, more knowledge can hurt Freddy because it reduces his flexibility to come up with the incorrect model that offsets the gambler's fallacy.

The most straightforward example of prior knowledge is when Freddy is aware that the state is constant ($\sigma_\eta^2 = 0$). The prototypical occurrence of this is when people observe the flips of a coin they know is fair. The state can then be defined as the probability of heads or tails, and it is known and constant.

When Freddy knows that $\sigma_\eta^2 = 0$, he obviously cannot develop a belief in the time-varying state. Therefore, his predictions are influenced only by the gambler's fallacy. This is consistent with the experimental evidence: when, for example, subjects know that they are dealing with fair coins they tend to predict reversals. Of course, our model matches the evidence by construction, but we believe that this is a strength of our approach (in taking the gambler's fallacy as a primitive bias and examining whether the hot-hand fallacy can follow as an implication). Indeed, one could argue that the hot-hand fallacy is a primitive bias, either unconnected to the gambler's fallacy or perhaps even generating it. But then, one would have to explain why such a primitive bias does not arise in experiments involving fair coins.

The hot-hand fallacy tends to arise in settings where people are uncertain about the mechanism generating the data, and where a belief in serially correlated variation is plausible a priori. Such settings are common when human skill is involved. For example, it is plausible - and often true - that the performance of a basketball player can fluctuate systematically over time because of mood, well-being, etc. Consistent with the evidence, our approach can generate the hot-hand fallacy in such settings. Indeed, we show below that Freddy's predictions can exhibit excessive persistence precisely when he allows for the possibility that the state can vary in a serially correlated manner, i.e., $\sigma_\eta^2 > 0$ and $\rho > 0$.

We next consider an example where Freddy's prior knowledge does not rule out serially correlated variation in the state. We assume that the state is constant, and that while Freddy is unsure about this, he knows the persistence parameter $\rho > 0$. This example involves an assumption about counterfactuals: if the state did fluctuate (which it does not), it would be persistent, and Freddy knows what the persistence is. To motivate the example, we return to the managerial-turnover interpretation of our model. Suppose that a mutual-fund's performance is determined by the average ability within the team of its managers, each manager's ability is constant over time, and a fraction of managers turn over in each period. Then, the state is time-varying if managers differ in ability because turnover can alter the average ability within the team. Suppose that in reality all managers are identical but Freddy is unsure about this. Freddy could, however, observe the turnover, in which case he knows with confidence what the state's persistence would be if there is time-variation. Proposition 6 characterizes Freddy's convergent beliefs:

Proposition 6 *Suppose that $\alpha > 0$, $\sigma_\eta^2 = 0$, $\rho > 0$, and Freddy considers parameter values in the set*

$$P_\rho \equiv \left\{ (\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu}) : \tilde{\sigma}_\eta^2 \in \mathbb{R}^+, \tilde{\rho} = \rho, \tilde{\sigma}_\epsilon^2 \in \mathbb{R}^+, \tilde{\mu} \in \mathbb{R} \right\}.$$

Then, any element of $m(P_\rho)$ *satisfies* $\tilde{\sigma}_\eta^2 > 0$, $\tilde{\sigma}_\epsilon^2 > 0$, *and* $\tilde{\mu} = \mu$. *Moreover,* $e(P_\rho) = 0$ *only when* $\rho = \delta - \alpha$.

Freddy ends up predicting the signals correctly ($e(P_\rho) = 0$) only in the knife-edge case where $\rho = \delta - \alpha$. Indeed, recall that in the absence of prior knowledge, Freddy gets around the gambler's fallacy by assuming that the state is time-varying with persistence parameter $\tilde{\rho} = \delta - \alpha$. This belief is consistent with the prior knowledge of ρ only when $\rho = \delta - \alpha$. When $\rho \neq \delta - \alpha$, Freddy still develops a belief in the time-varying state ($\tilde{\sigma}_\eta^2 > 0$) to explain why the signals do not exhibit systematic reversals. This belief, however, cannot fully offset the gambler's fallacy. Note that Freddy always converges to the true long-run mean μ.

To sharpen our characterization of Freddy's convergent beliefs, we consider the case where he is close to rational, i.e., α is small. Proposition 7 determines the convergent beliefs in closed form.

Proposition 7 *Suppose that* $\sigma_\eta^2 = 0$ *and* $\rho > 0$. *When* α *converges to zero, the set*

$$\left\{ \left(\frac{\tilde{\sigma}_\eta^2}{\alpha \tilde{\sigma}_\epsilon^2}, \tilde{\sigma}_\epsilon^2 \right) : \left(\tilde{\sigma}_\eta^2, \rho, \tilde{\sigma}_\epsilon^2, \mu \right) \in m(P_\rho) \right\}$$

converges (in the set topology) to the point (z, σ_ϵ^2), *where*

$$z \equiv \frac{(1-\rho^2)^2}{\rho(1-\rho\delta)}. \tag{15}$$

Proposition 7 implies that when α is small, Freddy ends up believing that the variance of the shocks to the state is $\tilde{\sigma}_\eta^2 \approx \alpha z \sigma_\epsilon^2$. Using this result, we can characterize Freddy's errors in predicting the signals. We examine, in particular, how Freddy predicts a signal that follows a streak of similar signals. We assume that the streak is between Periods t and $t'-1$, and all signals in the streak are identical.

Proposition 8 *Suppose that* α *is small,* $\sigma_\eta^2 = 0$, $\rho > 0$, *and Freddy considers parameter values in the set* P_ρ. *If* $\rho > \delta$, *then in steady state*

$$\sum_{t''=t}^{t'-1} \frac{dE_{t'-1}(s_{t'})}{ds_{t''}}$$

is negative for $t' = t+1$ *and becomes positive as* t' *increases. If* $\rho < \delta$, *then the opposite is true.*

When $\rho > \delta$, Freddy predicts a low signal following a short streak of high signals, but a high signal following a longer streak. Thus, the gambler's fallacy dominates for short streaks but the hot-hand fallacy dominates for longer ones. Intuitively, a high signal has a long-lasting impact on Freddy's estimate of the state because when ρ is large, Freddy knows the state to be very persistent. Therefore, a long streak of high signals has a large cumulative impact—and this belief in the increased state eventually overtakes the gambler's fallacy, leading Freddy to predict a high signal. At the same time, Freddy cannot predict a high signal after any streak of high signals because he would then be overcompensating for the gambler's fallacy: by adopting a smaller value of $\tilde{\sigma}_\eta^2$, he would hold a weaker belief in positive correlation, and his predictions would match the *i.i.d.* signals more closely. Therefore, the gambler's fallacy must dominate for short streaks. The results are reversed when $\rho < \delta$: the gambler's fallacy dominates for long streaks because Freddy believes in a large cumulative reversal of luck, and the hot-hand fallacy appears after short streaks.

Proposition 8 makes use of the closed-form solutions derived for small α. For general α, the fit-maximization problem can be solved through a simple numerical algorithm. The numerical results confirm the Proposition, with $\delta - \alpha$ taking the place of δ.

Summarizing, when signals are *i.i.d.*, our model confirms the intuition of previous researchers that the hot-hand fallacy can arise as a consequence of the gambler's fallacy. At the same time, we qualify this intuition in important ways. We find that the hot-hand fallacy can arise only when individuals have strong priors on the state's persistence but are open to learning about time-variation. We also show that the hot-hand fallacy does not always appear after long streaks: this requires the additional assumption that individuals expect the state to be persistent but luck to reverse quickly. Finally, we show that the endogenous belief in the hot hand cannot be so strong to always dominate the gambler's fallacy: otherwise individuals could improve their predictions by adopting a weaker such belief.

5 Serially Correlated Signals

In this section we consider Freddy's inference problem when the signals are serially correlated. Serial correlation arises when the state varies over time ($\sigma_\eta^2 > 0$) and is persistent ($\rho > 0$). To highlight the new effects relative to the *i.i.d.* case, we assume that Freddy is initially uncertain about the values of all parameters. Proposition 9 shows that unlike the *i.i.d.* case, Freddy can predict the signals correctly only when $\rho = \delta - \alpha$.

Proposition 9 *Suppose that $\alpha > 0$, $\sigma_\eta^2 > 0$, and $\rho > 0$. Then, $e(P_0) = 0$ only when $\rho = \delta - \alpha$.*

The intuition is that in the *i.i.d.* case, Freddy gets around the gambler's fallacy by taking the state's persistence to be $\tilde{\rho} = \delta - \alpha$ rather than the true value ρ. When the state is time-varying, the persistence parameter influences the signal process, and a discrepancy between $\tilde{\rho}$ and ρ leads

Freddy to incorrect predictions. Formally, consider the impact of a high signal in Period t on Freddy's forecast of the subsequent signals. From Lemma 2, this is

$$\frac{dE_t(s_{t'})}{ds_t} = \tilde{\rho}^{t'-t}G_1 - \alpha(\delta - \alpha)^{t'-t-1}G_2. \tag{16}$$

In the *i.i.d.* case, Freddy's predictions are correct because this expression can be made equal to zero for all $t' > t$, by setting $\tilde{\rho} = \delta - \alpha$. When signals are serially correlated, the same expression must be made equal to the impact of s_t under the true model

$$\frac{dE^*_t(s_{t'})}{ds_t} = \rho^{t'-t}G^*_1, \tag{17}$$

where $G^*_1 > 0$ because the state is time-varying. Equality is possible only in the knife-edge case where $\rho = \delta - \alpha$. Thus, although Freddy can develop a model that gets him around the gambler's fallacy when signals are *i.i.d.*, this is not feasible in a more complicated environment.

To determine Freddy's convergent beliefs when $\rho \neq \delta - \alpha$, we consider the case where he is close to rational, i.e., α is small. As in Section 4, this case allows for closed-form solutions that convey the main intuitions. In addition to α, we take the variance σ^2_η of the shocks to the state to be small, meaning that signals are close to *i.i.d.* Formally, we assume that both α and σ^2_η converge to zero, holding their ratio constant, and we set $\omega \equiv \sigma^2_\eta/(\alpha\sigma^2_\epsilon)$.[9] As we explain later in this section, our closed-form results are consistent with the numerical solutions derived in the general case.

When Freddy is close to rational, he converges to a model that generates predictions close to the true model's. Freddy might not, however, converge to the correct belief that the state is almost constant (σ^2_η small). Indeed, recall that *i.i.d* signals are generated either because the state is constant ($\sigma^2_\eta = 0$) or not persistent ($\rho = 0$). Therefore, when Freddy predicts that signals are close to *i.i.d.*, this might be because he converges to a model where $\tilde{\rho}$, and not $\tilde{\sigma}^2_\eta$, is small. Under this model, Freddy treats the state as approximately *i.i.d.* ($\tilde{\rho}$ small) and the signal as approximately equal to the state ($\tilde{\sigma}^2_\epsilon$ small). This model is, in fact, similar to the second model of Proposition 5 that gets around the gambler's fallacy by taking the sequence $\{\tilde{\epsilon}_t\}_{t\geq 1}$ to have zero variance. Condition 1 rules out convergence to the small-$\tilde{\rho}$ model, thus ensuring that Freddy converges correctly to the small-$\tilde{\sigma}^2_\eta$ model.

Condition 1 *Either*

$$\rho > \frac{\delta}{1 + \sqrt{1 - \delta^2}} \tag{18}$$

[9]The case where σ^2_η remains constant when α goes to zero can be derived by setting $\omega = \infty$ in our solutions, but the case $\omega < \infty$ is the more interesting.

or

$$\omega\rho^2\sqrt{1-\delta^2} > \left(1-\rho^2\right)^{\frac{3}{2}}. \tag{19}$$

The intuition behind Condition 1 is that the small-$\tilde{\rho}$ model does a good job in getting around the gambler's fallacy but a poor one in explaining the serially correlated signals. Therefore, it is dominated by the small-$\tilde{\sigma}_\eta^2$ model when the serial correlation is significant, which occurs when σ_η^2 and ρ are large. Condition 1 is indeed satisfied (through Equation (18)) when ρ is larger than δ, and also when it is not much smaller. It is also satisfied (through Equation (19)) for any value of ρ when the ratio $\omega \equiv \sigma_\eta^2/(\alpha\sigma_\epsilon^2)$ is large enough. We assume Condition 1 from now on because the small-$\tilde{\rho}$ model seems somewhat unappealing: its success relies heavily on the assumption that the gambler's fallacy does not apply to the sequence $\{\eta_t\}_{t\geq 1}$. Proposition 10 solves the fit-maximization problem.

Proposition 10 *Suppose that $\sigma_\eta^2 > 0$, $\rho > 0$, and Condition 1 is met. When α and σ_η^2 converge to zero, holding ω constant, the set*

$$\left\{\left(\frac{\tilde{\sigma}_\eta^2}{\alpha\tilde{\sigma}_\epsilon^2}, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu}\right) : \left(\tilde{\sigma}_\eta^2, \tilde{\rho}, \tilde{\sigma}_\epsilon^2, \tilde{\mu}\right) \in m(P_0)\right\}$$

converges (in the set topology) to the point $(z, r, \sigma_\epsilon^2, \mu)$, where

$$z \equiv \frac{(1-r^2)^2}{r(1-r\delta)} + \frac{\omega\rho(1-r^2)^2}{r(1-\rho^2)(1-\rho r)} \tag{20}$$

and r is the solution to

$$\frac{r-\delta}{(1-r\delta)^2} + \frac{\omega\rho(r-\rho)}{(1-\rho^2)(1-\rho r)^2} = 0. \tag{21}$$

Equation (21) implies that r is between ρ and δ. Therefore, when α and σ_η^2 are small, Freddy converges over time to a persistence parameter $\tilde{\rho}$ that is between the true value ρ and the memory parameter δ. Consider, for example, the case where $\rho > \delta$, i.e., Freddy expects luck to reverse quickly and yet sees highly persistent signals. To explain the absence of quick reversals, Freddy believes that shocks to the state are large but short-lived. Thus, he underestimates the persistence, tilting $\tilde{\rho}$ towards the value δ that characterizes the speed of the reversals. Of course, he tilts $\tilde{\rho}$ only partially towards δ because he must also account for the true persistence of the observed signals. In the opposite case where $\rho < \delta$, Freddy overestimates the persistence to explain the absence of slow reversals.

