Institute for New Economic Thinking

Working Papers
ineteconomics.org/workingpapers

Why Diagnostic Expectations Cannot Replace REH

Roman Frydman* and Halina Frydman **

Working Paper No. 175

January 18th, 2022

ABSTRACT

Gennaioli and Shleifer (GS) have proposed diagnostic expectations (DE) as an empirically-based approach to specifying participants' expectations, which, like REH, can be applied in every model. Beyond its supposedly general applicability, GS's formalization of DE implies that participants systematically and predictably overreact to news. Here, we present a formal argument that Kahneman and Tversky's compelling empirical findings, and those of other behavioral economists, do not provide a basis for a general approach to specifying participants' "predictable errors." We also show that the overreaction of participants' expectations is not a regularity, but rather an artifact of GS's particular specification of DE.

https://doi.org/10.36687/inetwp175

JEL Codes: D80, D84, E71.

Keywords: Rational Expectations Hypothesis, Diagnostic Expectations, Representativeness Heuristic.

* Department of Economics, New York University.

** Department of Technology, Operations and Statistics, Stern School of Business, New York University.

Roman Frydman is a participant in the Program on Knightian Uncertainty Economics at the Institute for New Economic Thinking (INET). The authors thank Anders Rahbek and Joshua Stillwagon for insightful comments on the earlier version of this paper. Roman Frydman is grateful to INET for its support of his research. Correspondence to: Roman Frydman, Department of Economics, New York University, 19 West 4th Street, New York, NY 10003, USA; e-mail rf3@nyu.edu

1 Introduction

An economic model rests on the premise that it formalizes the actual ("objective") uncertainty about market outcomes. Muth (1961, p. 316) proposed a general approach to representing market participants' expectations in macroeconomic and finance models: an economist can specify these expectations as being consistent with his own model's predictions of outcomes. Muth implemented his hypothesis in a model representing outcomes with a time-invariant stochastic process. It is this implementation that has come to be known as the rational expectations hypothesis (REH). Because REH can be used in any model, imposing it has become a standard for specifying participants' expectations.

The *raison d'être* of the behavioral-finance models is that factors such as representativeness, framing, intuition, or market sentiment "distort" participants' assessment of the "objective" uncertainty that they actually face, as specified by an economist's REH model.[1] As a result, according to an economist's model, market participants commit systematic forecast errors.

Lacking a unified approach to represent such errors, early behavioral-finance models formalized them with myriad context-specific psychological insights. Many of these insights seem relevant for understanding how individuals cope with uncertainty. However, as Thaler (2017 p. 489-490) put it in his Nobel lecture, these early behavioral models amounted to formalization of interesting "stories" rather than "a research paradigm" that could, like REH, be applied to specify participants' expectations in any model.

Kahneman and Tversky (KT) hypothesized that a psychological mechanism, which they called the representativeness heuristic, can explain how experimental subjects assess uncertainty in diverse settings.[2] Thaler conjectured

[1] For a seminal review of "framing effects" and related "biases," see Tversky and Kahneman (1981). For an extensive discussion of behavioral-finance models and further references, see Shleifer (2000), Barberis and Thaler (2003), and Gennaioli and Shleifer (2018).

[2] Tversky and Kahneman (1983) provide an extensive review of many experimental findings and further references to the voluminous literature spurred by their seminal 1974 article in *Science*,

that the representativeness heuristic would turn the hodgepodge of early behavioral models into "something resembling [a] science...[of] *predictable* errors" (pp. 489-490, emphasis in the original).

To be sure, a "science of predictable errors," were it possible, would fundamentally alter the foundations of macroeconomics and finance theory. It would replace Muth's hypothesis with a general approach to specifying macroeconomic expectations on the basis of compelling empirical evidence amassed by psychologists and behavioral economists.

However, this paper presents a formal argument that Kahneman and Tversky's findings do not provide a basis for a general approach to specifying market participants' "predictable errors" in macroeconomics and finance models.

Remarkably, Tversky and Kahneman (TK, 1983, p. 313) interpreted their experimental findings as not supporting the "'truth plus error' model, which assumes a coherent [REH-implied] system of beliefs that is perturbed by various sources of distortion and error." As they put it, "perception is not usefully analyzed into a process that produces accurate precepts and a distorting process that produces errors and illusions" (p. 313).

Early on, Kahneman and Tversky (KT, 1972, p. 431) pointed out why one would not expect representativeness to support a general "truth plus error" model of how individuals assess uncertainty: Whether an event is considered representative of another event depends on subjects' interpretation of the context. As a result, "No general definition [of representativeness] is available." Indeed, TK (1983, p. 296) left open the choice of uncertain events that underpin subjects' assessment of representativeness and adopted context-specific assumptions to demonstrate how the heuristic could explain their findings in different experimental settings.

