Discussion Paper No. 1249

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

SMOOTH DIAGNOSTIC EXPECTATIONS

Francesco Bianchi
Cosmin Ilut
Hikaru Saijo

July 2024

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Smooth Diagnostic Expectations*

Francesco Bianchi — Johns Hopkins University, CEPR and NBER
Cosmin Ilut — Duke University, NBER
Hikaru Saijo — UC Santa Cruz

July 4, 2024

Abstract

We show that in the formalization of representativeness (Kahneman and Tversky (1972)) developed by Gennaioli and Shleifer (2010), overreaction and confidence are affected by uncertainty, as a *news effect* interacts with an *uncertainty effect*. In the time series domain, this interaction emerges in a *smooth* version of Diagnostic Expectations (DE). Under smooth diagnosticity, agents overreact to new information. Since new information typically changes not just the conditional mean, but also the conditional uncertainty, changes in uncertainty surrounding current and past beliefs affect the severity of the DE distortion and confidence. Smooth DE implies a joint and parsimonious micro-foundation for key properties of survey data: (1) overreaction of conditional mean to news, (2) stronger overreaction for weaker signals and longer forecast horizons, and (3) overconfidence in subjective uncertainty. An analytical RBC model featuring Smooth DE accounts for overreaction and overconfidence in surveys, as well as three salient properties of the business cycle: (1) asymmetry, (2) countercyclical micro volatility, and (3) countercyclical macro volatility.

JEL Codes: D91, E32, E71.

Keywords: Diagnostic expectations, uncertainty, learning, overreaction, overconfidence

*Emails: francesco.bianchi@jhu.edu, cosmin.ilut@duke.edu, and hsaijo@ucsc.edu. We thank Joel Flynn, Spencer Kwon, Chen Lian, Karthik Sastry, Andrei Shleifer, Stephen Terry, Alonso Villacorta, as well as conference participants at the ASSA 2024 Annual Meeting. Saijo gratefully acknowledges the financial support of the Grants-in-Aid for Scientific Research (JP21H04397) from the Japan Society for the Promotion of Science.

1 Introduction

There has been a growing interest in psychological foundations that enrich models of belief formation in economics. A prominent example is the widely-documented "representative heuristic" of Kahneman and Tversky (1972), which serves as the underpinning of a recent and expanding literature on the paradigm of Diagnostic Expectations (DE). According to this heuristic, when new information arrives, as measured with respect to a reference distribution based on past data, memory selectively recalls more vividly past events that are more associated with, or representative of, that current news. Models of DE formalize the details on how memory retrieval distorts the subjective probability of uncertain events away from its objective, "kernel of truth," frequency (see Bordalo et al. (2022) for an overview).

One immediate manifestation of the kernel of truth logic is that when the new information completely eliminates uncertainty over the variable to be forecasted, there is objectively no room for memory to distort conditional judgements (Gennaioli and Shleifer (2010)). While in existing DE models this logic holds in its extreme version of no conditional uncertainty, the literature has so far largely ignored the fact that representativeness, as formalized in Gennaioli and Shleifer (2010) and Bordalo et al. (2016), also features a deeper and pervasive relation between diagnosticity and uncertainty. Specifically, representativeness implies that the distance between the subjective distorted distribution and its objective counterpart depends on both the amount of information received *and* the precision of the revised objective distribution. Uncertainty affects the strength of the representativeness distortion and leads to variation in confidence, accommodating both cases of under- and overconfidence.

After illustrating the connection between representativeness and uncertainty in a simple, static, two-state example, we argue that in the time series domain this connection manifests itself as a "smoothed" version of Diagnostic Expectations, in which the severity of the DE distortion and confidence depend on conditional uncertainty. As a result, Smooth Diagnostic Expectations (Smooth DE) end up connecting two vastly popular branches of Economics that have largely proceeded in parallel: the Diagnostic Expectations literature and the Uncertainty literature (Bloom (2009, 2014), Bloom et al. (2018), Baker et al. (2024)).

Representativeness and Uncertainty. We first study the relation between uncertainty and representiveness in a simple static two-state categorical distribution. We build on the classic example of inferring the probability of someone having red hair. In this example, representativeness implies that if the person of interest is revealed to be Irish, agents overreact and tend to overstate the probability that the person has red hair. This is because red hair is more representative of the Irish population with respect to the world population. We use this example to emphasize that the strength of the overreaction is not monotonic, but

rather hump-shaped. This is because two channels are at work: A *news channel*, measured by the distance between the current and reference distribution, and an *uncertainty channel*, measured by the entropy of the current distribution.

Suppose that there are only two categories - *red* and *non-red* hair color - and that, before any information is revealed about the particular person, the hair color red has a low probability. As information is revealed that makes the red hair more likely (say, Irish nationality), overreaction initially increases with the conditional probability assigned to the color red. This is because initially the news effect and the uncertainty effect work in the same direction: the color red is more likely *and* there is more uncertainty on the hair color, as an ex-ante low-probability hair color became more likely, but not certain. However, if the new information makes the objective probability cross the point of maximum uncertainty (50:50 in a two-state categorical distribution), the uncertainty effect starts working in the opposite direction. There is now progressively less uncertainty on the hair color and the distance between subjective and objective distributions declines. In the limit, when the probability of red hair goes to 1, uncertainty is fully removed, the uncertainty channel dominates the news channel, and the subjective and objective distribution coincide.

This interaction between the two channels also generate variation in the level of confidence. For small increases in the probability of the color red, the less likely hair color, the distorted distribution features more uncertainty than the objective distribution because, by overreacting, agents bring the perceived distorted probability closer to the point of maximum uncertainty (50:50). Eventually, overconfidence emerges, as agents' overreaction leads to attribute more than .5 probability to the color red even if the objective distribution has not crossed the point of maximum uncertainty. As the probability of the hair color red approaches 1, the severity of overreaction starts declining and eventually goes to zero.

Smooth Diagnostic Expectations. The natural adaptation of representativeness to the time series domain involves a "smoothed" version of Diagnostic expectations. Under Smooth DE the severity of the DE distortion and confidence depend on conditional uncertainty, in a way similar to what we documented for categorical distributions. Agents overreact to new information, defined as the difference between the current information set and a previous information set. Since new information typically changes not just the conditional mean, but *also* the conditional uncertainty, changes in uncertainty surrounding current and past beliefs affect the extent of the DE distortion.

Smooth DE stems from a minimal, but conceptually important change to the baseline DE paradigm developed by Bordalo et al. (2018) (BGS) and it aligns well with the original "representative heuristic" of Kahneman and Tversky (1972). In the BGS formulation of DE, the reference distribution is centered on the conditional mean under the true density formed at

some given past time, but shares the same uncertainty as the true distribution conditional on current information. Instead, since we condition exclusively on the past information set, the reference distribution reflects the level of uncertainty at that past time in which expectations were formed. This is in line with the original formalization of representativeness developed by Gennaioli and Shleifer (2010) and Bordalo et al. (2016) for categorical distributions, in which information sets between current and reference distributions are kept distinct.

Smooth DE is thus built on the key informational difference between conditional (or posterior) and unconditional (prior) distribution information. In this sense, our approach relates to recent work in Bordalo et al. (2020), which features a sampling by similarity framework that under conditional probability assessments yields a result reminiscent of DE. In a similar spirit to Smooth DE, in that setting it is important to keep track of the entire prior distribution, which plays a major role in memory interference. In fact, a version of what we label Smooth DE appears in chapter 5 of Gennaioli and Shleifer (2018). However, despite this early appearance, the growing DE literature has focused on the simplified version proposed in BGS. In this paper, we argue that representativeness features a fundamental connection between memory recall and uncertainty. As a result, Smooth DE has distinctive properties that render it an important point of contact with the uncertainty literature and help to make sense of several stylized facts.

When the current and reference distributions are Normal, the baseline BGS formulation delivers a distorted distribution that is also Normal, but in which only the mean is affected by DE. In comparison, under Smooth DE we uncover two key novel properties of the distorted distribution. First, the severity of the Smooth DE distortion decreases as the current level of uncertainty decreases compared to past uncertainty. Put differently, we obtain a smooth version of the logic expressed by Gennaioli and Shleifer (2010), as now an agent is less prone to overreact to the new information the more precise the current information is with respect to past information. In the limit, as uncertainty is fully resolved by the new information, the distortion vanishes, as in the baseline DE. However, with Smooth DE, the extent of the distortion varies smoothly as current uncertainty increases with respect to past uncertainty, while the baseline DE features a discontinuity once current uncertainty goes to zero.

Second, Smooth DE delivers a disconnect between the *objective* and *subjective* level of uncertainty. This is because under Smooth DE, not only the mean, but also the variance of the DE distribution is distorted. When agents experience a reduction in uncertainty with respect to the reference distribution, agents over-state the precision of their forecasts, leading to *overconfidence*. In other words, in that case the DE distribution features a variance lower than under Rational Expectations (RE). Given that typically events close in the future are easier to predict than events far into the future, agents' beliefs will typically feature such

overconfidence. However, the Smooth DE paradigm can also accommodate *underconfidence* following an increase in uncertainty, like in response to an uncertainty shock (Bloom (2009)).

A parsimonious micro-foundation for survey evidence. As the traditional DE, Smooth DE is characterized by a primitive stochastic environment and two parameters controlling (i) the severity of the distortion, $\theta > 0$, and (ii) the lag of the reference distribution, $J \geq 1$. Thus, Smooth DE makes use of *no additional* degree of freedom. Instead, by allowing the reference distribution to be based only on the information set available at some given past time, the kernel of truth logic endogenously generates predictions for the *effective* distortion. Under Smooth DE, the primitive parameter $\theta > 0$ measures the severity of the DE distortion for a *given* level of relative uncertainty, while the effective severity changes with the relative uncertainty. These disciplined predictions allow Smooth DE to offer a parsimonious micro-foundation for a wide range of stylized facts.