Proposition 10 implies that when α and σ_η^2 are small, Freddy believes that the variance of the shocks to the state is $\tilde{\sigma}_\eta^2 \approx \alpha z \sigma_\epsilon^2 = (z/\omega)\sigma_\eta^2$. One might expect the ratio z/ω to be greater than one because the gambler's fallacy can lead Freddy to overestimate the state's time-variation. When $\rho > \delta$, z/ω is indeed greater than one, and so is the ratio $\frac{z(1-\rho^2)}{\omega(1-r^2)}$. Thus, Freddy overestimates both the variance of the shocks to the state ($\tilde{\sigma}_\eta^2 > \sigma_\eta^2$) and the state's unconditional variance ($\frac{\tilde{\sigma}_\eta^2}{1-\tilde{\rho}^2} > \frac{\sigma_\eta^2}{1-\rho^2}$). Somewhat surprisingly, however, Freddy can underestimate both variances when ρ is sufficiently smaller than δ. Indeed, when $\rho < \delta$, Freddy overestimates the state's persistence, and this can compensate for the gambler's fallacy more effectively than overestimating the variance.

The intuition for the above results, and for some of the subsequent ones, can be seen through a graphical representation of the fit-maximization problem. In the proof of Proposition 10 we show that the error $e(\tilde{p})$ is approximately equal to

$$e(\tilde{p}) \approx \sigma_\epsilon^2 \sum_{k=1}^{\infty} \left(\frac{\tilde{s}_\eta^2}{1-\tilde{\rho}^2}\tilde{\rho}^k - \alpha\delta^{k-1} - \frac{s_\eta^2}{1-\rho^2}\rho^k \right)^2 + (\tilde{\mu} - \mu)^2, \tag{22}$$

where $s_\eta^2 \equiv \sigma_\eta^2/\sigma_\epsilon^2$ and $\tilde{s}_\eta^2 \equiv \tilde{\sigma}_\eta^2/\tilde{\sigma}_\epsilon^2$. The infinite sum in Equation (22) has an intuitive interpretation. Recall from Equations (16) and (17) that following a high signal in Period t, Freddy's forecast of the signal in Period $t' = t + k$, $k \geq 1$, differs from the true forecast by

$$\frac{dE_t(s_{t+k})}{ds_t} - \frac{dE_t^*(s_{t+k})}{ds_t} = \tilde{\rho}^k G_1 - \alpha(\delta - \alpha)^{k-1} G_2 - \rho^k G_1^*. \tag{23}$$

The regression coefficient G_1^* measures the signal's impact on the true forecast of the state. When the state is almost constant (σ_η^2 small), this coefficient is small and approximately equal to $s_\eta^2/(1-\rho^2)$. Likewise, when $\tilde{\sigma}_\eta^2$ is small, the regression coefficient G_1 under Freddy's model is approximately equal to $\tilde{s}_\eta^2/(1-\tilde{\rho}^2)$. Finally, when α is small, $(\delta-\alpha)^k \approx \delta^k$. Therefore, the infinite sum in Equation (22) concerns the "term structure" of Freddy's forecast errors as of Period t: it is the sum of squared differences between Freddy's forecast and the true forecast, over all forecast horizons $k \geq 1$.[10]

Figure 1 plots the two forecasts as a function of the horizon $k \geq 1$. The thin solid line represents the true (Tommy's) forecast, which decays at the rate ρ^k. The thick solid line represents Freddy's forecast, which is derived by subtracting the effect of the gambler's fallacy (represented by the thin dashed line and decaying at the rate δ^k) from the belief in the time-varying state (represented by the thick dashed line and decaying at the rate $\tilde{\rho}^k$). The optimization problem consists in choosing the parameters $\tilde{\sigma}_\eta^2$ and $\tilde{\rho}$ that characterize the amplitude and decay rate of the belief in the time-varying state. The objective is to bring the two solid lines as close as possible in terms of the sum of squared differences.

[10]The definition of $e(\tilde{p})$ in Theorem 1 concerns only the forecast of next period's signal, and not of all future signals. Because of linearity, however, the forecast error can be broken into independent errors generated by each of the past signals. Moreover, for small σ_η^2, $\tilde{\sigma}_\eta^2$, and α, the error generated by the Period $t-k$ signal on the forecast of the Period t signal is approximately equal to the error that the Period t signal induces on the forecast of the Period $t+k$ signal.

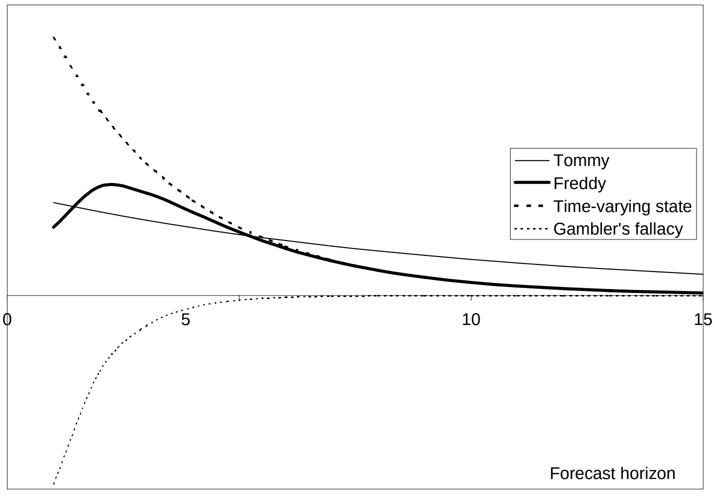

Figure 1: Effect of a high signal in Period t on the forecasts of future signals. The x-axis represents the forecast horizon $k \geq 1$, and the y-axis the forecast of the Period $t+k$ signal. The figure is drawn for $\rho > \delta$.

Figure 1 confirms that Freddy tilts his persistence estimate away from the true value ρ and towards δ. Indeed, the figure is drawn for $\rho > \delta$, and shows that Freddy's belief in the time-varying state must decay faster than Tommy's ($\tilde{\rho} < \rho$) to better counter the gambler's fallacy. At the same time, Freddy's belief must have a larger initial amplitude, meaning that Freddy overestimates the variance.[11]

Figure 1 reveals the time-pattern of Freddy's forecast errors: Freddy forecasts below Tommy for short horizons, above for intermediate horizons, and below again for long horizons. This pattern is, in fact, general, holding also for $\rho < \delta$. The intuition is similar as for our next result, which concerns Freddy's prediction of a signal that follows a streak of similar signals. To state the result, we assume that the streak is between Periods t and $t'-1$, and all signals in the streak are identical.

Proposition 11 *Suppose that α and σ_η^2 are small, $\sigma_\eta^2 > 0$, $\rho \notin \{0, \delta\}$, Condition 1 is met, and*

[11] Freddy's belief in the time-varying state has a larger initial amplitude than Tommy's if $\frac{\tilde{s}_\eta^2 \tilde{\rho}}{1-\tilde{\rho}^2} > \frac{s_\eta^2 \rho}{1-\rho^2}$. Since $\tilde{\rho} < \rho$ and $\tilde{\sigma}_\epsilon^2 \approx \sigma_\epsilon^2$, this implies that $\frac{\tilde{\sigma}_\eta^2}{1-\tilde{\rho}^2} > \frac{\sigma_\eta^2}{1-\rho^2}$ and $\tilde{\sigma}_\eta^2 > \sigma_\eta^2$, i.e., Freddy overestimates the variance.

The intuition why Freddy can underestimate the variance when $\rho < \delta$ can be seen graphically as follows. Suppose that Freddy estimates the state's unconditional variance correctly, i.e., $\frac{\tilde{\sigma}_\eta^2}{1-\tilde{\rho}^2} = \frac{\sigma_\eta^2}{1-\rho^2}$. Then, Equations $\tilde{\rho} > \rho$ and $\tilde{\sigma}_\epsilon^2 \approx \sigma_\epsilon^2$ imply that $\frac{\tilde{s}_\eta^2 \tilde{\rho}^k}{1-\tilde{\rho}^2} > \frac{s_\eta^2 \rho^k}{1-\rho^2}$, meaning that the thick dashed line is above the thin solid line. Suppose now that α is small but δ is close to one, in which case the thin dashed line is close to zero but decays slowly. Then, Freddy's forecasts are above Tommy's for short horizons (because the thin dashed line is close to zero), and below for long horizons (because the thin dashed line decays more slowly than all others). If $\tilde{\sigma}_\eta^2$ is increased, this will barely reduce Freddy's under-prediction in the long term because the thick dashed line has decayed to zero. Freddy's over-prediction in the short term, however, will worsen significantly. Therefore, the sum of squared differences can be reduced by reducing $\tilde{\sigma}_\eta^2$, meaning that Freddy underestimates the variance.

Freddy initially entertains all parameter values ($P = P_0$). Then, in steady state

$$\sum_{t''=t}^{t'-1} \frac{dE_{t'-1}(s_{t'})}{ds_{t''}} - \sum_{t''=t}^{t'-1} \frac{dE^*_{t'-1}(s_{t'})}{ds_{t''}}$$

is negative for $t' = t+1$, becomes positive as t' increases, and then becomes negative again.

Proposition 11 shows that Freddy under-predicts the next signal after a short streak of high signals, over-predicts after a longer streak, and under-predicts again after a very long streak. While this result holds both for $\rho > \delta$ and $\rho < \delta$, the intuition is different in the two cases.

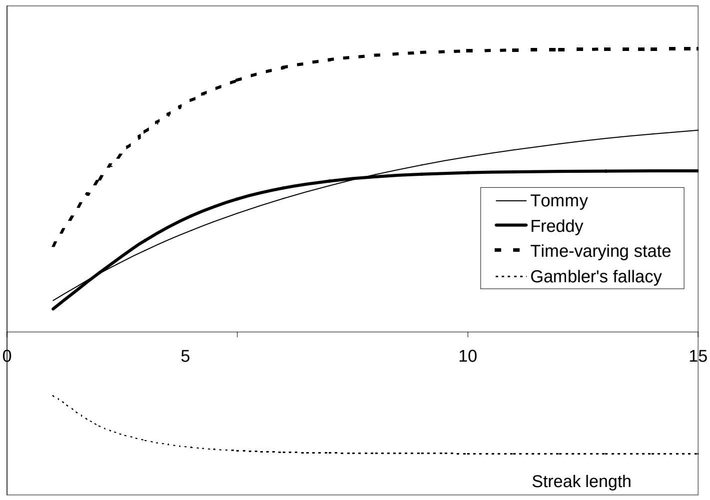

Figure 2: Effect of a streak of high signals on the forecast of the signal following the streak. The x-axis represents the streak's length, and the y-axis the forecast of the next signal. The figure is drawn for $\rho > \delta$ and the same parameter values as Figure 1.

Figure 2 considers the case $\rho > \delta$. The thin solid line represents Tommy's forecast of the signal following a streak of high signals, and the thick solid line represents Freddy's forecast. The latter is generated by Freddy's belief in the increased state (thick dashed line), and his expectation of a reversal in luck (thin dashed line).[12] Because Freddy expects luck to reverse quickly, he under-predicts the signal following a short streak. The belief in quick reversals, however, leads him to assume that the reversal following a streak is mainly generated by the streak's last signals, thus not increasing substantially with the streak's length. On the other hand, Freddy's forecast of the state increases more substantially with streak length: since he believes that the state mean-reverts relatively slowly ($\tilde{\rho} > \delta$), he assumes that even the early signals in a long streak are informative about the current state. (In terms of Figure 2, the thin dashed line levels off faster than the thick

[12]The thick dashed line is Freddy's forecast when the regression-coefficient vector $G = (G_1, G_2)'$ is replaced by $(G_1, 0)'$, and the thin dashed line is the forecast when G is replaced by $(0, G_2)'$.

dashed line.) Therefore, after a long streak, Freddy's belief in the increased state overtakes the gambler's fallacy, leading to over-prediction of the next signal. Finally, Freddy under-predicts after a very long streak because he underestimates the state's persistence: because the true persistence is high, Tommy's forecast increases even after very long streaks, overtaking Freddy's.

Figure 3 considers the case $\rho < \delta$. Recall that in this case, Freddy compensates for the gambler's fallacy mainly by overestimating the state's persistence rather than the variance. Therefore, after a short streak, his belief in the increased state is not much higher than Tommy's (and can even be lower when Freddy underestimates the variance). In combination with the gambler's fallacy, this generates under-prediction. Because Freddy overestimates the state's persistence, however, his forecast of the state increases faster than Tommy's as the streak gets longer. This generates over-prediction after a long streak. Finally, because Freddy believes that luck reverses slowly, he expects a large cumulative reversal after a very long streak, thus under-predicting the next signal.