In an influential book, Gennaioli and Shleifer (GS, 2018, p. 11, Chapter 5) sidestep KT's context-dependent interpretation of their findings, arguing that they *do* provide a basis for a general approach to specifying participants' expectations in macroeconomics and finance models. GS proposed a formalization of how the representativeness heuristic distorts participants' expectations, which, like REH, can be applied in any context. Moreover, GS (p.11) argue that, beyond its general applicability, their approach, which they call diagnostic expectations (DE), implies that participants systematically and predictably "overreact to news" about payoff-relevant variables (Bordalo, *et al.*, 2020, p. 2749), regardless of the context and the process driving these variables. This implication leads GS (p.11) to suggest that DE can replace REH in macroeconomic model-building and that, in contrast to earlier behavioral-finance models, it is not subject to Lucas's (1976) critique.[3]

However, as we show here, the overreaction supposedly implied by DE is not a regularity. Rather, it is an artifact of GS's particular specification of DE, which rests on their assumption that how the representativeness heuristic leads participants to *deviate* from REH can be formalized with the *REH-implied* forecast revisions.

Gennaioli and Shleifer's "REH-like" specification of DE represents the representativeness-driven "distortion" of participants' forecasts as being based on an "objective" process driving outcomes, as formalized by an economist's model. Because this "objective" process, according to Muth's model-consistency hypothesis, underpins REH, DE's supposed overreaction, relative to REH, is driven *solely* by news about the payoff-relevant outcomes.

GS's specification of DE-implied distortion as being driven solely by news appears to be at odds with behavioral economists' empirical evidence. For example, the seminal behavioral-finance model of Barberis, *et al.* (1998) analyzes extensive evidence about how psychological influences drive market participants' assessment of uncertainty about stock returns. The representativeness heuristic is one of the main psychological mechanisms underpinning their model's specification of participants' expectations. However, they argue that Kahneman and Tversky's results, and other empirical findings, do not provide a basis for specifying how news about the payoff-relevant variables drives the overreaction or underreaction of participants' expectations. As they put it,

> Unfortunately, the psychological evidence *does not* tell us quantitatively what kind of information is strong and salient (and hence

[3] All citations only to page numbers refer to GS (2018).

> is overreacted to) and what kind of information is low in weight (and hence is underreacted to) (p. 317, emphasis added).

This assessment of empirical evidence stands in sharp contrast to GS's assumption that the quantitative forecast error caused by participants' reliance on the representativeness heuristic can be specified solely in terms of the news. Barberis, *et al.* argue (p. 318) that the deviation of participants' expectations from their REH counterpart arises from "the investor...using the wrong model to form expectations [of returns]."

In this paper, we propose an alternative specification of DE that builds on Barberis, *et al.*'s model. Once we acknowledge the relevance of behavioral economists' findings, DE no longer implies the regularity of overreaction. Depending on the values of the model parameters, and the realizations of payoff-relevant variables, DE overreacts to news about payoff-relevant variables in some periods and underreacts in other periods, relative to the REH-implied forecast.

The plan of the paper is as follows. Section 2 provides a formal overview of the DE approach in the experimental context and highlights the key steps in implementing DE. Section 3 presents GS's formalization of DE in macroeconomics and finance models, including the definition of DE's overreaction (underreaction). Using this definition, Section 4 points out that the supposed regularity of overreaction is an artifact of GS's "REH-like" specification of DE. Section 5 reformulates DE based on Barberis, *et al.*'s (1998) empirically-based specification of participants' expectations and demonstrates that DE no longer implies the regularity of overreaction. Section 6 concludes the paper with a discussion of the difficulties inherent in developing a general approach – one that could replace Muth's hypothesis – to specifying participants' expectations on the basis of behavioral economists' compelling empirical findings.

2 The DE Approach in the Linda Experiment

Kahneman and Tversky formulated the representativeness heuristic to explain their findings in what has come to be known as the Linda experiment. Here, we follow Gennaioli and Shleifer (2018) and provide an overview of the main concepts underpinning their DE approach in the context of the experiment.

The simplicity of the experimental setting enables us to highlight a difficulty overlooked by GS, but which is inherent in any application of the representativeness heuristic in economic models: the operationalization of the heuristic depends on the *subjects*' interpretation of the context within which they assess uncertainty. As KT (1972, p. 431) put it, "Representativeness, like perceptual similarity, is easier to assess than to characterize. In both cases, no general definition is available." Indeed, in explaining their findings in different settings, KT relied on context-specific assumptions regarding how subjects assess the representativeness of uncertain events.[4]

2.1 An Overview of the Linda Experiment

The Linda experiment features a fictitious 31-year-old woman who currently works as a bank teller. As a college student, Linda engaged in "progressive" activities, including opposing discrimination, advocating for social justice, and participating in anti-nuclear demonstrations. We treat the set of 31-year-old women who graduated from college as a population, which we denote with W. We denote the subset of those who engaged in progressive activities while in college with $H^p \subset W$.

Tversky and Kahneman (TK, 1983 p. 297) presented the following statements to their experiment's subjects:

- Linda is a bank teller, places her among individuals in the set $T \subset W$.
- Linda is a bank teller who is also active in the feminist movement (the set F), which places her among the individuals comprising the intersection $T \cap F \subset W$.

Kahneman and Tversky asked the subjects whether it was more or less probable that Linda is among the bank tellers who are also active in the

[4]See TK (1983) for an extensive overview of how they relied on the context-specific operationalizations of representativeness in interpreting their findings in various versions of Linda-like experiments

feminist movement (in $T \cap F$) than that she is among generic bank tellers (in T). An overwhelming majority of subjects responded that it is more probable that Linda is in $T \cap F$ than that she is in T. This finding was then replicated in many Linda-like experiments in a variety of contexts.