The novel property that the effective overreaction to news is stronger when relative uncertainty is higher helps to account for the stylized survey fact that overreaction increases with the horizon of the survey forecast (see for example Bordalo et al. (2019), d'Arienzo (2020), Bordalo et al. (2020), Augenblick et al. (2021), and Bordalo et al. (2023)). For standard stationary processes the same piece of information is less informative about horizons further in the future. Critically, under Smooth DE this relatively smaller reduction in conditional uncertainty leads to a relatively stronger overreaction to news for longer horizons forecasts, consistent with the stylized findings.

The property that Smooth DE implies a disconnect between subjective and measured uncertainty makes the proposed framework relevant for a separate literature on overconfidence. Recent work documents that in survey data firms (i) *overreact* to news and (ii) are *overconfident* in their subjective forecasts (see, Barrero (2022), Born et al. (2022), and the reviews in Altig et al. (2020) and Born et al. (2022)). While the baseline DE model can account for overreaction, it is silent on overconfidence. Smooth DE can instead account for both these seemingly separate properties since it distorts both the mean and the variance of agents' expectations in a way to typically generate both overreaction and overconfidence.

More broadly, the overreaction and overconfidence properties have been typically studied as two distinct behavioral departures from full rationality (see Barberis (2018) for an overview). While overreaction has been typically the focus on the standard DE literature, a separate literature (including for example De Bondt and Thaler (1995) and Daniel et al. (1998, 2001)) is motivated by extensive psychological evidence for overconfidence and argues that models based on this behavioral property are promising in accounting for asset market puzzles. Our results indicate that Smooth DE can offer a *joint micro-foundation*, based on the representativeness heuristic, of these two-widely documented and studied departures

from standard Bayesian updating.

Business cycle implications. We leverage our theoretical insights to study a parsimonious business cycle model with time-varying uncertainty to illustrate how state-dependent overreaction from Smooth DE generate important cyclical implications. We consider an island economy subject to economy-wide and island-specific TFP shocks. Following Bloom et al. (2018), we assume that the island-specific TFP shocks are subject to time-varying volatility that is negatively correlated with economy-wide TFP innovations. We show that this parsimonious model can account for Barrero (2022)'s survey evidence on overreaction and overconfidence, as well as three key empirical properties of the business cycle: (1) *asymmetry* (recessions are deeper than expansions), (2) *countercyclical micro volatility* (cross-sectional variances of microeconomic variables rise in recessions), and (3) *countercyclical macro volatility* (time-series variances of macroeconomic variables rise in recessions).[1]

First, consider the asymmetry property. A negative economy-wide TFP shock generates higher uncertainty about the island-specific TFP shocks. Hence, agents overreact to the economy-wide TFP shock more than usual, leading to a sharper fall in hours, consumption, and output. In contrast, a positive TFP shock reduces agents' uncertainty, and the rise in economic activity is mild. Second, consider countercyclical micro volatility. In recessions, agents face higher uncertainty, so they overreact to the island-specific TFP and as a result, the cross-sectional variances of island-level hours, output, and consumption increase. Conversely, during expansions, agents' overreactions are milder, and hence the cross-sectional dispersion decreases. Third, consider countercyclical macro volatility. The state-dependent overreaction implies that in recessions, economic activity responds strongly to an economy-wide shock to TFP, while in expansions the responses are more muted. As a result, the aggregate volatility rises in recessions even when there is no change in the volatility of economy-wide shocks. These mechanisms highlight that the micro-level uncertainty and macroeconomic volatility are tightly linked through the agent's state-dependent overreaction. As a result, a novel policy implication emerges: a redistributive policy that reduces idiosyncratic uncertainty could be beneficial for macroeconomic stabilization because it dampens this state-dependent overreaction.

[1]These properties have been extensively documented in the literature. For instance, Neftci (1984), Hamilton (1989), Sichel (1993), McKay and Reis (2008), and Morley and Piger (2012) show macroeconomic asymmetries using various econometric approaches. Bloom (2009), Fernández-Villaverde et al. (2011), Ilut et al. (2018), Jurado et al. (2015), Basu and Bundick (2017), and Bloom et al. (2018) document that volatility or uncertainty rise in recessions at the micro and macro levels.

2 Representativeness and Uncertainty

The building block for our Smooth DE model is the concept of representativeness. Here we directly use the definition of representativeness introduced in recent work on selective memory by Gennaioli and Shleifer (2010), Bordalo et al. (2016), Bordalo et al. (2020). This definition, detailed below, captures formally Tversky and Kahneman (1975) own definition of representativeness: "an attribute is representative of a class if it is very diagnostic; that is, the relative frequency of this attribute is much higher in that class than in a relevant reference class" (p. 296).

To understand how we later build on this definition in a time series dimension, we first consider some simple examples, within a very standard probability environment that is static and discrete. These examples illustrate how representativeness implies that overreaction to the relative frequency is shaped by the entire distribution of attributes, both in the class of interest and the relevant reference class, consistently with the above definition by Tversky and Kahneman (1975). We highlight how the amount of *uncertainty* characterizing these distributions is of particular importance in shaping the effective manifestation of representativeness. This focus on the role of uncertainty will allow us to draw an explicit connection between the building block of representativeness and Smooth DE as developed in Section 3.

Representativeness. Consider a simple environment where (Ω, P) is a discrete probability space, with two random variables defined on this space: X, the trait that the agent seeks to assess, and D, the available data. The agent looks to form the conditional probability of a given trait $\widehat{x} \in X$, given a particular realized data $d \in D$. A Bayesian agent would simply use the conditional probability $P(\widehat{x}|d) = P(\widehat{x} \cap d)/P(d)$. The environment here is static, without a sense of repeated accumulation of information.

In assessing conditional probabilities, the agent subject to the representativeness heuristic distorts the Bayesian belief. Here we follow the formalization of representativeness in Gennaioli and Shleifer (2010), Bordalo et al. (2016), Bordalo et al. (2020). This work explains in detail how representativeness can be intuitively understood as the tendency to overweight representative traits, arising due to limited memory and the fact that representative traits are easier to recall. Formally, the representativeness of a given trait $\widehat{x}$ conditional on a particular data d compared to another conditioning data realization, d^{ref} (referred to as the reference data) is defined as the relative frequency of that trait across the two data groups:

$$rep(\widehat{x}|d, d^{ref}) \equiv \frac{P(\widehat{x}|d)}{P(\widehat{x}|d^{ref})} \tag{1}$$

The conditional belief under representativeness is modelled as a weight that distorts the

underlying absolute frequency of $\widehat{x}$ conditional on the data d

$$P^{\theta}(\widehat{x}|d, d^{ref}) = P(\widehat{x}|d)weight(\widehat{x}|d, d^{ref}). \tag{2}$$

The weight reflects the effect of the relative frequency in equation (1), and is given by

$$weight(\widehat{x}|d, d^{ref}) = \left[rep(\widehat{x}|d, d^{ref})\right]^{\theta} \frac{1}{Z(d, d^{ref})}. \tag{3}$$

where $Z(d, d^{ref})$ is a constant of integration, ensuring that the distorted probability integrates to one:

$$Z(d, d^{ref}) = \sum_{\widehat{x} \in X} P(\widehat{x}|d) \left[rep(\widehat{x}|d, d^{ref})\right]^{\theta} \tag{4}$$

In this formal representation representativeness, the parameter $\theta \geq 0$ measures the extent to which the relative frequency across two conditioning data groups affects judgements. When $\theta = 0$, the agent's memory retrieval is perfect and beliefs collapse to the standard frictionless model. When $\theta > 0$, memory is limited and the agent's judgments are shaped by representativeness.

Illustrations. Our simple illustrations below let Ω be the universe of people, X be different hair colors, and $d \in D$ take on specific Nationalities values, with one of them being a reference $d^{ref} = \{World\}$.[2] We evaluate the effects of representativeness of the event $\widehat{x}$ conditional on a specific Nationality group, compared to the reference group d^{ref}.