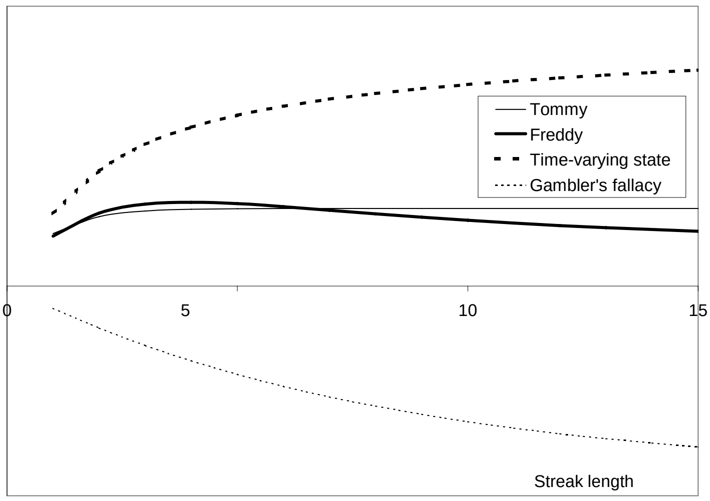

Figure 3: Effect of a streak of high signals on the forecast of the signal following the streak. The x-axis represents the streak's length, and the y-axis the forecast of the next signal. The figure is drawn for $\rho < \delta$.

Proposition 11 makes use of the closed-form solutions derived for small α. Our numerical results for general α confirm the under/over/under-prediction pattern of the proposition.

The analysis of prior knowledge has the same flavor as in the *i.i.d* case. In particular, the results are identical when Freddy knows the state's persistence parameter (ρ) but is uncertain about the extent of time-variation (σ_η^2): when $\rho > \delta - \alpha$, Freddy under-predicts after a short streak and over-predicts after a longer streak, while the opposite holds when $r < \delta - \alpha$.

Summarizing, under serially correlated signals, the hot-hand fallacy can arise even when individuals infer the persistence parameter from the data. In that case, however, our model generates under-prediction after very long streaks. Our model has also the general implication that as the environment becomes more complicated (from *i.i.d.* to serial correlation), individuals have greater difficulty to develop the false model that gets them around the gambler's fallacy.

6 Conclusion

This paper develops a model of belief in the gambler's fallacy, and explores the link with the seemingly opposite bias of the hot hand. We show that an individual who updates rationally except for the gambler's fallacy tends to overestimate the time-variation of an underlying state. This error, however, does not always generate a hot hand: in some cases it offsets the gambler's fallacy, leading to correct predictions. We show that a hot hand can develop—overtaking the gambler's fallacy—when the individual knows confidently the state's persistence parameter or when signals are serially correlated. In each case we determine whether the hot hand arises after long or short streaks of signals.

Our model takes the gambler's fallacy as the only primitive bias. This parsimony allows us to determine endogenously what other biases arise in a variety of environments and under a variety of assumptions about individuals' knowledge of the environment. Our model's tractable normal-linear structure allows the use of recursive-filtering techniques, and can provide a flexible framework for studying the gambler's fallacy in a range of economic and finance settings.

Indeed, we believe that our model's implication that people may come to believe in predictability even in *i.i.d.* environments has consequences in a broad range of settings. Consider, for example, a financial market where returns are *i.i.d.* Because rational Tommies will eventually learn the *i.i.d.* property, they will share a common expectation of future returns. But because Freddies believe that past returns help predict future returns, they will differ in their predictions if they observe different subsets of the return history. Combined with a theory of how trade based on different beliefs arising from statistical errors might occur, such intrinsic differences of opinion among investors who observe only asset returns could be relevant for understanding the large volume of trade in actual markets.

More interesting than the implications of "passive observation" are possible implications for how investors choose to acquire financial information based on erroneous beliefs about the value of that information. When Freddy believes that past returns are useful for predicting the future, he believes technical analysis is useful, even when returns are truly *i.i.d.* Moreover, he might be willing to pay for real-time price information, or keep observing prices for the "right time" to enter or exit a market.[13] Perhaps the main implication of "fallacious predictability" is the potential

[13]Of course, information on current prices is useful whenever an investor's portfolio strategy depends on wealth.

role in explaining exaggerated belief in financial expertise. Investors often seem to rely heavily on the opinions of experts, such as stockbrokers or managers of actively-managed funds. This seems somewhat puzzling: for example, it is well-documented that actively-managed funds do not outperform their market benchmarks on average.[14] Our model can readily generate a belief in non-existent expertise. If Freddy falsely believes in predictability but does not have the time to observe the detailed history of returns, he would treat agents who specialize in observing the market as useful experts.

Additional implications would follow if Freddies constitute a big enough fraction of a market as to affect prices. Suppose, for example, that the signal consists of a firm's *i.i.d.* earnings growth, and Freddies know that $\rho > \delta - \alpha$. Then, because they will base their investment behavior on the under-prediction of earnings after short streaks of good news and over-prediction after longer streaks, expected returns would be high after short streaks and low after longer streaks. This is consistent with the evidence of short-run momentum and long-run reversals in the stock market.[15]

Finally, some interesting implications may follow from Freddy's false belief about variance. When Freddy predicts the signals erroneously, he also overestimates their variance (Proposition 3) because he attributes his prediction error to signal noise. With *i.i.d.* signals, the belief in excess variance arises because of—and, surprisingly, in spite of—a false belief in predictability. Thus, when returns are truly *i.i.d.*, Freddy can hold both a false belief in predictability and a belief that the market is excessively volatile. Because of the latter, Freddy can require a high expected return to enter the market. This could perhaps speak to the equity-premium puzzle (Mehra and Prescott (1985)) that stocks' expected return is high relative to risk.

We believe that the gambler's fallacy may help provide a unified behavioral explanation for the diverse phenomena listed above. Moreover, our model and techniques could provide a flexible framework for pursuing the implications of the gambler's fallacy.

Our point is that investors might be willing to pay for such information for purely speculative reasons.

[14]This holds both for the performance of the average fund in a given year, and for the performance of a given fund over time. See Fama's (1991) survey.

[15]Of course, our model would not make this prediction for all parameter values. For references to the empirical literature on momentum and reversals, see Barberis, Shleifer and Vishny (1998). Our explanation of momentum and reversals is similar in spirit with the one in BSV, although we rely only on the gambler's fallacy to derive other biases endogenously rather than the exogenous amalgam of context-specific biases they base their model on.

A Extensions

A.1 The Case $\delta < \alpha$

When signals are *i.i.d.*, Freddy cannot get around the gambler's fallacy by developing a belief in a time-varying and persistent state. Indeed, under such a belief, Freddy would interpret a high signal as evidence that the state has increased, and would expect future signals to be high. Under the gambler's fallacy, however, Freddy expects a high signal to be followed by an oscillatory pattern of low and high signals because $\delta < \alpha$. These patterns obviously cannot offset. Formally, the first model $\tilde{p}_1$ of Proposition 5 is not feasible because it requires $\tilde{\rho} = \delta - \alpha$, which is inconsistent with $\delta - \alpha < 0 \leq \tilde{\rho} \in [0, 1)$. The second model $\tilde{p}_2$ of Proposition 5 leads to correct predictions, and is the unique model that Freddy converges to. Under this model, the gambler's fallacy is not present because Freddy attributes all of the signal's variation to the state. Moreover, Freddy assumes that the state is not persistent ($\tilde{\rho} = 0$), and this leads him to predict correctly that signals are *i.i.d.*

When Freddy knows confidently the value of $\rho > 0$, he develops a belief in a time-varying and persistent state. To determine his predictions after streaks of signals, recall from Section 4.2 that when $\rho > \delta - \alpha$, the gambler's fallacy dominates after short streaks and the hot-hand fallacy dominates after longer ones. One might conjecture the same to hold when $\delta - \alpha < 0$ since ρ still exceeds $\delta - \alpha$. Our numerical results confirm this conjecture when $\delta - \alpha > -\rho$, i.e., when $\delta - \alpha$ is not too negative. When $\delta - \alpha < -\rho$, however, the hot-hand fallacy dominates after streaks of any length. The intuition is similar as for model $\tilde{p}_2$: Freddy attempts to minimize the effects of the gambler's fallacy by attributing most of the signal's variation to the state. A difference with $\tilde{p}_2$ is that Freddy's predictions are incorrect because he takes the state to be persistent. However, predictions are more accurate than when Freddy attributes most of the signal's variation to luck because he has difficulty explaining the oscillatory pattern implied by the gambler's fallacy.

When signals are serially correlated, the results have the same flavor. When $\delta - \alpha$ is not too negative, predictions after streaks are as in Section 5: Freddy under-predicts after a short streak, over-predicts after a longer streak, and under-predicts again after a very long streak. When $\delta - \alpha$ is sufficiently negative, however, the first under-prediction disappears: Freddy over-predicts after all but very long streaks.

A.2 Gambler's Fallacy on State Variation

Our formalism can be easily extended to the case where the gambler's fallacy applies to both $\{\epsilon_t\}_{t\geq 1}$ and $\{\eta_t\}_{t\geq 1}$. Suppose that according to Freddy,

$$\eta_t = \tilde{\eta}_t - \alpha \sum_{k=0}^{\infty} \delta^k \eta_{t-1-k}, \tag{A.1}$$

where the shocks $\tilde{\eta}_t$ are *i.i.d.* and normal with mean zero and variance $\tilde{\sigma}_\eta^2$. The assumption that α, δ are common to both $\{\epsilon_t\}_{t\geq 1}$ and $\{\eta_t\}_{t\geq 1}$ is for simplicity and can easily be dropped. To formulate the recursive-filtering problem, we expand the state vector to

$$x_t \equiv \left[\theta_t - \tilde{\mu}, \eta_t^\delta, \epsilon_t^\delta\right]',$$

where

$$\eta_t^\delta \equiv \sum_{k=0}^{\infty} \delta^k \eta_{t-k}.$$

We also set

$$A \equiv \begin{bmatrix} \tilde{\rho} & -\alpha & 0 \\ 0 & \delta - \alpha & 0 \\ 0 & 0 & \delta - \alpha \end{bmatrix},$$

$$w_t \equiv [\tilde{\eta}_t, \tilde{\eta}_t, \tilde{\epsilon}_t]',$$

$$C \equiv [\tilde{\rho}, -\alpha, -\alpha],$$

and $v_t \equiv \tilde{\eta}_t + \tilde{\epsilon}_t$. Under these definitions, the analysis of Freddy's inference in Section 3 carries through identical.

When signals are *i.i.d.* and Freddy initially entertains all parameter values, he converges to the unique model

$$\tilde{p}_0 \equiv \left(\frac{\alpha(1-\delta(\delta-\alpha))}{\delta}\left(\sigma_\eta^2 + \sigma_\epsilon^2\right), \delta, \frac{\delta-\alpha}{\delta}\left(\sigma_\eta^2 + \sigma_\epsilon^2\right), \mu\right),$$

predicting the signals correctly. This model is very similar to model $\tilde{p}_1$ of Proposition 5: Freddy gets around the gambler's fallacy by developing a belief in a time-varying and persistent state. The difference with Proposition 5 is that model $\tilde{p}_2$ no longer leads to correct predictions. Indeed, under that model, the state is *i.i.d.* and equal to the signal. Under the gambler's fallacy for $\{\eta_t\}_{t\geq 1}$, however, Freddy expects the *i.i.d.* state to exhibit reversals.

When Freddy knows confidently the value of $\rho > 0$, he ends up believing that $\tilde{\sigma}_\eta^2 > 0$. Furthermore, when α is small, $\tilde{\sigma}_\eta^2$ is small relative to $\tilde{\sigma}_\epsilon^2$. Therefore, the effects of the gambler's fallacy for $\{\eta_t\}_{t\geq 1}$ are small relative to $\{\epsilon_t\}_{t\geq 1}$, and the analysis is the same as when the gambler's fallacy applies only to $\{\epsilon_t\}_{t\geq 1}$. In particular, Freddy's limit beliefs are given by Proposition 7, and his predictions after streaks by Proposition 8. Moreover, our numerical results confirm the patterns of Proposition 8 for general values of α.

When signals are serially correlated, we can show as in Proposition 9 that Freddy's predictions differ from those of the true model, except in knife-edge cases. When α is small, Freddy converges to a model where $\tilde{\sigma}_\eta^2$ is small (and Condition 1 is no longer needed to rule out the small-$\tilde{\rho}$ model because the gambler's fallacy for $\{\eta_t\}_{t\geq 1}$ invalidates model $\tilde{p}_2$). Therefore, the effects of the gambler's fallacy for $\{\eta_t\}_{t\geq 1}$ are small relative to $\{\epsilon_t\}_{t\geq 1}$, and the analysis becomes identical to the case where the gambler's fallacy applies only to $\{\epsilon_t\}_{t\geq 1}$. In particular, Freddy's limit beliefs are given by Proposition 10, and his predictions after streaks by Proposition 11. Our numerical results confirm the pattern of Proposition 11 for general values of α, except in a parameter region where the pattern reverses to over/under/over-prediction. Intuitively, Freddy can over-predict after very long streaks because his estimate $\tilde{\rho}$ of the state's persistence can exceed both ρ and δ. Freddy can exaggerate the persistence to compensate for his belief that shocks to the state should exhibit reversals.