2.2 Representativeness in an Experimental Setting

TK (1983, pp. 296-297, 299) hypothesized that their findings could be explained by subjects' reliance on a psychological mechanism, which they called the representativeness heuristic and defined it in terms of the ratio of the relevant frequencies.

Definition 1 *"An attribute is representative of a class if it is very diagnostic, that is, if the relative frequency of this attribute is* much higher *in that class than in a relevant reference class*

This definition leaves open the specific choice of the events operationalizing "the attribute," "the class," and "the reference class." Indeed, in arguing that the representativeness heuristic can explain subjects' responses in different experimental settings, Kahneman and Tversky make specific assumptions regarding the events underpinning the subjects' interpretation of the context within which they make their assessment of uncertainty.

In the context of the Linda experiment, TK consider the event $T \cap F$ as an "attribute," H^p as a "class," and individuals who do not have a history of progressive activities, H^{np}, as a "reference class." The idea underpinning TK's operationalization of Definition 1 was that one would expect feminist bank tellers to be more prevalent among the individuals who, like Linda, have a progressive history, $f(T \cap F|H^p)$, than among the individuals who do not have that history, $f(T \cap F|H^{np})$.[5] It is this apparently much greater prevalence that TK referred to in describing $T \cap F$ as being "very diagnostic" of H^p, which they formalized with $\frac{f(T \cap F|H^p)}{f(T \cap F|H^{np})} >> 1$

[5] $f(T \cap F|H^p) = \frac{n(T \cap F \cap H^p)}{n(H^p)}$, $f(T \cap F|H^{np}) = \frac{n(T \cap F \cap H^{np})}{n(H^{np})}$, and $n(\cdot)$ stands for a number of individuals in a respective set.

We assume that the uncertainty about the events in the Linda experiment can be represented with a probability measure on a space Ω. Thus, we operationalize Definition 1 in terms of the ratio of the conditional probabilities:

$$R(A|C, C^{ref}) = \frac{P(A|C)}{P(A|C^{ref})}, \tag{1}$$

where, all events, $A, C\ C^{ref} \subset \Omega$. In general, $C \cap C^{ref} \neq \varnothing$. However, in the special case of TK's *design* of the Linda experiment, $\Omega = H^p \cup H^{np}$, $A = T \cap F \subset \Omega$, $C = H^p$, and $C^{ref} = H^{np}$. According to Definition 1, A "is representative" of C if it is "very diagnostic," that is, if

$$R(A|C, C^{ref}) >> 1. \tag{2}$$

2.3 Diagnostic Probabilities

GS (pp. 144-152) introduce DE in the context of the Linda experiment. They represent subjects' assessment of uncertainty with a so-called distorted probability measure and specify how representativeness distorts subjective probabilities (p. 148) as follows:

$$P^{DE}(A|C) = P(A|C)\left[R(A|C, C^{ref}))\right]^{\theta} Z, \tag{3}$$

where $P^{DE}(\cdot|\cdot)$ specifies a distorted (subjective) probability on the space Ω, which we refer to as a diagnostic probability, $P(\cdot|\cdot)$ is the "objective" probability, and $\theta > 0$ formalizes the degree of distortion. Z ensures that (3) specifies a well-defined probability. There is no distortion when $\theta = 0$ and $Z = 1$.

2.4 From the Laboratory to Real-World Markets

To operationalize how the representativeness heuristic distorts participants' assessment of uncertainty, an economist would specify the probability distribution of market outcomes (an analog of the attribute $T \cap F$) that he aims to explain in terms of a set of causal variables (an analog of the class H^p), usually called information available to participants. Because any formal economic model rests on the premise that it specifies the "objective" process driving outcomes, an economist, relying on Muth's (1961) hypothesis, can then represent a participant's assessment of uncertainty, and her REH forecasts, with the "objective" process driving outcomes, as specified by the economist's model.

However, as Kahneman and Tversky (1972, p. 431) pointed out neither their experimental findings nor theoretical considerations provide the basis for a "general definition" of the reference class. To be sure, in an experimental setting an investigator would phrase the description of the context to sway the subjects' assessment of representativeness. For example, by providing information to the subjects that Linda has a progressive history, H^p, TK (1983, p. 300) aimed to influence them to compare her to those who do not have that history, thereby considering H^{np} as the relevant reference class.

By contrast, an economist has no way to influence participants' interpretation of the context within which they assess the representativeness of uncertain events in real-world markets. Moreover, there is no basis for specifying an analog of $P(A|C^{ref})$ that could be applied in any model. As we show in the remainder of this paper, this generality is precisely what GS assert for their specification of the reference class and DE in real-world-world markets.

3 Representativeness in Macroeconomics and Finance Models

In contrast to the Linda experiment, the concept of representativeness in macroeconomic and finance models involves continuous random variables. To fix ideas, we consider a payoff-relevant variable $x_{t+1} = \ln \widetilde{x}_{t+1}$, and formalize an "attribute" (an analog of $A = T \cap F$ in (1)) with the measurable event, $x_{t+1} \in A \subset \mathbb{R}^+$, and a "class" with an event $x_t \in C \subset \mathbb{R}^+$. We also operationalize the "reference class" with an event $x_t^{ref} \in C^{ref} \subset \mathbb{R}^+$.