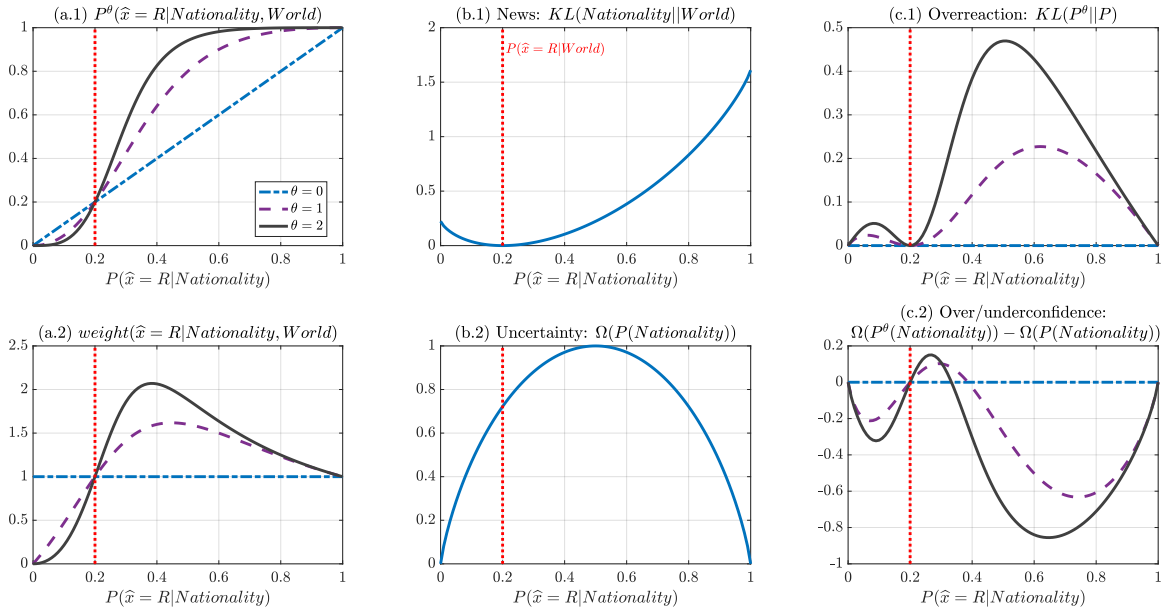
2.1 Two-state distribution

Consider the case where X assume only two values, "red" or "non-red/other", which we denote as $\widehat{x} = R$ and $\widehat{x} = NR$, respectively. The agent observes the nationality of an unknown individual and assess the probability of the hair color red based on $P^{\theta}(\widehat{x} = R|Nationality, World)$ as defined in equation (2). In particular, the agent twists the underlying probability $P(\widehat{x} = R|Nationality)$ by the weight $weight\,(\widehat{x} = R|Nationality, World)$ defined by equations (3) and (4).

We are interested in how the distorted probability *distribution* and its distance (appropriately defined) from the actual conditional distribution depend on the underlying probability distributions. We emphasize the following *three qualitative properties*.

[2]The particular numbers we will use are simply chosen to illustrate a series of key features of representativeness and are not necessarily reflecting the actual incidence of the hair colors in the world population. In fact, we will implicitly assume that there are nationalities for which the probability of the hair color red is 1.

Figure 1: News and uncertainty in a two-states discrete distribution



Notes: The figure displays how, as we vary the actual $P(\widehat{x}=R|Nationality)$, the following objects change: (a.1) the distorted probability of red hair, (a.2) the distorting weight on the red hair probability, (b.1) the news component, measured using the KL defined in equation (5), (b.2) the uncertainty component, measured using the normalized entropy of the actual distribution in equation (6), (c.1) the overall overreaction, measured by the KL divergence of the distorted probability distribution from the actual one, defined in equation (7), and (c.2) the entropy of the distorted distribution relative to the entropy of the actual distribution. We fix the reference probability of a red hair color at $P(\widehat{x}=R|World)$=0.2.

1. **Overreaction**. The distorted probability amplifies the relative change in frequency. This first and most immediate characteristic of representativeness can be immediately seen by equation (2). We show it graphically in Figure 1 (a.1), where we plot the distorted $P^\theta(\widehat{x}=R|Nationality,World)$ for different values of $P(\widehat{x}=R|Nationality)$ and diagnosticity parameter values $\theta=1,2$, as well as $\theta=0$ corresponding to no memory distortion. The frequency under the reference data is $P(\widehat{x}=R|World)=0.2$ marked by a vertical red dotted line. When the frequency of red hair for a given Nationality is higher (lower) than under the reference data, the distorted probability is higher (lower) than the actual frequency for that Nationality. The magnitude of the difference is increasing in θ.

2. **News and uncertainty.** Representativeness is characterized by an effective degree of overreaction that is a function of the entire shape of the distributions for the two data groups that get compared. This property has typically been outside the main interest of

recent applied work on representativeness and diagnostic expectations. However, it will play an important role in our development of Smooth DE.

We separate two channels that create this relation between overreaction and the amount of information present in the conditional distributions that get compared. On the one hand, equation (3) features a *news effect*. When the knowledge of the Nationality implies a larger revision in the frequency of red hair, the representativeness $rep(\widehat{x} = R|Nationality, World)$ changes by more. Holding fixed the constant of integration $Z(Nationality, World)$ in equation (3), the distorting weight is then larger in absolute terms, and therefore there is more overreaction to the news. On the other hand, there is an *uncertainty effect*: the more the knowledge of a given Nationality decreases the uncertainty over the hair color, the less room there is for overreaction to effectively manifest itself. This intuitive force mathematically appears in equation (3) through the constant of integration of equation (4). The news and uncertainty effects are jointly determined by the shapes of distributions getting compared through representativeness. We dissect these effects through a series of plots in Figure 1.

First, Panel (a.1) already shows that the difference between the distorted and underlying probability is *hump-shaped*. This difference is largest for intermediate values between the prior of 0.2 and the lower and upper bounds (0 and 1). By equation (2) the reason for the hump shape is the behavior of the distorting *weight* $(\widehat{x} = R|Nationality, World)$, which gets plotted in Figure 1 (a.2). Indeed, the weight is non-monotonic in the actual frequency: the weight is increasing from an underlying $P(\widehat{x} = R|Nationality) = 0$ to some intermediate value between the reference frequency of 0.2 and the upper bound of 1 and then starts decreasing. Formally, this is because the constant of integration in equation (4) becomes larger as the frequency $P(\widehat{x} = R|Nationality)$ approaches 1. Intuitively, as the actual frequency for a given Nationality becomes more certain towards one hair color, the room for distortion becomes smaller as the total probability has to sum to 1.

Second, we dissect the news and uncertainty channels by using information-theoretic measures. In Panel (b.1), we measure the size of the *news* component. This is captured by the change in the distribution induced by conditioning on a Nationality as opposed to conditioning on World, as measured by the Kullback-Leibler (KL) divergence of the two distributions:

$$KL(\text{Nationality}||\text{World}) = \sum_i P(\widehat{x} = i|Nationality) \ln \left(\frac{P(\widehat{x} = i|Nationality)}{P(\widehat{x} = i|World)} \right). \quad (5)$$

The KL divergence is U-shaped, with $KL = 0$ when the knowing the Nationality does not imply any revision in the distribution.

To measure *uncertainty* for the objective distribution of hair colors for a given Nationality,

Panel (b.2) in Figure 1 plots the normalized entropy

$$\Omega(P(Nationality)) = -\frac{\sum_i P(\widehat{x} = i|Nationality) \ln P(\widehat{x} = i|Nationality)}{\ln(2)}, \tag{6}$$

where the denominator ensures that the normalized entropy is between 0 and 1. This entropy measure of uncertainty is particularly suitable to capture uncertainty in this environment of categorical random variables. The entropy measure shows an *inverted* U-shape and peaks when $P(\widehat{x} = R|Nationality) = 0.5$, the point of maximum uncertainty in this two-state case.

To measure the extent to which overreaction distorts the perceived frequency with respect to the objective frequency, Panel (c.1) reports the KL divergence between the two distributions

$$KL(P^\theta||P) = \sum_i P^\theta(\widehat{x} = i|Nationality, World) \ln\left(\frac{P^\theta(\widehat{x} = i|Nationality, World)}{P(\widehat{x} = i|Nationality)}\right) \tag{7}$$

We can now see how the degree of overreaction shows two humps: a smaller one that peaks at the intermediate value between 0 and reference frequency $P(\widehat{x} = R|World) = 0.2$ and a larger one that peaks at the intermediate value between that reference of 0.2 and the upper bound of 1. The two humps are the result of the *interaction between news and uncertainty effects*, which affect overreaction, but display opposing patterns (U- and inverted U-shape). The particular sizes of the two humps are different because in panels (b.1) and (b.2) the trough of the news component and the peak of the uncertainty effect do not coincide.

3. **Over/under confidence**. The overreaction to information also affects the amount of uncertainty perceived by the agent compared to the actual frequency. Panel (c.2) reports the entropy $\Omega(P^\theta(Nationality))$ of the distorted distribution normalized with respect to the entropy $\Omega(P(Nationality))$ of the actual frequency. A relative entropy smaller than 0 implies overconfidence; the agent is too certain about the hair color compared to the objective distribution. The interaction between the news effect and uncertainty effect leads to an interesting pattern.

First, there is no over- or underconfidence if there is no news (i.e. the distributions conditional on Nationality or World coincide) or the news completely removes uncertainty ($P(\widehat{x} = R|Nationality) = \{0, 1\}$). Second, when the frequency of red hair is smaller for a given Nationality than for the World, news and uncertainty effects work in the same direction, and only over-confidence is possible. Third, when instead the relative frequency is larger, the two effects can work in opposite directions. On the one hand, uncertainty keeps increasing until the frequency is .5, at which point uncertainty starts decreasing again. As a result, for

small increases in the relative frequency of red hair, over-reaction to that relative frequency can lead to underconfidence. Eventually, the news effects starts dominating, leading to overconfidence. However, the strength of overconfidence declines as the news starts to fully remove uncertainty. The degree of overconfidence is more severe for an actual frequency on the right tail ($P(\widehat{x} = R|Nationality) > 0.8$) than on the left tail ($P(\widehat{x} = R|Nationality) < 0.2$). This is because the news effect is larger for a frequency on the right of a small probability event such as the one considered in this example.