B Proofs

Proof of Lemma 1: Define x_t and ϵ_t^δ as in Section 3.1. Equation (5) implies that $E_t(x_{t'}) = AE_t(x_{t'-1})$ for $t' > t$. Iterating between t' and t, we find

$$E_t(x_{t'}) = A^{t'-t}E_t(x_t). \tag{B.1}$$

Equation (B.1) implies that $E_{t-1}(\epsilon_{t'-1}^\delta) = (\delta-\alpha)^{t'-t}E_{t-1}(\epsilon_{t-1}^\delta)$. Therefore,

$$E_{t-1}(\epsilon_{t'}) = E_{t-1}(\tilde{\epsilon}_{t'} - \alpha\epsilon_{t'-1}^\delta) = -\alpha E_{t-1}(\epsilon_{t'-1}^\delta) = -\alpha(\delta-\alpha)^{t'-t}E_{t-1}(\epsilon_{t-1}^\delta).$$

∎

Proof of Proposition 1: Our formulation of the recursive-filtering problem is as in standard textbooks. For example, Equations (5) and (6) follow from (4.1.1) and (4.1.4) in Balakrishnan (1987) if x_{n+1} is replaced by x_t, x_n by x_{t-1}, A_n by A, U_n by 0, N_n^s by w_t, v_n by $s_t-\mu$, C_n by C, and N_n^0 by v_t. Equation (7) follows from (4.6.14), if the latter is written for $n+1$ instead of n, and $\overline{x}_{n+1}$ is replaced by $\overline{x}_t$, $\overline{x}_n$ by $\overline{x}_{t-1}$, and AK_n+Q_n by G_t. That G_t so defined is given by Equation (9), follows from (4.1.29) and (4.6.12) if H_{n-1} is replaced by Σ_{t-1}, G_nG_n' by V, and J_n by U. Equation (8) follows from (4.6.18) if the latter is written for $n+1$ instead of n, P_n is substituted from (4.1.30), and F_nF_n' is replaced by W. ∎

Proof of Proposition 2: It suffices to show (Balakrishnan, p.182-184) that the eigenvalues of $A - UV^{-1}C$ have modulus smaller than one. This matrix is

$$\begin{bmatrix} \frac{\tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_\eta^2+\tilde{\sigma}_\epsilon^2}\tilde{\rho} & \frac{\tilde{\sigma}_\eta^2}{\tilde{\sigma}_\eta^2+\tilde{\sigma}_\epsilon^2}\alpha \\ -\frac{\tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_\eta^2+\tilde{\sigma}_\epsilon^2}\tilde{\rho} & \delta - \frac{\tilde{\sigma}_\eta^2}{\tilde{\sigma}_\eta^2+\tilde{\sigma}_\epsilon^2}\alpha \end{bmatrix}.$$

The characteristic polynomial is

$$\begin{aligned} & \lambda^2 - \lambda \left[\frac{\tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}\tilde{\rho} + \delta - \frac{\tilde{\sigma}_\eta^2}{\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\eta^2}\alpha \right] + \frac{\tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}\tilde{\rho}\delta \\ \equiv \; & \lambda^2 - \lambda b + c \end{aligned}$$

Suppose that the roots λ_1, λ_2 of this polynomial are real, in which case $\lambda_1 + \lambda_2 = b$ and $\lambda_1 \lambda_2 = c$. Since $c > 0$, λ_1 and λ_2 have the same sign. If λ_1 and λ_2 are negative, they are both greater than -1, since $b > -1$ from $\alpha < 1$ and $\tilde{\rho}, \delta \geq 0$. If λ_1 and λ_2 are positive, then at least one is smaller than 1, since $b < 2$ from $\tilde{\rho}, \delta < 1$ and $\alpha \geq 0$. But since the characteristic polynomial for $\lambda = 1$ takes the value

$$(1-\delta)\left(1 - \frac{\tilde{\sigma}_\epsilon^2}{\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}\tilde{\rho}\right) + \frac{\tilde{\sigma}_\eta^2}{\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}\alpha > 0,$$

both λ_1 and λ_2 are smaller than 1. Suppose instead that λ_1, λ_2 are complex. In that case, they are conjugates and the modulus of each is $c < 1$. ∎

Lemma 3 determines $\overline{s}_t^*$, the true mean of s_t conditional on $\mathcal{H}_{t-1}$, and $\overline{s}_t(\tilde{p})$, the mean that Freddy computes under the parameter vector $\tilde{p}$. These are expressed as a function of the sequence of orthogonalized signals $\{\zeta_{t'}\}_{t'=1,..,t-1}$, where $\zeta_{t'} \equiv s_{t'} - \overline{s}_{t'}^*$. This sequence contains the same information as $\{s_{t'}\}_{t'=1,..,t-1}$, but has the advantage that the elements are uncorrelated. Using Lemma 3, we determine the error $e(\tilde{p}) \equiv \lim_{t\to\infty} E^* \left[\overline{s}_t^* - \overline{s}_t(\tilde{p})\right]^2$ in Lemma 4. To state both lemmas, we define the matrices $D_t \equiv A - G_t C$, $D \equiv A - GC$, and

$$J_{t,t'} \equiv \begin{cases} \prod_{k=t'}^{t} D_k & \text{for } t' = 1,..,t, \\ I & \text{for } t' > t. \end{cases}$$

We also use the superscript * for the matrices A, C, G, and Σ in the recursive-filtering problem under the true model. Finally, for simplicity we set the initial condition $\overline{x}_0 = 0$.

Lemma 3 *The true mean* $\overline{s}_t^*$ *is given by*

$$\overline{s}_t^* = \mu + \sum_{t'=1}^{t-1} C^*(A^*)^{t-t'-1} G_{t'}^* \zeta_{t'} \tag{B.2}$$

and Freddy's mean $\overline{s}_t(\tilde{p})$ *by*

$$\overline{s}_t(\tilde{p}) = \tilde{\mu} + \sum_{t'=1}^{t-1} CM_{t,t'}\zeta_{t'} + CM_t^{\mu}(\mu - \tilde{\mu}), \tag{B.3}$$

where

$$M_{t,t'} \equiv J_{t-1,t'+1}G_{t'} + \sum_{k=t'+1}^{t-1} J_{t-1,k+1}G_k C^*(A^*)^{k-t'-1}G_{t'}^*,$$

$$M_t^{\mu} \equiv \sum_{t'=1}^{t-1} J_{t-1,t'+1}G_{t'}.$$

Proof: Consider the recursive-filtering problem under the true model, and denote by $\overline{x}_t^*$ the mean of x_t conditional on $\mathcal{H}_t$. Equation (6) implies that

$$\overline{s}_t^* = \mu + C^*\overline{x}_{t-1}^*. \tag{B.4}$$

Equation (7) then implies that

$$\overline{x}_t^* = A^*\overline{x}_{t-1}^* + G_t^*(s_t - \overline{s}_t^*) = A^*\overline{x}_{t-1}^* + G_t^*\zeta_t.$$

Iterating between $t-1$ and zero, we find

$$\overline{x}_{t-1}^* = \sum_{t'=1}^{t-1}(A^*)^{t-t'-1}G_{t'}^*\zeta_{t'}. \tag{B.5}$$

Plugging into Equation (B.4), we find Equation (B.2).

Consider next Freddy's recursive-filtering problem under $\tilde{p}$. Equation (7) implies that

$$\overline{x}_t = (A - G_tC)\overline{x}_{t-1} + G_t(s_t - \tilde{\mu}).$$

Iterating between $t-1$ and zero, we find

$$\begin{aligned} \overline{x}_{t-1} &= \sum_{t'=1}^{t-1} J_{t-1,t'+1}G_{t'}(s_t - \tilde{\mu}) \\ &= \sum_{t'=1}^{t-1} J_{t-1,t'+1}G_{t'}(\zeta_t + \mu - \tilde{\mu} + C^*\overline{x}_{t-1}^*), \end{aligned} \tag{B.6}$$

where the second step follows from $s_t = \zeta_t + \overline{s}_t^*$ and Equation (B.4). Substituting $\overline{x}_{t'-1}^*$ from

Equation (B.5), and grouping terms, we find

$$\overline{x}_{t-1} = \sum_{t'=1}^{t-1} M_{t,t'}\zeta_{t'} + M_t^{\mu}(\mu - \tilde{\mu}).$$

Combining this with Equation

$$\overline{s}_t(\tilde{p}) = \tilde{\mu} + C\overline{x}_{t-1} \tag{B.7}$$

(which follows from (6)), we find (B.3). ∎

Lemma 4 *The error $e(\tilde{p})$ is given by*

$$e(\tilde{p}) \equiv \lim_{t\to\infty} E^*\left[\overline{s}_t^* - \overline{s}_t(\tilde{p})\right]^2 = \sigma_s^{*2}\sum_{k=0}^{\infty} N_k^2 + (N^{\mu})^2(\mu - \tilde{\mu})^2, \tag{B.8}$$

where

$$N_k \equiv C^*(A^*)^k G^* - C\left[D^k G + \sum_{k'=0}^{k-1} D^{k-1-k'} G C^*(A^*)^{k'} G^*\right], \tag{B.9}$$

$$N^{\mu} \equiv 1 - C\sum_{k=0}^{\infty} D^k G.$$

Proof: Lemma 3 implies that

$$\overline{s}_t^* - \overline{s}_t(\tilde{p}) = \sum_{t'=1}^{t-1} N_{t,t'}\zeta_{t'} + N_t^{\mu}(\mu - \tilde{\mu}), \tag{B.10}$$

where

$$N_{t,t'} \equiv C^*(A^*)^{t-t'-1} G_{t'}^* - CM_{t,t'},$$

$$N_t^{\mu} \equiv 1 - CM_t^{\mu}.$$

Therefore,

$$\left[\overline{s}_t^* - \overline{s}_t(p)\right]^2 = \sum_{t',t''=1}^{t-1} N_{t,t'}N_{t,t''}\zeta_{t'}\zeta_{t''} + (N_t^{\mu})^2(\mu - \tilde{\mu})^2 + 2\sum_{t'=1}^{t-1} N_{t,t'}N_t^{\mu}\zeta_{t'}(\mu - \tilde{\mu}). \tag{B.11}$$

Since the sequence $\{\zeta_{t'}\}_{t'=1,..,t-1}$ is independent with mean zero under the true measure P^*, we have

$$E^*\left[\overline{s}_t^* - \overline{s}_t(p)\right]^2 = \sum_{t'=1}^{t-1} N_{t,t'}^2 \sigma_{s,t'}^{*2} + (N_t^{\mu})^2(\mu - \tilde{\mu})^2. \tag{B.12}$$

We first determine the limit of $\sum_{t'=1}^{t-1} N_{t,t'}^2 \sigma_{s,t'}^{*2}$. We can write this as

$$\sum_{t'=1}^{t-1} N_{t,t'}^2 \sigma_{s,t'}^{*2} = \sum_{k=1}^{t-1} N_{t,t-k}^2 \sigma_{s,t-k}^{*2} = \sum_{k=1}^{\infty} \phi_{k,t},$$

defining the double sequence $\{\phi_{k,t}\}_{k,t\geq 1}$ as

$$\phi_{k,t} \equiv \begin{cases} N_{t,t-k}^2 \sigma_{s,t-k}^{*2} & \text{for } k = 1,..,t-1, \\ 0 & \text{for } k > t-1. \end{cases}$$

To determine the limit of $\sum_{k=0}^{\infty} \phi_{k,t}$ when t goes to ∞, we determine the limit of $\phi_{k,t}$ for given k and then apply the dominated convergence theorem. The definitions of $N_{t,t'}$ and $M_{t,t'}$ imply that

$$N_{t,t-k} = C^*(A^*)^{k-1} G_{t-k}^* - C\left[J_{t-1,t-k+1} G_{t-k} + \sum_{k'=1}^{k-1} J_{t-1,t-k+k'+1} G_{t-k+k'} C^*(A^*)^{k'-1} G_{t-k}^*\right]. \tag{B.13}$$

Equation (6) applied to the recursive-filtering problem under the true model implies that

$$\sigma_{s,t}^{*2} = C^* \Sigma_{t-1}^* (C^*)' + V^*.$$

When t goes to ∞, G_t^* goes to G^*, G_t to G, Σ_t^* to Σ^*, and $J_{t,t-k}$ to D^{k+1}. Therefore,

$$\lim_{t\to\infty} N_{t,t-k} = C^*(A^*)^{k-1} G^* - C\left[D^{k-1} G + \sum_{k'=1}^{k-1} D^{k-k'-1} G C^*(A^*)^{k'-1} G^*\right] = N_{k-1},$$

$$\lim_{t\to\infty} \sigma_{s,t-k}^{*2} = C^* \Sigma^* (C^*)' + V^* = C^* \Sigma^* (C^*)' + \sigma_\eta^{*2} + \sigma_\epsilon^{*2} \equiv \sigma_s^{*2}, \tag{B.14}$$

implying that

$$\lim_{t\to\infty} \phi_{k,t} = N_{k-1}^2 \sigma_s^{*2}.$$

The dominated convergence theorem will imply that

$$\lim_{t\to\infty} \sum_{k=1}^{\infty} \phi_{k,t} = \sum_{k=1}^{\infty} \lim_{t\to\infty} \phi_{k,t} = \sigma_s^{*2} \sum_{k=0}^{\infty} N_k^2, \tag{B.15}$$

if there exists a sequence $\{\overline{\phi}_k\}_{k\geq 1}$ such that $\sum_{k=1}^{\infty} \overline{\phi}_k < \infty$ and $|\phi_{k,t}| \leq \overline{\phi}_k$ for all $k, t \geq 1$. To construct such a sequence, we note that the eigenvalues of A^* have modulus smaller than one, and so do the eigenvalues of $D \equiv A - GC$ (Balakrishnan, Theorem 4.2.3, p.111). Denoting by $a < 1$ the maximum of the moduli, we can construct a dominating sequence $\{\overline{\phi}_k\}_{k\geq 1}$ that decays geometrically at the rate a^{2k}.