GS (p. 154) define $x_{t+1} \in A$'s representativeness of x_t, relative to x_t^{ref}, in terms of the ratio of conditional probability density functions (pdfs), as follows:

$$R^{gs}(x_{t+1}|x_t, x_t^{ref}) = \frac{f(x_{t+1}|x_t)}{f^{ref}(x_{t+1}|x_t^{ref})} > 1,\ x_{t+1} \in A, x_t \in C, x_t^{ref} \in C^{ref} \quad (4)$$

where $f(x_{t+1}|x_t)$ is the "objective" (conditional) pdf of x_{t+1}, as hypothesized by an economist's model.[6] We refer to $f^{ref}(x_{t+1}|x_t^{ref})$ as a (conditional) reference pdf, which is assumed by an economist to characterize the reference class that participants consider relevant.

Remark 2 *Importantly, GS's specification of the reference pdf,* $f^{ref}(x_{t+1}|x_t^{ref})$. *assumes that it differs from the "objective" pdf,* $f(x_{t+1}|x_t)$, *only by the choice of the conditioning variable,* $x_t^{ref} = x_{t-1}$, *rather than* x_t. *As we point out in Section 4, this assumption is tantamount to assuming that participants overreact to "news," realization of* x_t, *regardless of the context within which they assess* $x_{t+1} \in A$*'s representativeness of* x_t.

TK (1983, p. 296) define representativeness in terms of probabilities (or, equivalently frequencies of discrete events), which for continuous variables can be written as

$$R(x_{t+1}|x_t, x_t^{ref}) = \frac{\int_A f(x_{t+1}|x_t)dx_{t+1}}{\int_A f^{ref}(x_{t+1}|x_t^{ref})dx_{t+1}} > 1,\ x_t \in A, x_t^{ref} \in C^{ref}. \quad (5)$$

However, if the ratio of "objective" and reference pdfs satisfies (4), there exists an event $x_{t+1} \in A$, which is representative of x_t, relative to x_t^{ref}, in the sense that (5) holds.

3.1 Tractable Specification

To render the operationalization in (4) tractable in deriving the testable predictions of macroeconomic and finance models, GS (p. 155) specify the "objective" pdf of x_{t+1}, conditional on x_t as

[6] In addition to x_t, an economist's model typically specifies the conditioning set to include other relevant information (such as realizations of the model's variables) up to time t. Allowing for such a larger information set would not alter any of our conclusions here.

$$f(x_{t+1}|x_t) = \frac{1}{\sigma_{t+1|t}\sqrt{2\pi}} \exp\left[-\frac{(x_{t+1} - m_{t+1|t})^2}{2\left(\sigma_{t+1|t}\right)^2}\right], \ x_{t+1} \in A, x_t \in C, \qquad (6)$$

where $m_{t+1|t}$ and $\left(\sigma_{t+1|t}\right)^2$ denote the conditional mean and the variance. GS (p. 155) also assume that the reference class that underpins participants' assessment of x_{t+1}'s representativeness can be characterized with the normal pdf:

$$f^{ref}(x_{t+1}|x_t^{ref}) = \frac{1}{\sigma_{t+1|t}^{ref}\sqrt{2\pi}} \exp\left[-\frac{(x_{t+1} - m_{t+1|t}^{ref})^2}{2\left(\sigma_{t+1|t}^{ref}\right)^2}\right], \ x_{t+1} \in A, \ x_t^{ref} \in C^{ref}, \qquad (7)$$

where $m_{t+1|t}^{ref}$ and $\left(\sigma_{t+1|t}^{ref}\right)^2$ denote the conditional mean and variance.

3.2 Diagnostic Expectations

Using (4), GS (p. 154) specify the "distorted" pdf of x_{t+1} in the class x_t:

$$f^{de}(x_{t+1}|x_t) = f(x_{t+1}|x_t)\left[R^{gs}(x_{t+1}|x_t, x_t^{ref})\right]^{\theta} Z(\theta, x_t, x_{t-1}), \qquad (8)$$

where, we refer to $f^{de}(x_{t+1}|x_t)$ as the diagnostic pdf, $\theta > 0$, and $Z(\theta, x_t, x_{t-1})$ is specified to ensure that $f^{de}(x_{t+1}|x_t)$ integrates to 1. We denote the conditional mean of a diagnostic density with $m_{t+1|t}^{de}$. GS call $m_{t+1|t}^{de}$ a diagnostic expectation (DE) of x_{t+1}, conditional on x_t.
GS's Proposition 5.1. (p. 155), which we restate here, provides the basis for their assertion that DE implies the regularity of overreaction.

Proposition 3 *Suppose that, as specified in (6) and (7), the "objective," and reference (conditional) pdfs underpinning representativeness, in (4), are normal. Then, provided that* $(1+\theta)\left(\sigma_{t+1|t}^{ref}\right)^2 > \theta\left(\sigma_{t+1|t}\right)^2$, *there exists* $Z(\theta, x_t, x_{t-1})$ *that renders the diagnostic pdf,* $f^{de}(x_{t+1}|x_t)$ *in (8), a well-defined normal pdf* *with the following conditional mean and variance,*

$$m^{de}_{t+1|t} = m_{t+1|t} + \gamma \left(m_{t+1|t} - m^{ref}_{t+1|t} \right), \tag{9}$$

$$\left(\sigma^{de}_{t+1|t}\right)^2 = \frac{\gamma \left(\sigma^{ref}_{t+1|t}\right)^2}{\theta}, \tag{10}$$

where

$$\gamma = \theta \frac{\left(\sigma_{t+1|t}\right)^2}{\left(\sigma^{ref}_{t+1|t}\right)^2 + \theta \left[\left(\sigma^{ref}_{t+1|t}\right)^2 - \left(\sigma_{t+1|t}\right)^2 \right]} > 0. \tag{11}$$

Proof: GS (pp. 217-19).