2.2 Three-state distribution

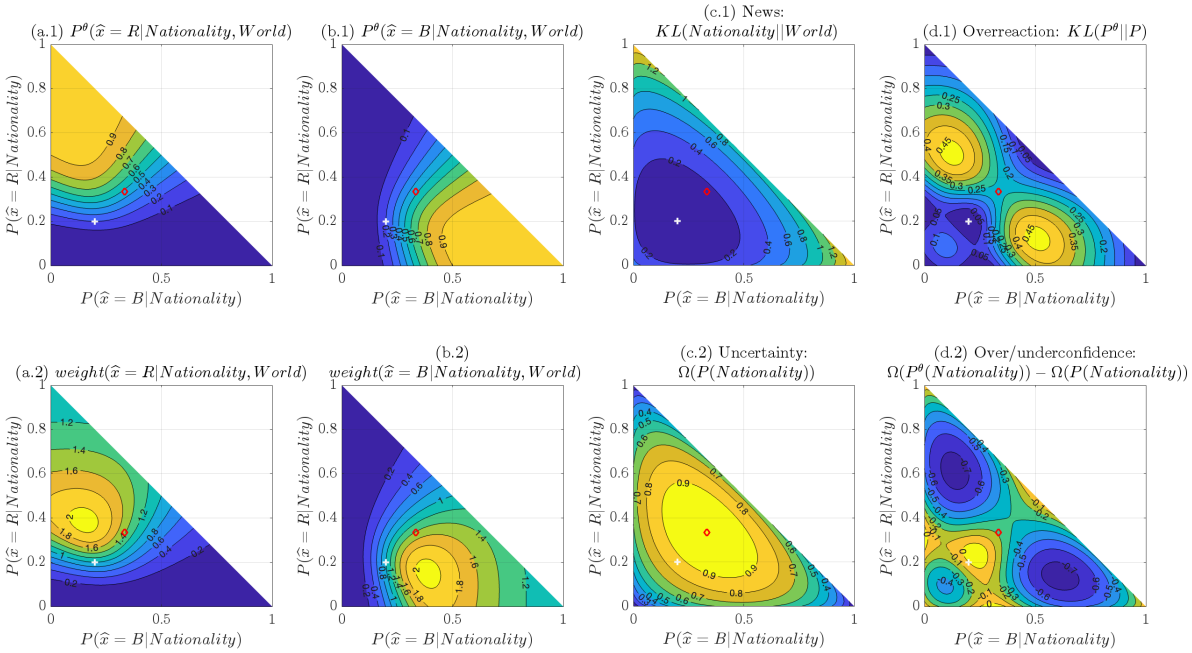
Consider now an extension of the event space to three hair colors - "red," "blond," and "dark," denoted by R, B, and D, respectively. We consider a symmetric reference frequency for R and B so that conditional on World, their frequencies are equal to 0.2, while for D it is equal to 0.6. This three-state example is instructive because the actual frequencies of R and B for a given Nationality jointly determine the news and uncertainty component, thus underscoring the importance of the *shape* of the joint probability *distribution* in affecting the degree of overreaction.

In Figure 2, we use contour plots to examine how news and uncertainty affect overreaction as we vary $P(\widehat{x} = R|Nationality)$ and $P(\widehat{x} = B|Nationality)$ when $\theta = 2$. To ease interpretation, we mark the coordinates corresponding to the reference frequencies with white pluses, and the coordinates representing the equal probabilities of 1/3 for all three outcomes with red diamonds. We can recover the three properties discussed for the two-state example:

1. **Overreaction**. Panels (a.1) and (a.2) report the distorted probability and the weight associated with the red hair color. As in the two-states example, the distorted probabilities overreact to the relative frequency of red hair for the given Nationality compared to the World. The distorting weight of the red probability peaks when $P(\widehat{x} = R|Nationality)$ takes an intermediate value between 0.2 and 1 and $P(\widehat{x} = B|Nationality)$ takes an intermediate value between 0 and 0.2. Similarly, Panel (b.1) and (b.2) report the distorting probability and the weight associated with a blond hair color.

2. **News and Uncertainty**. The three state example is particularly useful here to see how the shape of the distributions matter. Panel (c.1) displays the news component, measured again as the KL(Nationality||World) of equation (5). The news component increases as the two distributions differ and peaks when either of the hair color outcomes reaches certainty ($P(\widehat{x} = R|Nationality) = 1$ or $P(\widehat{x} = B|Nationality) = 1$). The uncertainty component, Ω(Nationality), reported in Panel (c.2) peaks when the distribution for a given Nationality assigns equal probabilities for all three outcomes.

Figure 2: News and uncertainty in a three-states discrete distribution



Notes: The figure displays how, as we vary the probability of red and blond hair colors for a given Nationality, the following objects change: (a.1) the distorted probability of red hair, (a.2) the distorting weight on the red hair probability, (b.1) the distorted probability of blond hair, (b.2) the distorting weight on the blond hair probability, (c.1) the news component, measured using the KL defined in equation (5), (c.2) the uncertainty component, measured using the normalized entropy of the actual distribution in equation (6), (d.1) the overall overreaction, measured using the KL divergence of the distorted probability distribution from the actual one, defined in equation (7), and (d.2) the entropy of the distorted probability distribution relative to the entropy of the actual distribution. We mark the coordinates corresponding to the reference probabilities with white pluses, and those representing equal probabilities for all three outcomes with red diamonds.

Panel (d.1) reports the $KL(P^\theta||\mathrm{P})$ divergence of equation (7). The overreaction displays twin peaks resembling the double humps in the two-states example. However, in the current three-state example how the actual frequencies are split matters for the overreaction. Indeed, the overreaction is mild as we increase both the actual probabilities of red and blond hair simultaneously in equal proportions. This is because both hair colors are becoming more representative for the given Nationality compared to the World.

3. **Over/under confidence**. Finally, Panel (d.2) reports the entropy $\Omega(P^\theta(Nationality))$ with respect to the underlying entropy $\Omega(P(Nationality))$. The former is smaller in most regions, indicating overconfidence, except for the neighborhood of the reference frequency

and the coordinate representing equal probabilities for all three outcomes. Overconfidence is most severe around the three corners of the triangle, when the actual frequency puts most probability weights on one hair color. Intuitively, as one specific hair color becomes more representative, the agent overestimates the probability of that hair color and becomes too certain. As before, overconfidence eventually disappears as uncertainty is fully removed.

To summarize, both the two-and three-states example underscore a nuanced but tight connection between news/uncertainty effects and overreaction/overconfidence.

3 Smooth Diagnostic Expectations

In this section, we show that the key principles uncovered for categorical distributions naturally find their counterparts under Smooth DE, which allows for continuous distributions and accumulation of information. In particular, in order to model the representativeness heuristic into time-series, we mirror the logic of equation (1), and interpret groups as different information sets that become available over time.

3.1 Smooth DE density: representativeness and information sets

We introduce the definition of representativeness and information sets with some generality. Consider the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ where $(\mathcal{F}_t)_{t\geq 0}$ is an increasing family of sub-σ-algebras of $\mathcal{F}$, with $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. Here the filtration $(\mathcal{F}_t)_{t\geq 0}$ represents the evolution of information over time. A stochastic process $(X_t)_{(t\geq 0)}$ on $(\Omega, \mathcal{F}, P)$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if, for each $t \geq 0$, X_t is F_t-measurable. Let the conditional density function of X_{t+h}, for some horizon $h \geq 1$, given the filtration $\mathcal{F}_t$, be denoted by $f(x_{t+h}|\mathcal{F}_t)$. This function describes the probability density of X_{t+h}, given all the information available up to time t.

In our approach, the counterpart of equation (1) is obtained by defining the reference group as a past available information set. As information flows and is accumulated, representativeness is then defined as follows:

Definition 1 ***(Representativeness and information sets)*** *Given the current information set $\mathcal{F}_t$ and a reference information set $\mathcal{F}_{t-J}$, formed $J \geq 1$ periods ago, the representativeness of a random event $\widehat{x}_{t+h}$ for some future horizon $h \geq 1$ is defined as*

$$rep\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right) \equiv \frac{f\left(\widehat{x}_{t+h}|\mathcal{F}_t\right)}{f\left(\widehat{x}_{t+h}|\mathcal{F}_{t-J}\right)} \tag{8}$$

Mirroring the language of Tversky and Kahneman (1975) in Section 2, in the time-series domain an event is representative for the current information set if the relative frequency of

this event is higher conditional on the current information set than conditional on some past reference information set.[3]

Based on this definition of the representativeness of an event, we then follow the same construction of the distorted belief as introduced early in equation (2) in Section 2. In particular, we build the following distorted conditional density, which we refer to as *Smooth Diagnostic Expectations*, or in a more abbreviated form as Smooth DE.

Definition 2 ***(Smooth Diagnostic Expectations)****. In the time series domain, the conditional density $f^{\theta}\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right)$, distorted by representativeness as defined in Definition 1, is constructed as follows*

$$f^{\theta}\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right) = f\left(\widehat{x}_{t+h}|\mathcal{F}_t\right)\left[rep\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right)\right]^{\theta}\frac{1}{Z} \tag{9}$$

where Z is a constant of integration and the parameter $\theta \geq 0$ measures the distortion severity.

Our formalization of Smooth DE strongly connects to that of Diagnostic Expectations (DE) introduced in Bordalo et al. (2018) (BGS) to incorporate representativeness in time-series. As we detail further below, the only, but consequential difference is in the construction of the reference group in the definition of representativeness. Given such a definition, the functional form for the distorted density in equation (9) is identical to DE.

3.2 Smooth DE with Normal densities

In the rest of this paper we follow BGS and focus on normal densities. As we show below, this focus leads to significant gains in tractability and in the range of possible applications. This is because when the true conditional density is Normal, expression (9) delivers a distorted distribution that is also Normal. However, it is important to emphasize that while this a natural point of departure, Smooth DE can easily accommodate alternative distributions, as illustrated in Section 2.