We next determine the limit of N_t^μ. We can write this as

$$N_t^\mu = 1 - C \sum_{k=1}^{t-1} J_{t-1,t-k+1} G_{t-k} = 1 - C \sum_{k=1}^{\infty} \chi_{k,t},$$

defining the double sequence $\{\chi_{k,t}\}_{k,t\geq 1}$ as

$$\chi_{k,t} \equiv \begin{cases} J_{t-1,t-k+1} G_{t-k} & \text{for } k = 1,..,t-1, \\ 0 & \text{for } k > t-1. \end{cases}$$

It is easy to check that the dominated convergence theorem applies for the sequence $\{\chi_{k,t}\}_{k,t\geq 1}$ as well, and thus

$$\lim_{t\to\infty} N_t^\mu = 1 - C \lim_{t\to\infty} \left[\sum_{k=1}^{\infty} \chi_{k,t} \right] = 1 - C \sum_{k=1}^{\infty} \lim_{t\to\infty} \chi_{k,t} = 1 - C \sum_{k=0}^{\infty} D^k G = N^\mu. \tag{B.16}$$

The lemma follows by combining Equations (B.12), (B.15), and (B.16). ∎

Proof of Theorem 1: Equation (12) implies that

$$\frac{2\log L_t(\mathcal{H}_t|\tilde{p})}{t} = -\log(2\pi) - \frac{\sum_{t'=1}^{t} \log \sigma_{s,t'}^2(\tilde{p})}{t} - \frac{1}{t} \sum_{t'=1}^{t} \frac{[s_{t'} - \overline{s}_{t'}(\tilde{p})]^2}{\sigma_{s,t'}^2(\tilde{p})}. \tag{B.17}$$

To determine the limit of the second term, we note that Equation (6) applied to Freddy's recursive-filtering problem under $\tilde{p}$ implies that

$$\sigma_{s,t}^2(\tilde{p}) = C\Sigma_{t-1}C' + V.$$

Therefore,

$$\lim_{t\to\infty} \sigma_{s,t}^2(\tilde{p}) = C\Sigma C' + V \equiv \sigma_s^2(\tilde{p}), \tag{B.18}$$

$$\lim_{t\to\infty} \frac{\sum_{t'=1}^{t} \log \sigma_{s,t'}^2(\tilde{p})}{t} = \lim_{t\to\infty} \log \sigma_{s,t}^2 = \log \sigma_s^2(\tilde{p}). \tag{B.19}$$

To determine the limit of the third term, we write it as

$$\underbrace{\frac{1}{t} \sum_{t'=1}^{t} \frac{\zeta_{t'}^2}{\sigma_{s,t'}^2(\tilde{p})}}_{X} + \underbrace{\frac{1}{t} \sum_{t'=1}^{t} \frac{2\zeta_{t'}[\overline{s}_{t'}^* - \overline{s}_{t'}(\tilde{p})]}{\sigma_{s,t'}^2(\tilde{p})}}_{Y} + \underbrace{\frac{1}{t} \sum_{t'=1}^{t} \frac{[\overline{s}_{t'}^* - \overline{s}_{t'}(\tilde{p})]^2}{\sigma_{s,t'}^2(\tilde{p})}}_{Z}.$$

The terms (X, Y, Z) are averages of random variables, and to determine their limits we apply a law of large numbers (LLN). An appropriate LLN in our setting is that of McLeish (1975) because it deals with random variables that are non-independent and non-identically distributed. To apply

McLeish's LLN, we must define a probability space $(\Omega, \mathcal{F}, P)$, a sequence $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ of σ-algebras, and a sequence $\{U_t\}_{t\geq 1}$ of random variables. The pair $(\{\mathcal{F}_t\}_{t\in\mathbb{Z}}, \{U_t\}_{t\geq 1})$ is a mixingale (McLeish, Definition 1.2, p.830) if and only if there exist sequences $\{c_t\}_{t\geq 1}$ and $\{\psi_k\}_{k\geq 0}$ of nonnegative constants, with $\lim_{k\to\infty}\psi_k = 0$, such that for all $t \geq 1$ and $k \geq 0$:

$$\|E_{t-k}U_t\|_2 \leq \psi_k c_t, \tag{B.20}$$

$$\|U_t - E_{t+k}U_t\|_2 \leq \psi_{k+1}c_t, \tag{B.21}$$

where $\|.\|_2$ denotes the L_2 norm, and $E_{t'}U_t$ the expectation of U_t conditional on $\mathcal{F}_{t'}$. McLeish's LLN (Corollary 1.9, p.832) states that if $(\{\mathcal{F}_t\}_{t\in\mathbb{Z}}, \{U_t\}_{t\geq 1})$ is a mixingale, then

$$\lim_{t\to\infty}\frac{1}{t}\sum_{t'=1}^{t} U_{t'} = 0$$

almost surely, provided that $\sum_{t=1}^{\infty} c_t^2/t^2 < \infty$ and $\{\psi_k\}_{k\geq 0}$ is of size $-1/2$. A sufficient condition for a monotone sequence $\{\psi_k\}_{k\geq 0}$ to be of size $-1/2$ is (p.831) that $\sum_{k=1}^{\infty}\psi_k < \infty$.

In our model, we take the probability measure to be the true measure P^*, and define the sequence $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ as follows: for $t \leq 0$, $\mathcal{F}_t = \{\Omega, \emptyset\}$, and for $t \geq 1$, $\mathcal{F}_t$ is the σ-algebra generated by $\mathcal{H}_t$. We first show that $(\{\mathcal{F}_t\}_{t\in\mathbb{Z}}, \{X_t\}_{t\geq 1})$ is a mixingale, where

$$X_t \equiv \frac{\zeta_t^2}{\sigma_{s,t}^2(\tilde{p})} - E^*\left[\frac{\zeta_t^2}{\sigma_{s,t}^2(\tilde{p})}\right] = \frac{\zeta_t^2}{\sigma_{s,t}^2(\tilde{p})} - \frac{\sigma_{s,t}^{*2}}{\sigma_{s,t}^2(\tilde{p})}.$$

Equation (B.21) trivially holds since $E_{t+k}^*X_t = X_t$. To show Equation (B.20), we note that $E_t^*X_t = X_t$, and $E_{t-k}^*X_t = E^*X_t = 0$ for $k \geq 1$ since the sequence $\{\zeta_t\}_{t\geq 1}$ is independent. Therefore, Equation (B.21) holds with $\psi_0 = 1$, $\psi_k = 0$ for $k \geq 1$, and $c_t = \sup_{t\geq 1}\|X_t\|_2$ for $t \geq 1$. McLeish's LLN implies that

$$\lim_{t\to\infty}\frac{1}{t}\sum_{t'=1}^{t} X_{t'} = 0$$

almost surely. Therefore,

$$\lim_{t\to\infty} X = \lim_{t\to\infty}\frac{1}{t}\left[\sum_{t'=1}^{t} X_{t'} + \sum_{t'=1}^{t}\frac{\sigma_{s,t'}^{*2}}{\sigma_{s,t'}^2(\tilde{p})}\right] = \lim_{t\to\infty}\frac{\sigma_{s,t}^{*2}}{\sigma_{s,t}^2(\tilde{p})} = \frac{\sigma_s^{*2}}{\sigma_s^2(\tilde{p})} \tag{B.22}$$

almost surely.

We next show that $(\{\mathcal{F}_t\}_{t\in\mathbb{Z}}, \{Y_t\}_{t\geq 1})$ is a mixingale, where

$$Y_t \equiv \frac{2\zeta_t[\overline{s}_t^* - \overline{s}_t(\tilde{p})]}{\sigma_{s,t}^2(\tilde{p})}.$$

Equation (B.21) trivially holds since $E_{t+k}^* Y_t = Y_t$. To show Equation (B.20), we note that $E_t^* Y_t = Y_t$, and

$$E_{t-k}^* Y_t = \frac{E_{t-k}^*\left[\zeta_t[\overline{s}_t^* - \overline{s}_t(\tilde{p})]\right]}{\sigma_{s,t}^2(\tilde{p})} = \frac{E_{t-k}^*[\zeta_t] E_{t-k}^*\left[\overline{s}_t^* - \overline{s}_t(\tilde{p})\right]}{\sigma_{s,t}^2(\tilde{p})} = 0$$

for $k \geq 1$, where the second step follows because $\overline{s}_t^* - \overline{s}_t(\tilde{p})$ depends only on $\mathcal{H}_{t-1}$ and thus is independent on ζ_t. Therefore, Equation (B.21) holds with $\psi_0 = 1$, $\psi_k = 0$ for $k \geq 1$, and $c_t = \sup_{t\geq 1} \|Y_t\|_2$ for $t \geq 1$. McLeish's LLN implies that

$$\lim_{t\to\infty} Y = \lim_{t\to\infty} \frac{1}{t}\sum_{t'=1}^{t} Y_{t'} = 0 \tag{B.23}$$

almost surely.

We finally show that $(\{\mathcal{F}_t\}_{t\in\mathbb{Z}}, \{Z_t\}_{t\geq 1})$ is a mixingale, where

$$Z_t \equiv \frac{[\overline{s}_t^* - \overline{s}_t(\tilde{p})]^2}{\sigma_{s,t}^2(\tilde{p})} - E^*\left[\frac{[\overline{s}_t^* - \overline{s}_t(\tilde{p})]^2}{\sigma_{s,t}^2(\tilde{p})}\right].$$

Equation (B.21) trivially holds since $E_{t+k}^* Z_t = Z_t$. To show Equation (B.20), we note that

$$\|E_{t-k}^* Z_t\|_2 = \frac{\left[Var^*\left[E_{t-k}^*[\overline{s}_t^* - \overline{s}_t(\tilde{p})]^2\right]\right]^{\frac{1}{2}}}{\sigma_{s,t}^2(\tilde{p})}, \tag{B.24}$$

where Var^* denotes the variance under the true measure P^*. Equation (B.11) implies that

$$E_{t-k}^*[\overline{s}_t^* - \overline{s}_t(\tilde{p})]^2 = \sum_{t'=t-k+1}^{t-1} N_{t,t'}^2 \sigma_{s,t'}^{*2} + \sum_{t',t''=1}^{t-k} N_{t,t'} N_{t,t''} \zeta_{t'} \zeta_{t''} + (N_t^\mu)^2(\mu - \tilde{\mu})^2 + 2\sum_{t'=1}^{t-k} N_{t,t'} N_t^\mu \zeta_{t'}(\mu - \tilde{\mu}).$$

Since

$$[Var(U_1 + U_2)]^{\frac{1}{2}} \leq [Var(U_1)]^{\frac{1}{2}} + [Var(U_2)]^{\frac{1}{2}}$$

for any two random variables (U_1, U_2), we have

$$\left[Var^*\left[E_{t-k}^*[\overline{s}_t^* - \overline{s}_t(p)]^2\right]\right]^{\frac{1}{2}} \leq \sum_{t',t''=1}^{t-k} |N_{t,t'} N_{t,t''}| \left[Var^*(\zeta_{t'}\zeta_{t''})\right]^{\frac{1}{2}} + 2\sum_{t'=1}^{t-k} |N_{t,t'} N_t^\mu (\mu - \tilde{\mu})| \left[Var^*(\zeta_{t'})\right]^{\frac{1}{2}}.$$

Denoting by $a < 1$ the maximum of the moduli of the eigenvalues of A^* and D, it is easy to show that $\sum_{t'=1}^{t-k} |N_{t,t'}|$ is bounded by a sequence ϕ_k that decays geometrically at the rate a^k. The same is true for $\left[Var^* \left[E^*_{t-k}[\bar{s}^*_t - \bar{s}_t(p)]^2\right]\right]^{\frac{1}{2}}$ and $\|E^*_{t-k} Z_t\|_2$. Therefore, we can choose $\psi_k = \phi_k$ and $c_t = 1$. McLeish's LLN implies that

$$\lim_{t\to\infty} \frac{1}{t} \sum_{t'=1}^{t} Z_{t'} = 0$$

almost surely. Therefore,

$$\lim_{t\to\infty} Z = \lim_{t\to\infty} \frac{1}{t} \left[\sum_{t'=1}^{t} Z_{t'} + \sum_{t'=1}^{t} \frac{E^*[\bar{s}^*_{t'} - \bar{s}_{t'}(\tilde{p})]^2}{\sigma^2_{s,t'}(\tilde{p})} \right] = \lim_{t\to\infty} \frac{E^*[\bar{s}^*_t - \bar{s}_t(\tilde{p})]^2}{\sigma^2_{s,t}(\tilde{p})} = \frac{e(\tilde{p})}{\sigma^2_s(\tilde{p})} \tag{B.25}$$

almost surely. The proposition follows from Equations (B.17) (B.22), (B.23), and (B.25). ∎