3.2.1 DE's Overreaction

Adopting Muth's (1961) hypothesis, the conditional mean and variance of the pdf characterizing the REH forecast are the same as their "objective" counterparts, that is, $m^{reh}_{t+1|t} = m_{t+1|t}$ and $\sigma^{reh}_{t+1|t} = \sigma_{t+1|t}$. This implies that $m^{de}_{t+1|t}$ in (9) can be written as

$$m^{de}_{t+1|t} = m^{reh}_{t+1|t} + \gamma \left(m^{reh}_{t+1|t} - m^{ref}_{t+1|t} \right), \tag{12}$$

where γ in (11) is redefined accordingly. GS (p.155) refer to $m^{de}_{t+1|t} > m^{reh}_{t+1|t}$ ($m^{de}_{t+1|t} < m^{reh}_{t+1|t}$) as the "overreaction" of DE, relative to REH.

4 Overreaction as an Artifact of the REH-like Specification of the Reference PDF

Gennaioli and Shleifer proposed DE as a new approach to specifying forecasts in behavioral-finance models that aimed to explain empirical findings that participants' forecasts do not conform to REH. However, their specification of the reference pdf shares a key feature with its REH counterpart: both are based on the "objective" process driving outcomes, as formalized by an economist's model. However, in contrast to the REH forecast, which is conditional on x_t, GS (p. 154) specified the mean of the reference pdf, $m^{ref}_{t+1|t}$ as conditional on x_{t-1}. We state this key assumption of GS's specification of DE as follows:

Assumption 4 *The pdf characterizing the reference class that participants consider relevant is based on the "objective" pdf, which underpins REH, which we formally state as follows*

$$m^{ref}_{t+1|t} = m_{t+1|t-1}. \tag{13}$$

Because $m_{t+1|t-1}$ is a conditional mean of the "objective" pdf, we refer to this choice of the reference pdf as REH-like and denote it with $m^{reh}_{t+1|t-1}$.

Thus, Assumption 4 is equivalent to assuming that the "distorting" influence of the representativeness heuristic on participants' expectations, $m^{reh}_{t+1|t} - m^{ref}_{t+1|t}$, is given by

$$m^{de}_{t+1|t} - m^{reh}_{t+1|t} = \gamma\left(m^{reh}_{t+1|t} - m^{ref}_{t+1|t}\right) = \gamma\left(m^{reh}_{t+1|t} - m^{reh}_{t+1|t-1}\right). \tag{14}$$

Thus, the supposed regularity of overreaction is in fact generated by a well-known property of REH-implied expectations: *by design*, the revision of such an expectation is driven solely by the time-t realization of news about x_t. We refer to GS's specification of DE, in (14), as REH-like.

Example 5 *GS (p.174) illustrate their argument that DE implies what Bordalo et al. (2020) refer to as the* regularity *of "overreaction in macroeconomic expectations" in the context of the following standard AR(1) model,*

$$X_{t+1} = \rho X_t + \mu + \varepsilon_{t+1}, \tag{15}$$

where $0 < \rho < 1$ *and* μ *are constants, and* $\varepsilon_t \sim iidN(0, \sigma^2)$*. Thus, according to Muth's hypothesis,*

$$m^{reh(gs)}_{t+1|t} = E(X_{t+1}|x_t) = \rho x_t + \mu \tag{16}$$

$$= \rho^2 x_{t-1} + (1+\rho)\mu + \rho e_t, \tag{17}$$

$$\left(\sigma^{reh(gs)}_{t+1|t}\right)^2 = \sigma^2, \tag{18}$$

where e_t, in (17), denotes the realization of ε_t.

Furthermore, according to Assumption 4, the mean and the variance of the reference pdf, in (7), are given by

$$\begin{aligned} m_{t+1|t}^{ref(gs)} &= E(X_{t+1}|x_{t-1}) = \\ &= \rho^2 x_{t-1} + (\rho+1)\mu, \qquad (19) \\ \left(\sigma_{t+1|t}^{ref(gs)}\right)^2 &= (1+\rho^2)\sigma^2. \qquad (20) \end{aligned}$$

Because the "objective" and reference pdfs are normal and $\left(\sigma_{t+1|t}^{ref(gs)}\right)^2 > \left(\sigma_{t+1|t}^{reh(gs)}\right)^2$, Proposition 3 holds, which together with Assumption 4, implies that

$$m_{t+1|t}^{de(gs)} - m_{t+1|t}^{reh(gs)} = \gamma^{(gs)} \left(m_{t+1|t}^{reh} - m_{t+1|t-1}^{reh}\right) = \gamma^{(gs)} \rho e_t, \qquad (21)$$

where $\gamma^{(gs)} = \frac{\theta}{(1+\rho^2)(1+\theta)}$, and e_t is the realization of ε_t.