With Normal densities, two moments, the conditional mean $\mu_{t+h|s}$ and conditional variance $\sigma^2_{t+h|s}$, summarize the whole conditioning information:

$$f\left(\widehat{x}_{t+h}|\mathcal{F}_s\right) = \mathcal{N}\left(\widehat{x}_{t+h}; \mu_{t+h|s}, \sigma^2_{t+h|s}\right), \quad \text{for} \quad s \leq t$$

It follows that given Definition 1, the representativeness of the event $\widehat{x}_{t+h}$ is evaluated as

$$rep\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right) = \frac{N(\widehat{x}_{t+h}; \mu_{t+h|t}, \sigma^2_{t+h|t})}{N(\widehat{x}_{t+h}; \mu_{t+h|t-J}, \sigma^2_{t+h|t-J})}, \tag{10}$$

[3]The approach is easily extended when the available information is incomplete or noisy - see Section E.

capturing its relative frequency under the current information set $\mathcal{F}_t$, compared to some past information set $\mathcal{F}_{t-J}$, with $J \geq 1$.

Under the representativeness characterizing normal densities in equation (10), the Smooth DE density in equation (9) has a closed-form solution given by Proposition 1 below.

Proposition 1 *Denote the ratio of variances for the current and reference groups as*

$$R_{t+h|t,t-J} \equiv \sigma^2_{t+h|t} / \sigma^2_{t+h|t-J} \tag{11}$$

If $R_{t+h|t,t-J} < (1+\theta)/\theta$*, the Smooth DE density* $f^{\theta}\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right)$ *in equation (9) is Normal with conditional mean*

$$\mathbb{E}^{\theta}_t\left(x_{t+h}\right) = \mu_{t+h|t} + \theta \frac{R_{t+h|t,t-J}}{1+\theta\left(1-R_{t+h|t,t-J}\right)}\left(\mu_{t+h|t} - \mu_{t+h|t-J}\right) \tag{12}$$

and conditional variance

$$\mathbb{V}^{\theta}_t\left(x_{t+h}\right) = \frac{\sigma^2_{t+h|t}}{1+\theta\left(1-R_{t+h|t,t-J}\right)} \tag{13}$$

Proof. See Appendix. ■

The condition $R_{t+h|t,t-J} < (1+\theta)/\theta$ guarantees that the variance of the resulting distorted Normal distribution is finite and positive. As the ratio of conditional variances between the current and reference distribution approaches this limiting value, the variance of the Smooth DE distribution approaches infinity and the corresponding Normal distribution approaches a degenerate, flat distribution. Thus, the condition requires that the current uncertainty with respect to a future event ($\sigma^2_{t+h|t}$) is not too high with respect to the past uncertainty about the same event ($\sigma^2_{t+h|t-J}$). The condition typically holds in stationary environments with homoskedastic innovations in which events closer into the future are easier to predict than events far into the future. However, the condition also allows for the possibility of an increase in uncertainty, for example as a result of heteroskedasticity, as long as the increase is not too large with respect to the DE distortion.[4]

[4]In Appendix B we discuss an approach that deals with this threshold condition by implementing an upper bound on the Smooth DE overreaction in the conditional mean. The approach guarantees that both the mean and variance distortions remain finite and non-decreasing as the ratio $R_{t+h|t,t-J}$ goes to infinity.

3.3 Three key properties

Smooth DE is characterized by three important properties. To understand them, it is helpful to define the effective overreaction of the conditional mean to news in equation (12) as

$$\widetilde{\theta}_{t,t-J} \equiv \theta \frac{R_{t+h|t,t-J}}{1+\theta\left(1-R_{t+h|t,t-J}\right)}. \tag{14}$$

Corollary 1 *Assume the presence of residual uncertainty with respect to a future event:* $\sigma^2_{t+h|t} > 0$. *Compared to the RE forecast* ($\theta = 0$), *the conditional forecast under Smooth DE* ($\theta > 0$), *characterized in Proposition 1, exhibits*

1. *overreaction of the conditional mean to new information, since*

$$\widetilde{\theta}_{t,t-J} > 0 \tag{15}$$

2. *an effective overreaction of the conditional mean to new information that is monotonically increasing in the ratio* $R_{t+h|t,t-J}$ *between current and past uncertainty*

$$\frac{\partial \widetilde{\theta}_{t,t-J}}{\partial R_{t+h|t,t-J}} > 0 \tag{16}$$

3. *overconfidence when* $R_{t+h|t,t-J} < 1$, *since then by equation (13)*

$$\mathbb{V}_t^{\theta}\left(x_{t+h}\right) < \sigma^2_{t+h|t} \tag{17}$$

or underconfidence when $R_{t+h|t,t-J} > 1$, *since then by equation (13)*

$$\mathbb{V}_t^{\theta}\left(x_{t+h}\right) > \sigma^2_{t+h|t}. \tag{18}$$

On the one hand, these properties mirror the qualitative ones for the static, discrete state example introduced and discussed in Section 2.1 - properties labelled there 'overreaction', 'news and uncertainty', and 'over/under confidence', respectively. The formalism and closed-form solution of Normal densities make these properties particularly transparent. This connection is useful, indicating that Smooth DE maintains similar insights when extending the concept or representativeness and over-reaction to new information to dynamic models. Additionally, the tractability of the Normal distribution in the time series domain will allow us to connect these three properties to stylized survey facts in Section 4.

The first property in Corollary 1, overreaction of the conditional mean, is an immediate

manifestation of the amplification of the relative frequency of an event that becomes more likely. This property is shared with the standard DE approach, as we detail further below. The other two properties are novel and reflect the tight connection between uncertainty and the degree of over-reaction, in line with the analysis presented in Section 2.

In particular, like in the static, discrete case, the *shapes* of the current and reference densities matter for the effective degree of overreaction. In the Normal case, the shape is summarized only by the conditional uncertainty. Indeed, Smooth DE captures formally, by equation (16), how the severity of the mean distortion *increases* as the ratio $R_{t+h|t,t-J}$ of today's uncertainty to past uncertainty increases. Smooth DE thus micro-founds in a time-series domain an inverse smooth link between overreaction of conditional mean to news and the precision of the news compared to the reference distribution. This is captured by the ratio $R_{t+h|t,t-J}$: The more the new information reduces uncertainty $\sigma^2_{t+h|t}$ compared to $\sigma^2_{t+h|t-J}$, the lower is the role of memory in distorting probability judgements, and thus the lower is the effective overreaction to news. Conversely, everything else equal, the larger today's uncertainty, the larger the Smooth DE distortion.

Similar to the basic intuition of the static, discrete case, Smooth DE also has important implications for the level of subjective *confidence* that agents express with respect to their expectations. As summarized in Corollary 1, if agents experience a reduction of uncertainty with respect to the reference distribution, so that $R_{t+h|t,t-J} < 1$, Smooth DE implies *overconfidence*, i.e. agents overstate the precision of their expectations. Under this scenario, independently of the direction and size of the mean distortion, agents are overconfident about the precision of their expectations. If agents do not experience a change in uncertainty, and $R_{t+h|t,t-J} = 1$, we do not observe a change in confidence with respect to $\sigma^2_{t+h|t}$. Finally, if agents experience an increase in uncertainty, so that $R_{t+h|t,t-J} > 1$, they will be less confident than under the true density.

3.4 Standard DE

The standard approach to apply representativeness in the time-series domain follows the BGS formulation, in which the representativeness of $\widehat{x}_{t+h}$ is given by the relative frequency

$$rep^{BGS}\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right) \equiv \frac{N(\widehat{x}_{t+h}; \mu_{t+h|t}, \sigma^2_{t+h|t})}{N(\widehat{x}_{t+h}; \mu_{t+h|t-J}, \sigma^2_{t+h|t})}, \tag{19}$$

In equation (19), the reference density uses the mean conditional on the information $\mathcal{F}_{t-J}$ available J periods ago, $\mu_{t+h|t-J}$, but the uncertainty conditional on the current information set $\mathcal{F}_t$, $\sigma_{t+h|t}$. The reference density can thus be understood as a mixture of information sets, meant to keep the conditional uncertainty the same for the current and reference density.

Instead, under Smooth DE, the reference density captures entirely and solely the role of new information, as the difference between information sets $\mathcal{F}_t$ and $\mathcal{F}_{t-J}$. Under Normality, comparing the definition of representativeness in equations (10) and (19), this difference formally manifests itself in the *variance* of the reference distribution.

This BGS assumption delivers the following closed form standard DE density:

Proposition 2 *(BGS implementation for DE). Consider the BGS definition of representativeness in equation (19). When $\sigma^2_{t+h|t} > 0$, the resulting DE density $f^\theta\left(\widehat{x}_{t+h}|\mathcal{F}_t, \mathcal{F}_{t-J}\right)$ defined by equation (9) has a Normal distribution with mean:*

$$\mathbb{E}^\theta_t\left(x_{t+h}\right) = \mu_{t+h|t} + \theta\left[\mu_{t+h|t} - \mu_{t+h|t-J}\right]. \tag{20}$$

and variance:

$$\mathbb{V}^\theta_t\left(x_{t+h}\right) = \sigma^2_{t+h|t}. \tag{21}$$

When $\sigma^2_{t+h|t} = 0$, the DE conditional mean $\mathbb{E}^\theta_t\left(x_{t+h}\right)$ collapses to $\mu_{t+h|t}$.