Proof of Corollary 1: Consider $\tilde{p} \notin m(P)$, and $\hat{p} \in P$ such that $F(\hat{p}) > F(\tilde{p})$. Equation (13) implies that

$$\pi_t(\tilde{p}) = \frac{\pi_0(\tilde{p}) L_t(\mathcal{H}_t|\tilde{p})}{\sum_{\tilde{p}' \in P} \pi_0(\tilde{p}') L_t(\mathcal{H}_t|\tilde{p}')} < \frac{\pi_0(\tilde{p}) L_t(\mathcal{H}_t|\tilde{p})}{\pi_0(\hat{p}) L_t(\mathcal{H}_t|\hat{p})}.$$

Since from Theorem 1

$$\frac{L_t(\mathcal{H}_t|\tilde{p})}{L_t(\mathcal{H}_t|\hat{p})} \sim \exp\left[t\left[F(\tilde{p}) - F(\hat{p})\right]\right] \to 0,$$

$\pi_t(\tilde{p}) \to 0$ almost surely. ∎

Proof of Proposition 3: Consider $\tilde{p} \in P$ such that $e(\tilde{p}) = e(P)$ and $\sigma^2_s(\tilde{p}) = \sigma^{*2}_s + e(\tilde{p})$. We will show that $F(\tilde{p}) \geq F(\hat{p})$ for any $\hat{p} = (\hat{\sigma}^2_\eta, \hat{\rho}, \hat{\sigma}^2_\epsilon, \hat{\mu}) \in P$. Denote by $\hat{\Sigma}$ and $\hat{G}$ the steady-state variance and regression coefficient for the recursive-filtering problem under $\hat{p}$, and by $\hat{\Sigma}_\lambda$ and $\hat{G}_\lambda$ those under $\hat{p}_\lambda \equiv (\lambda\hat{\sigma}^2_\eta, \hat{\rho}, \lambda\hat{\sigma}^2_\epsilon, \hat{\mu})$ for $\lambda > 0$. It is easy to check that $\lambda\hat{\Sigma}$ solves Equation (10) for $\hat{p}_\lambda$. Since this equation has a unique solution, $\hat{\Sigma}_\lambda = \lambda\hat{\Sigma}$. Equation (9) then implies that $\hat{G}_\lambda = \hat{G}$, and Equations (B.8) and (B.18) imply that $e(\hat{p}_\lambda) = e(\hat{p})$ and $\sigma^2_s(\hat{p}_\lambda) = \lambda\sigma^2_s(\hat{p})$. Therefore,

$$F(\hat{p}_\lambda) = -\frac{1}{2}\left[\log\left[2\pi\lambda\sigma^2_s(\hat{p})\right] + \frac{\sigma^{*2}_s + e(\hat{p})}{\lambda\sigma^2_s(\hat{p})}\right]. \tag{B.26}$$

Since this function is maximized for

$$\lambda^* = \frac{\sigma^{*2}_s + e(\hat{p})}{\sigma^2_s(\hat{p})},$$

we have

$$F(\hat{p}) \leq F(\hat{p}_{\lambda^*}) = \frac{1}{2}\left[\log\left[2\pi\left[\sigma^{*2}_s + e(\hat{p})\right]\right] + 1\right] \leq \frac{1}{2}\left[\log\left[2\pi\left[\sigma^{*2}_s + e(\tilde{p})\right]\right] + 1\right] = F(\tilde{p}),$$

where the second inequality follows from $e(\tilde{p}) = e(P)$ and the second equality from $\sigma^2_s(\tilde{p}) = \sigma^{*2}_s + e(\tilde{p})$. The proof of the converse is along the same lines. ∎

Lemma 5 shows that any convergent model must satisfy $\tilde{\mu} = \mu$.

Lemma 5 *For all $\tilde{p} \in m(P)$, $\tilde{\mu} = \mu$.*

Proof: It suffices to show that $N^{\mu} > 0$: Lemma 4 will then imply that if $\tilde{\mu} \neq \mu$, $e(\tilde{p})$ can be reduced by setting $\tilde{\mu} = \mu$. From the definition of N^{μ} it follows that

$$N^{\mu} = 1 - C(I-D)^{-1}G = 1 - C(I - A + CG)^{-1}G.$$

Replacing A and C by their values, and denoting the components of G by G_1 and G_2, we find after some algebra

$$1 - C(I - A + CG)^{-1}G = \frac{(1-\tilde{\rho})(1-\delta+\alpha)}{[1-\tilde{\rho}(1-G_1)](1-\delta+\alpha) - \alpha(1-\tilde{\rho})G_2}.$$

This is non-zero since $\tilde{\rho}, \alpha, \delta \in [0, 1)$. ∎

Lemma 6 determines when a model can predict the signals equally well as the true model.

Lemma 6 *The error $e(\tilde{p})$ is zero if and only if $\tilde{\mu} = \mu$ and*

$$CA^kG = C^*(A^*)^kG^* \tag{B.27}$$

for all $k \geq 0$.

Proof: From Lemmas 4 and 5 it suffices to show that $\{N_k\}_{k\geq 0} = 0$ if and only if $CA^kG = C^*(A^*)^kG^*$ for all $k \geq 0$. Setting $a_k \equiv C^*(A^*)^kG^* - CA^kG$ and

$$u_k \equiv A^kG - \left[D^kG + \sum_{k'=0}^{k-1} D^{k-1-k'}GC^*(A^*)^{k'}G^* \right],$$

we can write $N_k = a_k + Cu_k$. Simple algebra shows that

$$u_k = Du_{k-1} - Ga_{k-1}.$$

Iterating between k and zero, and using the initial condition $u_0 = 0$, we find

$$u_k = -\sum_{k'=0}^{k-1} D^{k-1-k'}Ga_{k'}.$$

Therefore,

$$N_k = a_k - \sum_{k'=0}^{k-1} CD^{k-1-k'}Ga_{k'}. \tag{B.28}$$

Equation (B.28) implies that$\{N_k\}_{k\geq 0} = 0$ if and only if $\{a_k\}_{k\geq 0} = 0$. ∎

Proof of Proposition 4: Tommy can achieve minimum error $e(P_0) = 0$ by using the vector of true parameters p. Since $e(P_0) = 0$, Proposition 3 and Lemma 6 imply that $\tilde{p} \in m(P_0)$ if and only if (i) $\tilde{\mu} = \mu$, (ii) $CA^k G = C^*(A^*)^k G^*$ for all $k \geq 0$, and (iii) $\sigma_s^2(\tilde{p}) = \sigma_s^{*2}$. Since $\alpha = 0$ for Tommy, we can write Condition (ii) as

$$\tilde{\rho}^{k+1} G_1 = \rho^{k+1} G_1^*. \tag{B.29}$$

We can also write the element (1,1) of Equation (10) as

$$\Sigma_{11} = \frac{(\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2)\tilde{\sigma}_\epsilon^2}{\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}, \tag{B.30}$$

$$\Sigma_{11}^* = \frac{(\rho^2 \Sigma_{11}^* + \sigma_\eta^2)\sigma_\epsilon^2}{\rho^2 \Sigma_{11}^* + \tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}, \tag{B.31}$$

and the first element of Equation (11) as

$$G_1 = \frac{\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2}{\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2}, \tag{B.32}$$

$$G_1^* = \frac{\rho^2 \Sigma_{11}^* + \sigma_\eta^2}{\rho^2 \Sigma_{11}^* + \sigma_\eta^2 + \sigma_\epsilon^2}, \tag{B.33}$$

where the first equation in each case is for $\tilde{p}$ and the second for p. Using Equations (B.18) and (B.14), we can write Condition (iii) as

$$\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2 = \rho^2 \Sigma_{11}^* + \sigma_\eta^2 + \sigma_\epsilon^2. \tag{B.34}$$

Suppose that $\rho\sigma_\eta^2 > 0$, and consider $\tilde{p}$ that satisfies Conditions (i)-(iii). Equation (B.33) implies that $G_1^* > 0$. Since Equation (B.29) must hold for all $k \geq 0$, we have $\rho = \rho^*$ and $G_1 = G_1^*$. We next write Equations (B.30)-(B.33) in terms of the normalized variables $\tilde{s}_\eta^2 \equiv \tilde{\sigma}_\eta^2/\tilde{\sigma}_\epsilon^2$, $S_{11} \equiv \Sigma_{11}/\tilde{\sigma}_\epsilon^2$, $s_\eta^2 \equiv \sigma_\eta^2/\sigma_\epsilon^2$, and $S_{11}^* \equiv \Sigma_{11}^*/\sigma_\epsilon^2$. Equations (B.30) and (B.31) then imply that $S_{11} = g(\tilde{s}_\eta^2)$ and $S_{11}^* = g(s_\eta^2)$ for the same function g, and Equations (B.32), (B.33), and $G_1 = G_1^*$ imply that $\tilde{s}_\eta^2 = s_\eta^2$. Equation (B.34) then implies that $\tilde{\sigma}_\epsilon^2 = \sigma_\epsilon^2$, and thus $\tilde{p} = p$.

Suppose next that $\rho\sigma_\eta^2 = 0$, and consider $\tilde{p}$ that satisfies Conditions (i)-(iii). If $\rho = 0$, Equation (B.29) implies that $\tilde{\rho}^{k+1} G_1 = 0$, and Equation (B.34) that $\tilde{\rho}^2 \Sigma_{11} + \tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2 = \sigma_\eta^2 + \sigma_\epsilon^2$. If $\sigma_\eta^2 = 0$, the same implications follow because $\Sigma^* = 0$ and $G^* = [0, 1]'$ from Equations (10) and (11). Equation $\tilde{\rho}^{k+1} G_1 = 0$ implies that either $\tilde{\rho} = 0$, or $G_1 = 0$ in which case $\tilde{\sigma}_\eta^2 = 0$. If $\tilde{\rho} = 0$, then $\tilde{\sigma}_\eta^2 + \tilde{\sigma}_\epsilon^2 = \sigma_\eta^2 + \sigma_\epsilon^2$. If $\tilde{\sigma}_\eta^2 = 0$, then $\tilde{\sigma}_\epsilon^2 = \sigma_\eta^2 + \sigma_\epsilon^2$. Therefore, $\tilde{p}$ is as in the proposition. Showing that all $\tilde{p}$ in the proposition satisfy Conditions (i)-(iii) is obvious. ∎

Proof of Proposition 5: We determine the parameter vectors $\tilde{p}$ that belong to $m(P_0)$ and satisfy

$e(\tilde{p}) = 0$. From Proposition 3 and Lemma 6, these must satisfy (i) $\tilde{\mu} = \mu$, (ii) $CA^kG = C^*(A^*)^kG^*$ for all $k \geq 0$, and (iii) $\sigma_s^2(\tilde{p}) = \sigma_s^{*2}$. Since $\rho\sigma_\eta^2 = 0$, we can write Condition (ii) as

$$\tilde{\rho}^{k+1}G_1 - \alpha(\delta-\alpha)^k G_2 = 0, \tag{B.35}$$

and Condition (iii) as

$$C\Sigma C' + V = \sigma_\eta^2 + \sigma_\epsilon^2. \tag{B.36}$$

Suppose first that $G_2 \neq 0$. Equation (B.35) then implies that $\tilde{\rho} = \delta - \alpha$. Multiplying Equation (11) by an arbitrary 1×2 vector v, and noting that $vA = (\delta-\alpha)A$, we find

$$vG = \frac{(\delta-\alpha)v\Sigma C' + vU}{C\Sigma C' + V}. \tag{B.37}$$

Writing Equation (10) as

$$\Sigma = A\Sigma A - G(C\Sigma A + U') + W,$$

multiplying by v, and noting that $vA = (\delta-\alpha)A$, we find

$$v\Sigma = \left[-vG(C\Sigma A + U') + vW\right][I - (\delta-\alpha)A]^{-1}. \tag{B.38}$$

Substituting $v\Sigma$ into Equation (B.37), we find

$$vG\left[1 + \frac{(\delta-\alpha)(C\Sigma A + U')[I-(\delta-\alpha)A]^{-1}C'}{C\Sigma C' + V}\right] = \frac{(\delta-\alpha)vW[I-(\delta-\alpha)A]^{-1}C' + vU}{C\Sigma C' + V}. \tag{B.39}$$

If v satisfies $vG = 0$, then Equation (B.39) implies that

$$(\delta-\alpha)vW[I-(\delta-\alpha)A]^{-1}C' + vU = 0. \tag{B.40}$$

Since $CG = 0$ from Condition (ii), Equation (B.40) holds for $v = C$:

$$(\delta-\alpha)CW[I-(\delta-\alpha)A]^{-1}C' + CU = 0. \tag{B.41}$$

Since $CG = 0$, Equations (B.36) and (B.38) imply that

$$CW[I-(\delta-\alpha)A]^{-1}C' + V = \sigma_\eta^2 + \sigma_\epsilon^2. \tag{B.42}$$

Substituting for $\tilde{\rho} = \delta - a$, A, C, V, W, and U, we find that the solution $(\tilde{\sigma}_\eta^2, \tilde{\sigma}_\epsilon^2)$ to the system of (B.41)-(B.42) is $\tilde{p} = \tilde{p}_1$.

Suppose next that $G_2 = 0$. Equation (B.40) then holds for $v \equiv (0,1)$ since $vA = (\delta-\alpha)v$ and

$vG = 0$. Solving this equation, we find $\tilde{\sigma}_\epsilon^2 = 0$. The unique solution of Equation (10) then is $\Sigma = 0$, and Equation (B.36) implies that $\tilde{\sigma}_\eta^2 = \sigma_\eta^2 + \sigma_\epsilon^2$. Equation (11) implies that $G_1 = 1$, and Equation $CG = 0$ implies that $\tilde{\rho} = 0$. Therefore, $p = p_2$.