GS (p. 155) refer to $e_t > 0$ ($e_t < 0$) as good (bad) news about the payoff-relevant outcome x_t. Expression (21) shows that the supposed regularity of overreaction, relative to REH, is an artifact of GS's Assumption 4: good (bad) news leads participants to overreact in the same direction and in the proportionately (predictable) magnitude as the REH forecast revision.

5 DE Sometimes Overreacts, Sometimes Underreacts to News

We have pointed out that the supposed overreaction implied by DE reflects GS's REH-like specification of the reference pdf. This specification assumes that the DE-implied "distortion" of participants' expectations, relative to REH, is driven solely by news. In their seminal pre-DE behavioral-finance paper on stock returns, Barberis *et al.* (1998, p. 317) presented extensive empirical evidence that "the psychological evidence does not tell us quantitatively what kind of information" causes market participants overreact or underreact. "For example, it does not tell us how long a sequence of earnings increases is required for its strength to cause significant overpricing. Nor does the evidence tell us the magnitude of the reaction (relative to a true Bayesian) to information that has high strength and weight, or low strength and weight."

Instead, Barberis, *et al.* argue (p. 318) that the deviation of participants' expectations from their REH counterpart arises from "the investor...using the wrong model form expectations [of returns]." While an economist's model assumes that earnings evolve according to a random walk, the investor "thinks that the world moves between two 'states' or 'regimes' and that there is a different model governing earnings in each regime." They formalize this assumption with the two-state stationary Markov chain.

Barberis, *et al.* argue (p. 318) that the deviation of participants' expectations from their REH counterpart arises from "the investor...using the wrong model [of earnings] to form expectations [of returns]." While an economist's model assumes that earnings evolve according to a random walk, the investor "thinks that the world moves between two 'states' or 'regimes' and that there is a different model governing earnings in each regime." Barberis, *et al.* formalize this assumption with the two-state stationary Markov chain.

Our alternative to GS's specification of DE adapts the key premise of Barberis, *et al.* and other pre-DE behavioral models. We assume that participants' assessment of representativeness is implied by the "wrong" reference pdf, rather than being driven solely by news, and show that such an empirically-based reformulation of DE no longer implies the regularity of overreaction. Depending on the values of the parameters of both the REH and reference pdfs, as well as the realizations of x_t, DE overreacts in some periods and underreacts in other periods.

5.1 A Behavioral Markov (BM) Specification of DE

To facilitate comparison with GS's REH-like specification of the reference pdf, we use an AR(1) process, in (15), to characterize the "objective" process driving x_t, which we restate here for convenience,

$$X_{t+1} = \rho X_t + \mu + \varepsilon_{t+1}, \tag{22}$$

However, adapting Barberis *et al.'s* assumption, we specify the reference process that participants consider relevant as being based on the following "wrong" version of the "objective" process in (22):

$$X_{t+1} = \rho_t^{(b)} X_t + \mu_{t+1}^{(b)} + \varepsilon_{t+1}, \tag{23}$$

where "b" in the superscript denotes that $\rho_t^{(b)}$ and $\mu_t^{(b)}$ specify the reference pdf in the behavioral markov (BM) specification of DE. Each of them evolves according to a Markov chain, which switches between two states, $\rho^{(i)}$, and $\mu^{(i)}$ $i = 1, 2$ with the transition probabilities, $p_{12,}$ and q_{12}, respectively and $\varepsilon_t \sim iidN(0, \sigma^2)$. We note that (23) implies that X_t is dependent on ($\mu_{t-j}^{(b)}$, $\rho_{t-j}^{(b)}$) for $j = 1, 2, ...$However, to simplify the presentation, we assume that X_t and ($\mu_{t+j}^{(b)}$, $\rho_{t+j}^{(b)}$) for $j = 0, 1, 2...$are independent.

Macroeconomic and finance models typically constrain the parameters of a Markov chain to remain unchanging over an infinite past and indefinite future. Thus, in the context of these models, the unconditional distribution of ($\mu_t^{(b)}$, $\rho_t^{(b)}$) eventually converges to a steady-state (stationary) probability distribution (Lawler, 2006, p. 15). Following Barberis *et al.*, we make the following assumption:

Assumption 6 *$\rho_t^{(b)}$ and $\mu_t^{(b)}$ are stationary Markov chains. For all t,*

$$P(\mu_t^{(b)} = \mu^{(1)}) = \pi_\mu^{(1)}, P(\mu_t^{(b)} = \mu^{(2)}) = \pi_\mu^{(2)} = \left(1 - \pi_\mu^{(1)}\right), \tag{24}$$

$$P(\rho_t^{(b)} = \rho^{(1)}) = \pi_\rho^{(1)}, P(\rho_t^{(b)} = \rho^{(2)}) = \pi_\rho^{(2)} = \left(1 - \pi_\rho^{(1)}\right) \tag{25}$$

5.1.1 A Mixture Characterization of the Behavioral Reference PDF

Allowing $\mu_t^{(b)}$ and $\rho_t^{(b)}$ to evolve according to a Markov chain implies that, in contrast to GS's Assumption 4, the reference pdf implied by (23) is a mixture of normal pdfs. We state this as a proposition.