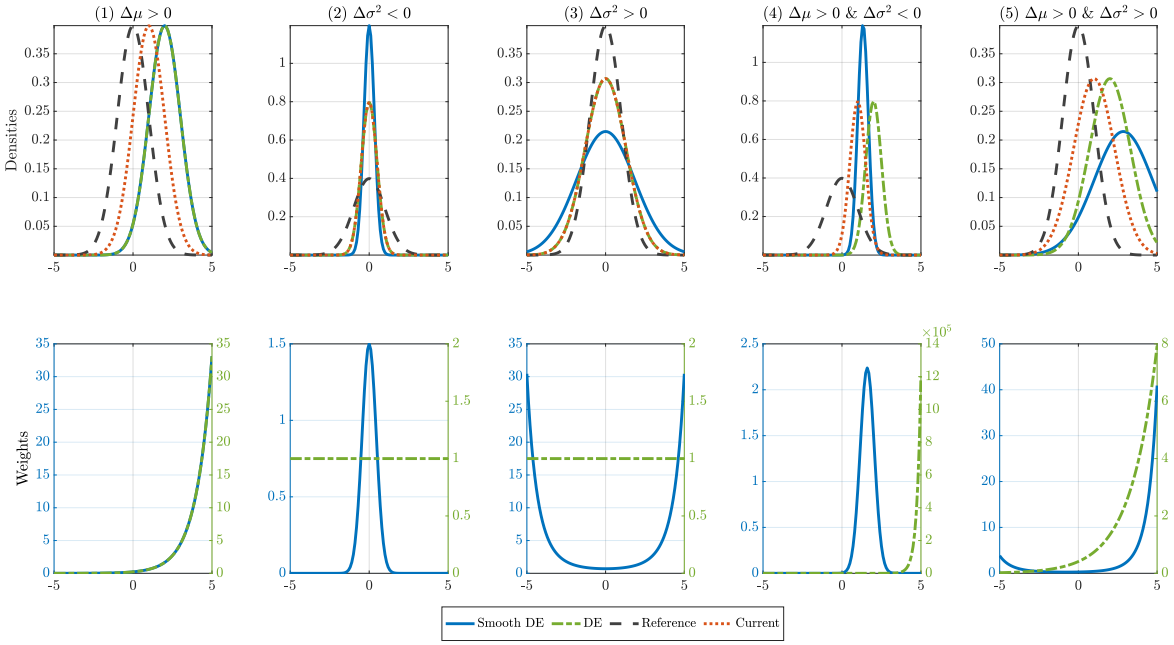
Proof. See Bordalo et al. (2018). ■

Smooth DE and the standard BGS formulation of DE coincide in two cases. First, in the limit of no conditional uncertainty. In particular, in both approaches, with $\theta > 0$, there is a distortion if and only if the conditional variance $\sigma^2_{t+h|t} > 0$. Intuitively, when $\sigma^2_{t+h|t} = 0$, the conditional likelihood of observing any other scenario for x_{t+h} than the one the agent is now fully informed on has become equal to zero. As noted by Gennaioli and Shleifer (2010), the lack of such conditional (or "residual") uncertainty leaves no room for memory to distort conditional forecasts. According to Proposition 1, Smooth DE formally nests that limiting possibility, which would amount to $R_{t+h|t,t-J} = 0$ and thus effectively no distortion even if $\theta > 0$. In the BGS formulation that limit is instead imposed through a discontinuity at $\sigma^2_{t+h|t} = 0$: in the language developed in Bordalo et al. (2018), to compute $\mathbb{E}^\theta_t\left(x_{t+h}\right)$ the realization x_{t+h} constitutes its infinitely representative state (see appendix in Bordalo et al. (2018) on Corollary 1), and the result is $\mathbb{E}^\theta_t\left(x_{t+h}\right) = \mu_{t+h|t}$. Under Smooth DE the effective overreaction $\widetilde{\theta}_{t,t-J}$ in equation (14) *smoothly* goes to zero as current uncertainty goes to zero.

Second, and more importantly, away from the zero conditional uncertainty case, in the original BGS formulation the ratio $R_{t+h|t,t-J}$ is always 1 by assumption. Thus, DE coincides with Smooth DE if and only if the stochastic process is characterized by information sets $\mathcal{F}_t$ and $\mathcal{F}_{t-J}$ that happen to deliver $\sigma^2_{t+h|t} = \sigma^2_{t+h|t-J}, \forall t$ and for any given J. Indeed, if $R_{t+h|t,t-J} = 1$ in Proposition 1 the effective overreaction $\widetilde{\theta}_{t,t-J} = \theta$, and formulas (12) and (13) collapse to their respective counterparts in equations (20) and (21).

In Figure 3 we use a series of illustrative examples to show the different effects at work

Figure 3: Smooth Diagnostic Expectations and standard Diagnostic Expectations densities



Notes: The figure is obtained using the following parameter values for the mean and variances of the current and reference distributions: $\mu_{t+h|t-J} = [0, 0, 0, 0, 0]$, $\mu_{t+h|t} = [1, 0, 0, 1, 1]$, $\sigma^2_{t+h|t-J} = [1, 1, 1, 1, 1]$, and $\sigma^2_{t+h|t-J} = [1, .5, 1.3, .5, 1.3]$.

in Corollary 1.[5] In the first row, we report the reference, current, DE, and Smooth DE distributions. In the second row, we report the weights that capture the belief distortion under DE and under Smooth DE. These are computed as:

$$weight\left(\widehat{x}_{t+h}\right) \equiv \left(\frac{N(\widehat{x}_{t+h}; \mu_{t+h|t}, \sigma^2_{t+h|t})}{N(\widehat{x}_{t+h}; \mu_{t+h|t-J}, \sigma^2_{t+h|t})}\right)^{\theta} \frac{1}{Z},$$

for the standard DE and

$$weight\left(\widehat{x}_{t+h}\right) = \left(\frac{N(\widehat{x}_{t+h}; \mu_{t+h|t}, \sigma^2_{t+h|t})}{N(\widehat{x}_{t+h}; \mu_{t+h|t-J}, \sigma^2_{t+h|t-J})}\right)^{\theta} \frac{1}{Z},$$

for the Smooth DE, where the constant of integration accounts for the respective formulation.

We organize the examples to follow the three key properties presented in Section 2 and formally characterized for the Normal distribution in Corollary 1.[6]

[5] We use the following parameter values for the mean and variances of the current and reference distributions: $\mu_{t+h|t-J} = [0, 0, 0, 0, 0]$, $\mu_{t+h|t} = [1, 0, 0, 1, 1]$, $\sigma^2_{t+h|t-J} = [1, 1, 1, 1, 1]$, and $\sigma^2_{t+h|t-J} = [1, .5, 1.3, .5, 1.3]$.

[6] Appendix D uses these examples to further and more generally discuss how the news and uncertainty

Overreaction of the conditional mean to new information. Consider first a case in which the reference and current distributions only differ in terms of the mean: $\mu_{t+h|t} > \mu_{t+h|t-J}$. The new information does not induce any change in uncertainty and $R_{t+h|t,t-J} = 1$. Smooth DE and DE lead to the same normal distribution, in which only the conditional mean is distorted. Both the SDE and DE means are shifted to the right with respect to the current distribution by $\theta(\mu_{t+h|t} - \mu_{t+h|t-J})$. The lower panel in the first column shows that the weights increase moving from left to right: The weights shift probability mass to the right of the current density, lowering the probability assigned to events that became less likely, and inflating the probability of events that became more likely. The weights keep increasing as elements in the right tail of the current density became much more likely to occur *in relative terms*. However, the true probability of these events goes to zero faster than the weights increase, preserving the normality of the SDE and DE distributions.

Over/underconfidence. In the second and third columns, the reference and current distributions *only differ in terms of variance*, while $\mu_{t+h|t} = \mu_{t+h|t-J}$. In the second column the current true distribution features a *lower* variance, i.e $\sigma^2_{t+h|t} < \sigma^2_{t+h|t-J}$ and thus $R_{t+h|t,t-J} < 1$. This case is typical for standard stochastic processes when new information leads to a reduction of uncertainty and more precise forecasts. Under the standard BGS implementation of DE, the fact that events close to the mean, and the mean itself, became more likely does not have any effect. Instead, under Smooth DE, the agent revises her beliefs in light of the new information. She inflates the probability of the mean and the other events that became more *representative*, while further downplaying the probability of events that became less likely. The result is an even narrower distribution than the current true distribution. Under Smooth DE, the new information leads to *overreaction* in terms of the decline in uncertainty and, as a result, to a novel implication: overconfidence.

In the third case the current distribution has a *larger* variance than the reference distribution, i.e. $\sigma^2_{t+h|t} > \sigma^2_{t+h|t-J}$, and thus $R_{t+h|t,t-J} > 1$. This situation could arise, for example, in response to a positive uncertainty shock. Now tail events become more representative under the revised density and receive a magnified weight under Smooth DE. The probability mass is moved from the center to the tails, but preserving normality.[7] Under DE, the weights are once again uniformly equal to 1, and the DE density coincides with the true density.

Changes in uncertainty affect the overreaction of the conditional mean. The fourth and fifth columns combine a revision in mean with a revision in variance. The key

effects introduced in Section 2 determine the effective overreaction of the Smooth DE density.

[7]This example also allows us to illustrate the role of the upper bound on $R_{t+h|t,t-J}$: As this ratio increases, more and more probability mass is moved to the tails, flattening the normal distribution. As $R_{t+h|t,t-J} \to (1+\theta)/\theta$, the variance of the Smooth DE density goes to infinity, as an increasing probability mass is moved to the tails.

observation here is illustrated by Corollary 1: under Smooth DE, a change in uncertainty affects the degree to which the distorted revision responds to the true revision. Since changes in information sets typically involve actual changes in both conditional moments, this is a particularly novel and important aspect of Smooth DE.