Summarizing, the only parameter vectors that can satisfy Conditions (i)-(iii) are $\tilde{p}_1$ and $\tilde{p}_2$. Showing that $\tilde{p}_2$ indeed satisfies (i)-(iii) is straightforward. Showing the same for $\tilde{p}_1$ follows by retracing the previous steps and noting that the term in brackets in Equation (B.39) is non-zero. Since $\tilde{p}_1$ and $\tilde{p}_2$ are the only parameter vectors to satisfy (i)-(iii), the proposition follows. ∎

Proof of Lemma 2: Equations (6) and (B.1) imply that

$$E_t(s_{t'}) = \tilde{\mu} + CE_t(x_{t'-1}) = \tilde{\mu} + CA^{t'-t-1}E_t(x_t)$$

for $t' > t$. Therefore,

$$\frac{dE_t(s_{t'})}{ds_t} = CA^{t'-t-1}\frac{dE_t(x_t)}{ds_t} = CA^{t'-t-1}G,$$

where the second step follows from the steady-state version of (7). ∎

Proof of Proposition 6: Proposition 5 implies that when $\rho \notin \{0, \delta - \alpha\}$, $e(P_\rho) > 0$. Lemma 5 implies that any element of $m(P_\rho)$ satisfies $\tilde{\mu} = \mu$. Therefore, the proposition will follow if we show that the error $e(\tilde{p})$ is not minimized for $\tilde{s}_\eta^2 \equiv \tilde{\sigma}_\eta^2/\tilde{\sigma}_\epsilon^2 \in \{0, \infty\}$. To compute $e(\tilde{p})$, we note that the assumption $\sigma_\eta^2 = 0$ generates the following sequence of implications: $\Sigma^* = 0$, $G^* = [0,1]'$, $C^*(A^*)^k G^* = 0$ for all $k \geq 0$, $N_k = -CD^kG$ for all $k \geq 0$, and

$$e(\tilde{p}) = (\sigma_s^*)^2 \sum_{k=0}^{\infty} (CD^kG)^2.$$

Consider the behavior of $e(\tilde{p})$ at $\tilde{s}_\eta^2 = 0$. Since $\Sigma = 0$ and $G = [0,1]'$,

$$D \equiv A - GC = \begin{bmatrix} \rho & 0 \\ -\rho & \delta \end{bmatrix},$$

$$D^k = \begin{bmatrix} \rho^k & 0 \\ -\rho\frac{\rho^k - \delta^k}{\rho - \delta} & \delta^k \end{bmatrix},$$

and $CD^kG = -\alpha\delta^k$. Therefore, the derivative of $e(\tilde{p})$ w.r.t. $\tilde{s}_\eta^2$ at $\tilde{s}_\eta^2 = 0$ is

$$\begin{aligned}
\frac{de(\tilde{p})}{d\tilde{s}_\eta^2} &= 2(\sigma_s^*)^2 \sum_{k=0}^{\infty} CD^kG \left(C\frac{dD^k}{d\tilde{s}_\eta^2}G + CD^k\frac{dG}{d\tilde{s}_\eta^2} \right) \\
&= 2(\sigma_s^*)^2 \sum_{k=0}^{\infty} CD^kG \left(-\sum_{k'=0}^{k-1} CD^{k'}\frac{dG}{d\tilde{s}_\eta^2}CD^{k-1-k'}G + CD^k\frac{dG}{d\tilde{s}_\eta^2} \right) \\
&= -2(\sigma_s^*)^2 \sum_{k=0}^{\infty} \alpha\delta^k \left(\sum_{k'=0}^{k-1} CD^{k'}\frac{dG}{d\tilde{s}_\eta^2}\alpha\delta^{k-1-k'} + CD^k\frac{dG}{d\tilde{s}_\eta^2} \right).
\end{aligned}$$

The error $e(\tilde{p})$ is not minimized for $\tilde{s}_\eta^2 = 0$ if the derivative is negative. Since

$$CD^k\frac{dG}{d\tilde{s}_\eta^2} = \left(\rho^{k+1} + \alpha\rho\frac{\rho^k - \delta^k}{\rho - \delta} \right) \frac{dG_1}{d\tilde{s}_\eta^2} - \alpha\delta^k\frac{dG_2}{d\tilde{s}_\eta^2},$$

the derivative is negative if $dG_1/d\tilde{s}_\eta^2 > 0$ and $dG_2/d\tilde{s}_\eta^2 < 0$. To show these inequalities, we differentiate Equations (10) and (11), after writing them in terms of $\tilde{s}_\eta^2$ and $S \equiv \Sigma/\tilde{\sigma}_\eta^2$. Dividing both sides by $\tilde{\sigma}_\epsilon^2$, we can write Equation (10) as

$$S = ASA' - \frac{1}{CSC' + 1 + \tilde{s}_\eta^2}\left[ASC' + (\tilde{s}_\eta^2, 1)'\right]\left[ASC' + (\tilde{s}_\eta^2, 1)'\right]' + \begin{bmatrix} \tilde{s}_\eta^2 & 0 \\ 0 & 1 \end{bmatrix}. \tag{B.43}$$

Likewise, we can write Equation (11) as

$$G \equiv \frac{1}{CSC' + 1 + \tilde{s}_\eta^2}\left[ASC' + (\tilde{s}_\eta^2, 1)'\right]. \tag{B.44}$$

Differentiating Equation (B.43) w.r.t. $\tilde{s}_\eta^2$ at $\tilde{s}_\eta^2 = 0$, we find

$$\begin{aligned}
\frac{dS_{11}}{d\tilde{s}_\eta^2} &= \frac{1}{1-\rho^2}, \\
\frac{dS_{12}}{d\tilde{s}_\eta^2} &= -\frac{1}{(1-\rho^2)(1-\rho\delta)}, \\
\frac{dS_{22}}{d\tilde{s}_\eta^2} &= \frac{1+\rho\delta}{(1-\rho^2)(1-\rho\delta)(1-\delta^2)}.
\end{aligned}$$

Differentiating Equation (B.44), we then find

$$\begin{aligned}
\frac{dG_1}{d\tilde{s}_\eta^2} &= \frac{1-\rho(\delta-\alpha)}{(1-\rho^2)(1-\rho\delta)} > 0, \\
\frac{dG_2}{d\tilde{s}_\eta^2} &= -\frac{(1+\rho\alpha)(1-\delta^2) + \delta\alpha(1+\rho\delta)}{(1-\rho^2)(1-\rho\delta)(1-\delta^2)} < 0.
\end{aligned} \tag{B.45}$$

Consider next the behavior of $e(\tilde{p})$ at $\tilde{s}_\eta^2 = \infty$. Since $\Sigma = 0$ and $G = [1, 0]'$,

$$D \equiv A - GC = \left[\begin{array}{cc} 0 & \alpha \\ 0 & \delta - \alpha \end{array} \right],$$

$$D^k = \left[\begin{array}{cc} 0 & \alpha(\delta - \alpha)^{k-1} \\ 0 & (\delta - \alpha)^k \end{array} \right],$$

$CG = \rho$, and $CD^kG = 0$ for $k > 0$. The derivative of $e(\tilde{p})$ w.r.t. $y \equiv 1/\tilde{s}_\eta^2$ at $y = 0$ is

$$\frac{de(\tilde{p})}{dy} = 2(\sigma_s^*)^2 CGC \frac{dG}{dy} = 2(\sigma_s^*)^2 \rho \left(\rho \frac{dG_1}{dy} - \alpha \frac{dG_2}{dy} \right).$$

The error $e(\tilde{p})$ is not minimized for $\tilde{s}_\eta^2 = \infty$ if the derivative is negative, which is ensured by $dG_1/dy < 0$ and $dG_2/dy > 0$. To show these inequalities, we write Equations (10) and (11) in terms of y and $Y \equiv S/\tilde{\sigma}_\epsilon^2$, and differentiate w.r.t. y. ∎

Proof of Proposition 7: Consider a sequence $\{\alpha_n\}_{n \in \mathbb{N}}$ converging to zero, and an element $\tilde{p}_n \equiv ((\tilde{\sigma}_\eta^2)_n, \rho, (\tilde{\sigma}_\epsilon^2)_n, \mu)$ from the set $m(P_\rho)$ corresponding to α_n. The proposition will follow if we show that $\frac{(\tilde{s}_\eta^2)_n}{\alpha_n} \equiv \frac{(\tilde{\sigma}_\eta^2)_n}{\alpha_n (\tilde{\sigma}_\epsilon^2)_n}$ converges to z and $(\tilde{\sigma}_\epsilon^2)_n$ converges to σ_ϵ^2. Denoting the limits of $((\tilde{\sigma}_\eta^2)_n, (\tilde{\sigma}_\epsilon^2)_n, \frac{(\tilde{s}_\eta^2)_n}{\alpha_n})$ by $(\ell_\eta, \ell_\epsilon, \ell_s)$, the point $(\ell_\eta, \rho, \ell_\epsilon, \mu)$ belongs to Tommy's $m(P_\rho)$. (If the sequences do not converge, we can extract converging subsequences.) Tommy's $m(P_\rho)$ consists of the unique element $(0, \rho, \sigma_\epsilon^2, \mu)$ because only this element belongs in both $m(P_0)$ (Proposition 4, case $\sigma_\eta^2 = 0$) and P_ρ. Therefore, $\ell_\eta = 0$ and $\ell_\epsilon = \sigma_\epsilon^2$.

Recall from Proposition 6 that when $\tilde{s}_\eta^2$ converges to zero, G converges to $[0, 1]'$, CD^kG converges to $-\alpha\delta^k$, and $G_1/\tilde{s}_\eta^2$ converges to $\frac{1-\rho(\delta-\alpha)}{(1-\rho^2)(1-\rho\delta)}$ (Equation (B.45)). Therefore, when $v \equiv (\alpha, \tilde{\sigma}_\eta^2, \tilde{\sigma}_\epsilon^2, \frac{\tilde{s}_\eta^2}{\alpha})$ converges to $\ell_v \equiv (0, 0, \sigma_\epsilon^2, \ell_s)$, CD^kG converges to zero and $G_1/\tilde{s}_\eta^2$ converges to $\frac{1}{1-\rho^2}$. Since

$$\frac{a_k}{\alpha} = -\frac{CA^kG}{\alpha} = -\frac{\rho^{k+1}G_1 - \alpha(\delta - \alpha)^k G_2}{\alpha} = (\delta - \alpha)^k G_2 - \rho^{k+1} \frac{\tilde{s}_\eta^2}{\alpha} \frac{G_1}{\tilde{s}_\eta^2},$$

we have

$$\lim_{v \to \ell_v} \frac{a_k}{\alpha} = \delta^k - \frac{\ell_s}{1 - \rho^2} \rho^{k+1}.$$

Equation (B.28) and $\lim_{v \to \ell_v} CD^kG = 0$ then imply that

$$\lim_{v \to \ell_v} \frac{N_k}{\alpha} = \lim_{v \to \ell_v} \frac{a_k}{\alpha} = \delta^k - \frac{\ell_s}{1 - \rho^2} \rho^{k+1}, \tag{B.46}$$

and thus

$$\lim_{n \to \infty} \frac{e(\tilde{p}_n)}{\alpha_n^2} = \lim_{v \to \ell_v} \frac{e(\tilde{p})}{\alpha^2} = \lim_{v \to \ell_v} \left[\sigma_s^2(\tilde{p}) \frac{\sum_{k=0}^\infty N_k^2}{\alpha^2} \right] = \sigma_\epsilon^2 \sum_{k=0}^\infty \left(\delta^k - \frac{\ell_s}{1 - \rho^2} \rho^{k+1} \right)^2 \equiv \sigma_\epsilon^2 F(\ell_s).$$

Since $\tilde{p}_n$ minimizes the error $e(\tilde{p}_n)$, ℓ_s must minimize the function $F(\ell_s)$. The first-order condition is

$$\sum_{k=0}^{\infty} \rho^k \left(\delta^k - \frac{\ell_s}{1-\rho^2} \rho^{k+1} \right) = 0$$

and implies $\ell_s = z$. ∎

Proof of Proposition 8: Equations (B.6) and (B.7) imply that in steady state

$$\overline{s}_{t'}(\tilde{p}) \equiv E_{t'-1}(s_t) = \tilde{\mu} + C \sum_{t'=1}^{t-1} D^{t'-t-1} G(s_t - \tilde{\mu}),$$

and thus

$$\sum_{t''=t}^{t'-1} \frac{dE_{t'-1}(s_{t'})}{ds_{t''}} = C \sum_{k=0}^{t'-t-1} D^k G. \tag{B.47}$$

When $\sigma_\eta^2 = 0$, $CD^kG = -N_k$. Equation (B.46) then implies that

$$\lim_{\alpha \to 0} \frac{CD^kG}{\alpha} = \left(\frac{z}{1-\rho^2} \rho^{k+1} - \delta^k \right) \equiv \Delta_k,$$

and thus

$$\sum_{t''=t}^{t'-1} \frac{dE_{t'-1}(s_{t'})}{ds_{t''}} = \alpha \Gamma_{t'-t-1} + o(\alpha),$$

where

$$\Gamma_k \equiv \sum_{k'=0}^{k} \Delta_{k'}.$$

Using Equation (15) to substitute z, we find

$$\Delta_0 = \Gamma_0 = \frac{\rho(\delta - \rho)}{1 - \rho\delta},$$

$$\Gamma_\infty = \frac{\rho - \delta}{(1 - \rho\delta)(1 - \delta)}.$$

When $\rho > \delta$, the function Δ_k is negative for $k = 0$ and positive for large k. Since it can change sign only once, it is negative and then positive. The function Γ_k is negative for $k = 0$, then decreases ($\Delta_k < 0$), then increases ($\Delta_k > 0$), and is eventually positive ($\Gamma_\infty > 0$). Therefore, Γ_k is negative and then positive. When $\rho < \delta$, the opposite conclusions hold. ∎

Proof of Proposition 9: Suppose that $\rho \neq \delta - \alpha$, and consider $\tilde{p}$ such that $e(\tilde{p}) = 0$. Lemma 6 implies that for all $k \geq 0$, $CA^kG = C^*(A^*)^kG^*$, i.e.,

$$\tilde{\rho}^{k+1} G_1 - \alpha(\delta - \alpha)^k G_2 = \rho^{k+1} G_1^*. \tag{B.48}$$

Since $\rho \neq \delta - \alpha$, Equation (B.48) can hold only if one of $(\delta - \alpha)^k$ and ρ^k has a zero coefficient. Since $\alpha, \rho, G_1^* > 0$ (the latter because $\sigma_\eta^2 > 0$), this is possible only if $G_2 = 0$. Proceeding as in Proposition 5, we can then show that $\tilde{\sigma}_\epsilon^2 = 0$, and thus $G_1 = 1$. Moreover, Equation (B.48) can hold only if $\tilde{\rho} = \rho$ and $G_1 = G_1^*$, which is a contradiction since $\sigma_\epsilon^2 > 0$ implies $G_1^* \in (0, 1)$.