Proposition 7 *Suppose that (22) characterizes the process driving x_{t+1}. Then, conditional on x_t, the reference pdf of x_{t+1}, denoted with $g^{ref}(x_{t+1}|x_t)$, is the* *following mixture of the four conditional normal pdfs:*

$$g^{ref}(x_{t+1}|x_t) = \sum_{i,j=1}^{2} \pi_\rho^{(j)} \pi_\mu^{(i)} f^{(i,j)}(x_{t+1}|x_t, \rho_t^{(b)} = \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}), \tag{26}$$

where $\pi_\rho^{(j)}$ *and* $\pi_\mu^{(i)}$, $j, i = 1, 2$, *are components of the respective stationary distributions. Furthermore, the conditional mean and variance of* $g^{ref}(x_{t+1}|x_t)$ *are given by*

$$m_{t+1|t}^{ref(b)} = E(\rho_t^{(b)})x_t + E(\mu_t^{(b)}), \tag{27}$$

$$\left(\sigma_{t+1|t}^{ref(b)}\right)^2 = \sigma^2 + V\left(\rho_t^{(b)}\right)x_t^2 + V\left(\mu_{t+1}^{(b)}\right) \tag{28}$$

where, from Assumption 6, $V(\mu_t) = \pi_\mu^{(1)}\pi_\mu^{(2)}\left(\mu^{(b,1)} - \mu^{(b,2)}\right)^2$ and $V(\rho_t) = \pi_\rho^{(1)}\pi_\rho^{(2)}\left(\rho^{(b,1)} - \rho^{(b,2)}\right)^2$.

Proof: By law of total probability,

$$g^{ref}(x_{t+1}|x_t) = \sum_{i,j=1}^{2} f^{rij}(x_{t+1}, \rho_t^{(b)} = \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}|x_t),$$

and

$$\begin{aligned} f^{rij}(x_{t+1}, \rho_t^{(b)} &= \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}|x_t) \\ &= f^{f^{rij}}(x_{t+1}|\rho_t^{(b)} = \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}, x_t) \\ \times f^{rij}(\rho_t^{(b)} &= \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}|x_t) \\ &= f^{ij}(x_{t+1}|\rho_t^{(b)} = \rho^{(j)}, \mu_{t+1}^{(b)} = \mu^{(i)}, x_t)\pi_\rho^{(j)}\pi_\mu^{(i)}, \end{aligned}$$

where the last step follows from the independence of $\left(\rho_t^{(b)}, \mu_{t+1}^{(b)}\right)$ and X_t. Summing the last expression over (i, j)'s establishes (26).

The expressions for the moments in (27) and (28) follow immediately from (23).

5.1.2 A Behavioral Markov DE May Overreact or Underreact to News

Because the mixture in (26) is a normal pdf, and (18) and (28) show that $\left(\sigma_{t+1|t}^{ref(b)}\right)^2 > \left(\sigma_{t+1|t}^{reh(gs)}\right)^2$, Proposition 3 holds. Thus, the diagnostic expectation implied by GS's specification of the time-invariant REH pdf, and the BM specification of the reference pdf is given by

$$m_{t+1|t}^{de(b)} = m_{t+1|t}^{reh(gs)} + \gamma^{(b)} \left(m_{t+1|t}^{reh(gs)} - m_{t+1|t}^{ref(b)} \right) \tag{29}$$

$$= m_{t+1|t}^{reh(gs)} + \gamma^{(b)} \left\{ \left[\rho - E(\rho_t^{(b)}) \right] x_t + \mu - E(\mu_t^{(b)}) \right\} \tag{30}$$

where, from (11), (18) and (28), $\gamma^{(b)} = \theta \frac{\sigma^2}{\sigma^2 + (1+\theta)\left[V(\rho_t^{(b)}) x_t{}^2 + V(\mu_t)\right]}$.

The expression in (30) shows that, according to the BM specification, whether DE overreacts, relative to its REH counterpart, depends on whether $\left[\rho - E(\rho_t^{(b)})\right] x_t + \mu - E(\mu_t^{(b)}) > 0$. The following lemma states this point explicitly:

Lemma 8 *Suppose that the specification of DE based on (22) and (23) characterizes how the representativeness heuristic leads participants away from forecasting according to REH.*

Letting $\rho - E(\rho_t^{(b)}) > 0$ implies that if and only if

$$x_t > \frac{E(\mu_t^{(b)}) - \mu}{\rho - E(\rho_t^{(b)})}.$$

holds, DE overreacts, that is, $m_{t+1|t}^{reh(gs)} > m_{t+1|t}^{ref(b)}$ holds. Conversely, letting $\rho - E(\rho_t^{(b)}) < 0$ implies that if and only if

$$x_t < \frac{E(\mu_t^{(b)}) - \mu}{\rho - E(\rho_t^{(b)})}.$$

holds, DE underreacts, that is, $m_{t+1|t}^{reh(gs)} < m_{t+1|t}^{ref(b)}$ holds.

This lemma shows that once we acknowledge the behavioral economists' evidence that participants' forecasts are not based on the "objective" process driving outcomes, DE no longer implies the regularity of overreaction. Depending on the values of the model parameters of both the REH and reference pdfs, $\left(\rho, \mu, \sigma^2, \rho^{(b,i)}, \mu^{(b,i)}, \pi_\rho^{(i)}, \pi_\mu^{(i)}\right)$, $i = 1, 2$, and the realizations of x_t, DE overreacts in some periods and underreacts in others periods.