In particular, the fourth case combines the first two, with an *increase in the conditional mean and a reduction of conditional uncertainty*. As in the first case, under DE we observe a shift of the probability mass to the right. Accordingly, the DE density moves to the right, but with no change in shape with respect to the true density. Under Smooth DE, instead, the agent recognizes that, despite the increase in the mean, the new information made tail events to the right less representative. Thus, for a given θ, the Smooth DE density still shifts to the right, but by a smaller amount, and becomes visibly narrower, as the weights take into account the change in uncertainty.[8]

Finally, new information can also bring a shift in the mean, but now with *more* uncertainty. The last column considers this case, where the shift in the mean is positive like in the fourth column. The agent's overreaction in terms of revisions to the conditional mean is now *stronger* than under standard DE because tail events have become more likely under the current distribution. Thus, the weights determine an even more significant shift of probability mass to the right (the scale for the SDE weights is on the left). In terms of distorted conditional uncertainty, in this case, we observe underconfidence. This is again in itself a form of *overreaction*, as the agent magnifies the increase in uncertainty as this appears large compared to the reference uncertainty.

3.5 AR(1) process

The Smooth DE density can be easily applied for a standard AR(1) process. For a simpler exposition we focus on the one-step-ahead horizon ($h = 1$) and recent past ($J = 1$). We extend this to the more general h and J in Section 4.1. Consider

$$x_{t+1} = \rho x_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$$

where $\rho \leq 1$ and $\sigma^2 > 0$. The true conditional density is simply

$$f\left(\widehat{x}_{t+1}|\mathcal{F}_t\right) = \mathcal{N}\left(\widehat{x}_{t+1}; \rho x_t, \sigma^2\right).$$

[8] As the current conditional uncertainty becomes continuously smaller and converges to zero, the shift in the distorted conditional mean also smoothly converges to zero. In contrast, under standard DE, there would be a discrete jump in the extreme case of $\sigma^2_{t+h|t=0}$.

while the reference density is

$$f\left(\widehat{x}_{t+1}|\mathcal{F}_{t-1}\right)=\mathcal{N}\left(\widehat{x}_{t+1};\rho^{2}x_{t-1},\left(1+\rho^{2}\right)\sigma^{2}\right)$$

As a result, $R_{t+1|t,t-1}$ defined in Proposition 2 takes the form

$$R_{t+1|t,t-1}=\frac{\sigma^{2}}{\left(1+\rho^{2}\right)\sigma^{2}}=\frac{1}{1+\rho^{2}} \tag{22}$$

The Smooth DE density $f^{\theta}\left(\widehat{x}_{t+1}\right)$ in Proposition 1 is then Normal, with a conditional mean

$$\mathbb{E}_{t}^{\theta}\left(x_{t+1}\right) \;=\; \rho x_{t}+\frac{\theta}{1+\rho^{2}(1+\theta)}\left(\rho x_{t}-\rho^{2}x_{t-1}\right) \tag{23}$$

and conditional distorted variance

$$\mathbb{V}_{t}^{\theta}\left(x_{t+1}\right) \;=\; \frac{\sigma^{2}}{1+\frac{\rho^{2}}{1+\rho^{2}}\theta} \tag{24}$$

This AR(1) example further illustrates some of the general principles behind Smooth DE. First, equation (23) implies that, as long as $\theta>0$, Smooth DE exhibits effective overreaction of the conditional mean to news. Second, given σ^{2}, a higher persistence parameter ρ implies that the new information determines a larger reduction in current uncertainty about the variable of interest x_{t+1} compared to the reference density. This lower variance ratio $R_{t+1|t,t-1}$ leaves less room for memory to distort probability judgements which makes the effective overreaction in equation (23) decrease in the persistence parameter ρ. Third, since new information at time t lowers the conditional uncertainty from $\left(1+\rho^{2}\right)\sigma^{2}$ to σ^{2}, the ratio $R_{t+1|t,t-1}<1$ in equation (22). Thus, the distorted density is characterized by overconfidence, so that $\mathbb{V}_{t}^{\theta}\left(x_{t+1}\right)<\sigma^{2}$, as seen in equation (24).

4 A parsimonious micro-foundation for survey evidence

Like in existing work on DE, we take the underlying $\theta\geq 0$ and $J\geq 1$ as primitive parameters characterizing the decision's maker limited memory and the effect that the representativeness heuristic has on agent's judgments. As discussed in Section 3, given θ and J, the Smooth DE density does not introduce any further degrees of freedom. Nevertheless, by allowing the density for the representative group to reflect the time $t-J$ conditional uncertainty, we find that Smooth DE can offer a *joint and parsimonious micro-foundation* for a range of observable implications consistent with survey data. These implications refer to the broad

properties of Smooth DE emphasized and collected by Corollary 1.

At its core, Smooth DE captures the intuition that new information that significantly reduces current uncertainty over the variable of interest leaves less room for memory and representativeness to distort judgements. As we discuss below, this implication of a stronger (weaker) effective overreaction of the conditional mean to new information that reduces less (more) current uncertainty helps to account for two sets of stylized survey facts.

4.1 Stronger overreaction for longer forecast horizons

The first set of over-identifying restrictions on our theory of overreaction relates to the model's implications for short- versus long-horizon forecasts. A strand of literature using survey data argues that overreaction appears to be increasing with the horizon of the forecast. For example, Bordalo et al. (2019) and Bordalo et al. (2023) point to such stronger overreaction for equity analysts' forecasts of long-term earnings growth and emphasize the potential for this type of overreaction to account for stock market volatility. Using professional forecasters' forecasts of interest rates, other contributions, including for example Bordalo et al. (2020), d'Arienzo (2020), find evidence of significant overreaction for expectations of long-term interest rates, but not for expectations of short-term interest rates. Augenblick et al. (2021) use field data from betting and financial markets to argue that compared to the Bayesian forecast there is relatively stronger overreaction to signals with a longer (shorter) time-to-resolution, conceptually similar to longer (shorter) forecast horizons.

Smooth DE is consistent with such evidence as it predicts that overreaction increases with the horizon of the forecast. The basic intuition appears in Section 3. Smooth DE formalizes an inverse relation between the informativeness of the new piece of information obtained by the decision-maker and the overreaction of her conditional forecasts (see for example equation (16) in Corollary 1). In the context of forecasting at different horizons, the same piece of information is less informative about horizons further in the future, leading to a smaller reduction in uncertainty and a stronger overreaction. Thus, Smooth DE naturally predicts that overreaction is relatively stronger for long-horizon forecasts.

The simplest environment to showcase this basic insight is the AR(1) process introduced with equation (3.5) in Section 3. For an horizon $h \geq 1$ and a $J-$ lagged reference distribution ($J \geq 1$), the conditional mean for the Smooth DE density $f^{\theta}\left(\widehat{x}_{t+h}\right)$ is

$$\mathbb{E}_t^{\theta}\left(x_{t+h}\right) = \rho^h x_t + \widetilde{\theta}_{t,t-J}\left(\rho^h x_t - \rho^{h+J} x_{t-J}\right) \tag{25}$$

where the effective severity $\widetilde{\theta}_{t,t-J}$ of DE distortion is given in equation (14).

Given the AR(1) process in equation (3.5), the ratio $R_{t+h|t,t-J}$ of conditional variances,

defined in Proposition 1, takes the particular form

$$R_{t+h|t,t-J} = \frac{\mathbb{V}_t\left[x_{t+h}\right]}{\mathbb{V}_{t-J}\left[x_{t+h}\right]} = \left[\begin{array}{c} \frac{1-\rho^{2h}}{1-\rho^{2(h+J)}} \text{ when } \rho^2 < 1 \\ \frac{h}{h+J} \text{ when } \rho^2 = 1 \end{array}\right]$$

Proposition 3 *The ratio $R_{t+h|t,t-J}$ increases in the forecast horizon h. Thus, the effective overreaction $\widetilde{\theta}_{t,t-J}$ of $\mathbb{E}_t^{\theta}\left(x_{t+h}\right)$ in equation (25) is stronger for longer forecast horizons.*

For a given lag J in the reference distribution, as the forecast horizon h increases, the effective horizon of the current RE forecast (h), and the effective horizon of the reference RE forecast ($h + J$) become increasingly similar. As a result, the levels of uncertainty associated with the two forecasts also become increasingly similar because the information set is implicitly more similar. Intuitively, the uncertainty around the two forecasts reflects a larger and larger number of the same shocks. In relative terms, the current information set is less and less informative for the variable that the agent is trying to predict. Given that under Smooth DE overreaction is increasing in the level of relative uncertainty, as h increases, so does the amount of overreaction to a given revision of the RE forecasts.

4.2 Overreaction *and* overconfidence

A recent literature studying the properties of survey responses, including Barrero (2022), Born et al. (2022), and the reviews in Altig et al. (2020) and Born et al. (2022), documents that while firms' forecasts are unconditionally unbiased, i.e forecast errors are on average not significantly different from zero, firms make conditionally predictable forecast errors. In particular, firms *overreact* to news and are *overconfident* in their subjective forecasts.

These stylized facts provide a challenge for models featuring standard rational belief updating. As a result, the overreaction and overconfidence empirical properties have been typically addressed in existing models through *two distinct* behavioral primitive assumptions that do not distort unconditional forecasts. Overreaction of conditional forecasts has been explained as an outcome of DE, modelled according to the original BGS formulation. Under DE, agents overreact only in presence of new information and in a symmetric way, preserving unbiased unconditional forecasts, but failing to account for overconfidence. Thus, the finding of overconfidence has been typically addressed with an additional overconfidence bias. An example of this approach is Barrero (2022), who uses both distinct features to account for the three survey facts.