Suppose next that $\rho = \delta - \alpha$. We can then proceed as in Proposition 5 to construct $\tilde{p}$ such that $e(\tilde{p}) = 0$. ∎

Proof of Proposition 10: Lemma 5 implies that $\tilde{\mu} = \mu$. Consider a sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ converging to zero, and an element $\tilde{p}_n \equiv ((\tilde{\sigma}_\eta^2)_n, \tilde{\rho}_n, (\tilde{\sigma}_\epsilon^2)_n, \mu)$ from the set $m(P_0)$ corresponding to α_n and $(\sigma_\eta^2)_n = \omega\alpha_n$. The proposition will follow if we show that $\frac{(\tilde{s}_\eta^2)_n}{\alpha_n}$ converges to z, $\tilde{\rho}_n$ converges to r, and $(\tilde{\sigma}_\epsilon^2)_n$ converges to σ_ϵ^2. Denoting the limits of $((\tilde{\sigma}_\eta^2)_n, \tilde{\rho}_n, (\tilde{\sigma}_\epsilon^2)_n, \frac{(\tilde{s}_\eta^2)_n}{\alpha_n})$ by $(\ell_\eta, \ell_\rho, \ell_\epsilon, \ell_s)$, the point $(\ell_\eta, \ell_\rho, \ell_\epsilon, \mu)$ belongs to Tommy's $m(P_0)$ (Proposition 4, case $\sigma_\eta^2 = 0$).

Suppose that $\ell_\rho > 0$. Since the only element of Tommy's $m(P_0)$ with persistence parameter ℓ_ρ is $(0, \ell_\rho, \sigma_\epsilon^2, \mu)$, we have $\ell_\eta = 0$ and $\ell_\epsilon = \sigma_\epsilon^2$. Since

$$\frac{a_k}{\alpha} = \frac{C^*(A^*)^k G^* - CA^k G}{\alpha} = \frac{\rho^{k+1}G_1^* + \alpha(\delta-\alpha)^k G_2 - \tilde{\rho}^{k+1}G_1}{\alpha} = \rho^{k+1}\omega\frac{G_1^*}{s_\eta^2} + (\delta-\alpha)^k G_2 - \tilde{\rho}^{k+1}\frac{\tilde{s}_\eta^2}{\alpha}\frac{G_1}{\tilde{s}_\eta^2},$$

the same argument as in the proof of Proposition 7 implies that

$$\lim_{n\to\infty}\frac{e(\tilde{p}_n)}{\alpha_n^2} = \sigma_\epsilon^2 \sum_{k=0}^{\infty}\left(\frac{\omega}{1-\rho^2}\rho^{k+1} + \delta^k - \frac{\ell_s}{1-\ell_\rho^2}\ell_\rho^{k+1}\right)^2 \equiv \sigma_\epsilon^2 H(\ell_\rho, \ell_s).$$

Since $\tilde{p}_n$ minimizes the error $e(\tilde{p}_n)$, (ℓ_ρ, ℓ_s) minimizes the function $H(\ell_\rho, \ell_s)$. Treating this function as one in $(\ell_\rho, \ell_s/(1-\ell_\rho^2))$, the first-order condition w.r.t. the second argument is

$$\sum_{k=0}^{\infty}\ell_\rho^k\left(\frac{\omega}{1-\rho^2}\rho^{k+1} + \delta^k - \frac{\ell_s}{1-\ell_\rho^2}\ell_\rho^{k+1}\right) = 0 \tag{B.49}$$

and w.r.t. the first is

$$\sum_{k=0}^{\infty}(k+1)\ell_\rho^k\left(\frac{\omega}{1-\rho^2}\rho^{k+1} + \delta^k - \frac{\ell_s}{1-\ell_\rho^2}\ell_\rho^{k+1}\right) = 0. \tag{B.50}$$

Computing the infinite sums, we can write Equation (B.49) as

$$\frac{\omega\rho}{(1-\rho^2)(1-\rho\ell_\rho)} + \frac{1}{1-\ell_\rho\delta} - \frac{\ell_s\ell_\rho}{(1-\ell_\rho)^2} = 0 \tag{B.51}$$

and Equation (B.50) as

$$\frac{\omega\rho}{(1-\rho^2)(1-\rho\ell_\rho)^2}+\frac{1}{(1-\ell_\rho\delta)^2}-\frac{\ell_s\ell_\rho}{(1-\ell_\rho)^3}=0. \tag{B.52}$$

Eliminating ℓ_s from Equations (B.51) and (B.52), we find $\ell_\rho = r$. Substituting into Equation (B.51), we find $\ell_s = z$. Thus, if $\ell_\rho > 0$, the limits $(\ell_\rho, \ell_\epsilon, \ell_s)$ equal $(r, \sigma_\epsilon^2, z)$.

We next show that ℓ_ρ cannot be zero. Suppose, by contradiction, that $\ell_\rho = 0$. Denoting the limits of $(\tilde{\rho}_n(G_1)_n/\alpha_n, (G_2)_n)$ by (ℓ_ϕ, ℓ_G), we have

$$\begin{aligned}
\lim_{n\to\infty}\frac{e(\tilde{p}_n)}{\alpha_n^2} &= \sigma_\epsilon^2\left(\frac{\omega}{1-\rho^2}\rho+\ell_G-\ell_\phi\right)^2+\sigma_\epsilon^2\sum_{k=1}^{\infty}\left(\frac{\omega}{1-\rho^2}\rho^{k+1}+\ell_G\delta^k\right)^2\\
&\geq \sigma_\epsilon^2\sum_{k=1}^{\infty}\left(\frac{\omega}{1-\rho^2}\rho^{k+1}+\ell_G\delta^k\right)^2\\
&\geq \sigma_\epsilon^2\sum_{k=1}^{\infty}\left(\frac{\omega}{1-\rho^2}\rho^{k+1}\right)^2,
\end{aligned}$$

where the last step follows because $G \geq 0$ for Tommy. Since $\tilde{p}_n$ minimizes the error, we have

$$H(\ell_\rho,\ell_s)\geq\sum_{k=1}^{\infty}\left(\frac{\omega}{1-\rho^2}\rho^{k+1}\right)^2\equiv\psi,$$

for all (ℓ_ρ, ℓ_s). Simple algebra shows that Equations

$$H(\rho,\omega)\geq\psi$$

and (19) are inconsistent, and so are Equations

$$\min_{\ell_s}H(\delta,\ell_s)\geq\psi$$

and (18). ∎

Proof of Proposition 11: Equation (B.47) implies that

$$\sum_{t''=t}^{t'-1}\frac{dE_{t'-1}(s_{t'})}{ds_{t''}}-\sum_{t''=t}^{t'-1}\frac{dE^*_{t'-1}(s_{t'})}{ds_{t''}}=C\sum_{k=0}^{t'-t-1}D^kG-C^*\sum_{k=0}^{t'-t-1}(D^*)^kG^*.$$

Writing that $N_k = 0$ for the true model, and subtracting this equation from (B.9), we find

$$CD^kG-C^*(D^*)^kG^*=\left[C^*\sum_{k'=0}^{k-1}(D^*)^{k-1-k'}G^*-C\sum_{k'=0}^{k-1}D^{k-1-k'}G\right]C^*(A^*)^{k'}G^*-N_k.$$

Since CD^kG and $C^*(D^*)^kG^*$ converge to zero, while $C^*(A^*)^kG^*/\alpha$ converges to a finite limit, we have

$$\lim_{\alpha\to 0}\frac{CD^kG - C^*(D^*)^kG^*}{\alpha} = -\lim_{\alpha\to 0}\frac{N_k}{\alpha} = -\lim_{\alpha\to 0}\frac{a_k}{\alpha} = \left(\frac{z}{1-r^2}r^{k+1} - \delta^k - \frac{\omega}{1-\rho^2}\rho^{k+1}\right) \equiv \Delta_k$$

and thus

$$\sum_{t''=t}^{t'-1}\frac{dE_{t'-1}(s_{t'})}{ds_{t''}} = \alpha\Gamma_{t'-t-1} + o(\alpha),$$

where

$$\Gamma_k \equiv \sum_{k'=0}^{k}\Delta_{k'}.$$

For $k=0$, the functions Δ_0 and Γ_0 take the value

$$\begin{aligned}\Delta_0 = \Gamma_0 &= \frac{zr}{1-r^2} - 1 - \frac{\omega\rho}{1-\rho^2}\\ &= \frac{1-r^2}{1-r\delta} + \frac{\omega\rho(1-r^2)}{(1-\rho^2)(1-\rho r)} - 1 - \frac{\omega\rho}{1-\rho^2}\\ &= \frac{r^2(\delta-r)(\rho-\delta)}{(1-r\delta)^2} < 0,\end{aligned}$$

where the second step follows from Equation (20), the third from (21), and the last because r is between ρ and δ. A similar calculation shows that

$$\begin{aligned}\Gamma_\infty &= \frac{zr}{(1-r^2)(1-r)} - \frac{1}{1-\delta} - \frac{\omega\rho}{(1-\rho^2)(1-\rho)}\\ &= \frac{(1-r)(\delta-r)(\rho-\delta)}{(1-r\delta)^2(1-\delta)(1-\rho)} < 0.\end{aligned}$$

Since r is between ρ and δ, Δ_k is negative for large k. Since Equation (B.49) can be written as

$$\sum_{k=0}^{\infty} r^k\left(\frac{\omega}{1-\rho^2}\rho^{k+1} + \delta^k - \frac{z}{1-r^2}r^{k+1}\right) = 0 \Leftrightarrow \sum_{k=0}^{\infty} r^k\Delta_k = 0, \tag{B.53}$$

Δ_k has to be positive for some k. Therefore, Δ_k is negative, then positive, and then negative again, because it can change sign at most twice. (The latter is because the derivative of Δ_k/ρ^k can change sign at most once, and thus Δ_k/ρ^k can change sign at most twice.)

The function Γ_k is negative for $k=0$, then decreases ($\Delta_k<0$), then increases ($\Delta_k>0$), then decreases again ($\Delta_k<0$), and is eventually negative ($\Gamma_\infty<0$). Therefore, Γ_k can either be (i) always negative or (ii) negative, then positive, and then negative again. To rule out (i), we write

Equation (B.53) as

$$\Gamma_0 + \sum_{k=1}^{\infty} \left(\Gamma_k - \Gamma_{k-1}\right) r^k = 0 \Leftrightarrow \sum_{k=0}^{\infty} \Gamma_k \left(r^k - r^{k+1}\right) = 0.$$

∎

References

Anderson, B. and J. Moore, 1979, *Optimal Filtering,* Prentice Hall, New York.

Balakrishnan, A., 1987, *Kalman Filtering Theory,* Optimization Software Inc., Publications Division, New York.

Barberis, N., A. Shleifer and R. Vishny, 1998, "A Model of Investor Sentiment," *Journal of Financial Economics,* 49, 307-343.

Camerer, C., 1989, "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review,* 79, 1257-1261.

Edwards, W., 1961, "Probability Learning in 1000 trials," *Journal of Experimental Psychology,* 62, 385-394.

Fama, E., 1991, "Efficient Capital Markets: II," *Journal of Finance,* 46, 1575-1617.

Gilovich, T., R. Vallone and A. Tversky, 1985, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology,* 17, 295-314.

McLeish, D., 1975, "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability,* 3, 829-839.

Mehra, R. and E. Prescott, 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics,* 15, 145-161.

Rabin, M., 2002, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics,* 117, 775-816.

Tversky, A. and D. Kahneman, 1971, "Belief in the Law of Small Numbers," *Psychological Bulletin,* 76, 105-110.

Tversky, A. and T. Gilovich, 1989a, "The Cold Facts about the 'Hot Hand' in Basketball," *Chance,* 2, 16-21.

Tversky, A. and T. Gilovich, 1989b, "The Hot Hand: Statistical Reality or Cognitive Illusion?" *Chance,* 2, 31-34.