6 Concluding Remarks

Muth's model-consistency hypothesis provides a general approach to specifying market participants' expectations in a macroeconomic or finance model. The difficulty inherent in specifying expectations with model-inconsistent representations is that there are myriad ways in which expectations could differ from the predictions of an economist's model. However, dispensing with Muth's hypothesis, behavioral economists have relied on their compelling empirical evidence that psychological and other non-fundamental factors significantly influence participants' expectations as a basis for specifying expectations in macroeconomic and finance models.

Early behavioral-finance models relied on context-specific empirical findings to formalize non-REH expectations.[7] These models aimed to explain particular historical episodes or market outcomes. However, as GS put it, "it takes *a theory* of expectations, [which can be applied in every context], to replace the existing [REH] theory" (GS, p. 9, emphasis added).

GS (p. 11) have argued that formalizing how the representativeness heuristic "distorts" market participants' assessment of uncertainty provides a general theory of expectations that can replace REH in specifying macroeconomic expectations. However, to claim such generality, GS had to sidestep a problem inherent in any application of representativeness: while REH enables an economist to specify "the objective" pdf of outcomes, as Kahneman and Tversky (1972, p. 431) acknowledged, neither their experimental findings nor theoretical considerations provide a basis for a "general definition" of the reference pdf. As a result, in sharp contrast to REH, "no general definition [of representativeness] is available," which can be applied in every macroeconomic or

[7]See Gennaioli and Sheifer (2018, pp. 9-10) for an extensive discussion and further references.

finance model.[8]

To evade this apparently insuperable problem, GS select the reference pdf to *ensure* that DE is applicable in every model. They assume that, like the pdf underpinning the REH forecast, the reference pdf is also based on the "objective" pdf, as formalized by an economist's model.

Thus, once an economist formulates his model, he does not need to address the problem of specifying the reference pdf that would characterize how market participants interpret the context within which they rely on representativeness to assess uncertainty. As a result, GS's specification of DE determines the representativeness-driven "distortion," and thus participants' expectations, *completely* on the basis of an economist's model. Regardless of the context, GS's REH-like formulation implies predictable "overreaction" to news.

GS's formulation of DE does not seem to be supported by empirical evidence presented in the behavioral-finance literature. In a widely cited paper, Barberis *et al.* (1998, p. 317) argue that Kahneman and Tversky's results, and other empirical findings, do not provide a basis for a "quantitative" specification of how news about payoff-relevant outcomes drives the overreaction or underreaction of participants' expectations.

Although Barberis *et al.* argue (p. 318) that the deviation of participants' expectations from their REH counterpart arises from "the investor...using the wrong model," they do not provide a way to select a particular non-REH specification of expectations from myriad potential alternatives.

In reflecting on the implications of the results of their pathbreaking research program, Tversky and Kahneman (1983, p. 313) cautioned against the presumption that participants commit systematic errors, let alone predictable errors, relative to some standard like REH. Our specification of the reference pdf based on Barberis *et al.*'s extensive examination of empirical evidence provides a formal demonstration of Tversky and Kahneman's interpretation

[8] As we pointed out in Section 2.2, in order to explain subjects' responses in different experimental settings, KT make specific assumptions regarding the "objective" and reference events underpinning the operationalization of the representativeness heuristic.

of their experimental findings.

References

[1] Barberis, Nicholas C., Andrei Shleifer, and Robert Vishny.(1998), "A Model of Investor Sentiment," *Journal of Financial Economics,* 49, 307-343.

[2] Barberis, Nicholas C. and Richard H. Thaler (2003), "A Survey of Behavioral Finance," in Constantinides, George, Harris, Milton and Rene Stulz (eds.), *Handbook of the Economics of Finance*, Amsterdam: North-Holland.

[3] Bordalo, Pedro, Gennaioli, Nicola, Ma, Yueran and Andrei Shleifer (2020), "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–2782.

[4] Gennaioli, Nicola and Andrei Shleifer (2018), *The Crisis of Beliefs: Investor Psychology and Financial Fragility,* Princeton University Press.

[5] Kahneman, Daniel and Amos Tversky (1972), "Subjective Probability: A Judgment of Representativeness," in *Cognitive Psychology,* 430-454.

[6] Knight, Frank H. (1921), *Risk, Uncertainty and Profit*, Houghton Mifflin.

[7] Lawler Gregory F. (2006), *Introduction to Stochastic Processes,* Chapman and Hall/CRC.

[8] Muth, John F. (1961), "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29, 315-335.

[9] Shleifer, Andrei (2000), "Inefficient Markets: An Introduction to Behavioural Finance," Oxford University Press.

[10] Thaler, Richard H. (2017), "From Cashews to Nudges: The Evolution of Behavioral Economics," The Nobel Prize Lecture, The Nobel Foundation.

[11] Tversky, Amos and Daniel Kahneman (1974),"Judgment under Uncertainty: Heuristics and Biases," *Science,* 185, 1124-1131..

[12] Tversky, Amos and Daniel Kahneman (1981),"The Framing of Decisions and the Psyxhology of Choice," *Science*, 211, 453-458.

[13] Tversky, Amos and Daniel Kahneman (1983), "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review*, 90, 293-315.