Smooth DE can instead account for all three stylized facts. Consider for example the Gaussian environment of Section 3 where a firm's fundamental (eg. productivity) follows a simple AR(1) process. Or, the arguably more empirically plausible extension, where those

fundamentals are not directly observed, but firms can learn about their realizations from noisy signals, like in the simple state-space model described in Section E. In either case, Corollary 1 and Proposition 10, respectively, describe how forecasts made under Smooth DE are characterized by precisely these two key properties: overreaction to news *and* overconfidence. Moreover, as in the RE case ($\theta = 0$), forecasts under DE are nevertheless unconditionally unbiased, being on average driven by the underlying rational forecasts.

The discussion and formalism in Section 3 indicate that, in contrast to the generality of the overreaction of the conditional mean, overconfidence is not a universal property of Smooth DE. However, we view it as a 'typical' property, because the necessary condition for overconfidence is simply that new information reduces uncertainty. This condition is ubiquitous as it holds in stationary, homoskedastic environments, where events closer into the future are naturally easier to predict than events far into the future. At the same time, the condition might not hold if new information entails a sufficiently large and unexpected increase in uncertainty, as indicated by some of our examples in Figure 3.

Finally, we further note the broader context of a large literature on overconfidence (eg. De Bondt and Thaler (1995) and Daniel et al. (1998, 2001)). This work has been motivated by extensive psychological evidence for overconfidence and argues that models based on this behavioral property are promising in accounting for asset market puzzles. Our key insight here is that Smooth DE emerges as a potential *parsimonious micro-foundation*, based on the representativeness heuristic, for overreaction and overconfidence, two behavioral features argued as important in understanding a variety of economic outcomes.[9]

5 Business cycle implications

We illustrate the business cycle implications of Smooth DE in a parsimonious RBC model with time-varying uncertainty. We argue that Smooth DE emerges as a novel behavioral propagation and amplification mechanism for time-varying uncertainty. We first show that the model replicates, without relying on additional frictions, several salient features of the data thanks to the state-dependent overreaction: (1) *asymmetry* (recessions are deeper than expansions), (2) *countercyclical micro volatility* (cross-sectional variances of microeconomic variables rise in recessions), and (3) *countercyclical macro volatility* (time-series variances of macroeconomic variables rise in recessions).[10] We also show that, the perceived increase of

[9]See further Barberis (2018) for a review of these two-widely documented, but typically studied separately, departures from standard Bayesian updating.

[10]As described in the Introduction, these properties have strong empirical support in the literature. The concept of asymmetries have a long tradition in macroeconomics, including Neftci (1984), Hamilton (1989), Sichel (1993), and more recently McKay and Reis (2008) and Morley and Piger (2012). The extensive

uncertainty in recessions is more than three times larger than the actual uncertainty increase. We then discuss a novel policy implication: a redistributive policy that reduces idiosyncratic uncertainty could be beneficial for macroeconomic stabilization because it dampens the state-dependent overreaction.

5.1 The model

To isolate the role of Smooth DE as a propagation mechanism, we keep the model simple and abstract from conventional frictions in the uncertainty shock literature, such as adjustment costs (Bloom (2009), Bloom et al. (2018)) and sticky prices (Basu and Bundick (2017), Fernández-Villaverde et al. (2015), and Bianchi et al. (2023a)). The economy consists of a continuum of islands $i \in [0, 1]$. In each island i, an agent has the per-period utility function

$$U(c_{i,t}, h_{i,t}) = \frac{c_{i,t}^{1-\gamma}}{1-\gamma} - \beta \frac{h_{i,t}^{1+\eta}}{1+\eta}.$$

where $c_{i,t}$ is consumption, $h_{i,t}$ is the amount of hours worked, γ is the coefficient of relative risk aversion, and η is the inverse of the Frisch labor elasticity. We simplify the algebra below by multiplying the disutility of labor by the discount factor β.

Output in each island is produced according to

$$y_{i,t} = z_{i,t} h_{i,t-1}. \tag{26}$$

The $t-1$ subscript on hours reflects the assumption that the labor input is chosen before the random realization of productivity $z_{i,t}$ is known. The island resource constraint is

$$c_{i,t} = y_{i,t}. \tag{27}$$

We obtain aggregate variables by simply adding up variables of all islands:

$$H_t = \int_0^1 h_{i,t} di, \quad Y_t = \int_0^1 y_{i,t} di, \quad C_t = \int_0^1 c_{i,t} di$$

The island productivity $z_{i,t+1}$ is the sum of aggregate and idiosyncratic components:

$$\ln z_{i,t+1} = A_{t+1} + a_{i,t+1},$$

literature of the macroeconomics of time-varying uncertainty, including Bloom (2009), Fernández-Villaverde et al. (2011), Ilut et al. (2018), Jurado et al. (2015), Basu and Bundick (2017), and Bloom et al. (2018) confirm that volatility or uncertainty is countercyclical at both micro and macro levels.

The economy-wide TFP shock A_{t+1} is common across all islands and follows the process

$$A_{t+1} = \rho_A A_t + u_{A,t+1}, \quad u_{A,t+1} \sim i.i.d.N(0, \sigma_A^2).$$

The idiosyncratic TFP $a_{i,t+1}$ is instead specific to island i, and it is composed of a predictable component $s_{i,t} \sim i.i.d.N(0, \sigma_s^2)$, known one-period-in-advance, and an unpredictable component $u_{a,i,t+1} \sim i.i.d.N(0, \sigma_{a,t}^2)$ realized at $t+1$: $a_{i,t+1} = s_{i,t} + u_{a,i,t+1}$. Following Bloom et al. (2018), we assume the volatility $\sigma_{a,t}$ is time-varying and negatively correlated with the economy-wide TFP. In particular, as we describe in Section 5.4, $\sigma_{a,t}$ increases when there is a negative innovation to the economy-wide TFP, and vice versa. We use $\sigma_{a,t}$ to denote the volatility of the period $t+1$ innovation to reflect the assumption that the volatility of the next period's innovation is known one-period-in-advance. We also assume that the volatility of the predictable component, $s_{i,t}$, is constant. This implies that the cross-sectional dispersion of labor is driven only by the news effect of the uncertainty shock. If we were to relax the assumption of constant volatility of $s_{i,t}$, the cross-sectional dispersion would also depend on its realized volatility, but none of the main qualitative properties of the model would change.

The mean and variance of $a_{i,t+1}$ conditional on the predictable component $s_{i,t}$ are

$$\mathbb{E}_{i,t}\left[a_{i,t+1}\right] = s_{i,t}, \quad \mathbb{V}_{i,t}\left[a_{i,t+1}\right] = \sigma_{a,t}^2.$$

We can define the residual uncertainty (posterior variance relative to ex-ante uncertainty) as in David et al. (2016) as $\sigma_{a,t}^2/(\sigma_s^2 + \sigma_{a,t}^2)$, which is increasing in $\sigma_{a,t}^2$. Intuitively, in times of low aggregate TFP and higher uncertainty $\sigma_{a,t}^2$, the predictable component $s_{i,t}$ serves as a weaker signal in forecasting $a_{i,t+1}$ relative to times of lower uncertainty.

5.2 Rational Expectations solution

We first characterize the equilibrium under RE. The island i agent's Bellman equation is

$$\mathcal{V}(h_{i,t-1}, z_{i,t}) = \max_{h_{i,t}} \left\{U(c_{i,t}, h_{i,t}) + \beta \mathbb{E}_{i,t}\left[\mathcal{V}(h_{i,t}, z_{i,t+1})\right]\right\}.$$

Combining the first order condition for labor with the envelope condition, we obtain

$$(h_{i,t})^{\eta} = \mathbb{E}_{i,t}\left[(c_{i,t+1})^{-\gamma} z_{i,t+1}\right]. \tag{28}$$

The optimality condition equates the current marginal disutility of working with its expected benefit. The latter is given by the marginal product of labor weighted by the marginal utility of consumption. We log-linearize the condition and use the method of undetermined

coefficients to obtain the RE solution. Hats denote log-deviations from steady state.

Proposition 4 *The equilibrium under RE is given as follows:*

1. *Individual hours worked are given by*

$$\widehat{h}_{i,t} = \varepsilon \left[\rho_A A_t + s_{i,t} \right],$$

where $\varepsilon = \frac{1-\gamma}{\eta+\gamma}$. *Equilibrium output and consumption follow immediately as*

$$\widehat{y}_{i,t} = A_t + a_{i,t} + \widehat{h}_{i,t-1} = \widehat{c}_{i,t}.$$

2. *Equilibrium aggregate variables are given by*

$$\widehat{H}_t = \varepsilon \rho_A A_t, \quad \widehat{Y}_t = A_t + \widehat{H}_{t-1} = \widehat{C}_t$$

Proof. See Appendix. ■

The response of individual and aggregate hours to news about expected economy-wide productivity $\rho_A A_t$ and island-specific productivity $s_{i,t}$ is affected by the intertemporal elasticity of consumption (IES), which here also equals the inverse of the coefficient of relative risk aversion. When the IES is large enough, so that $\gamma^{-1} > 1$ and thus $\varepsilon > 0$, an increase in expected productivity raises hours. In that case the intertemporal substitution effect dominates the wealth effect that would lower hours through the effect on marginal utility.

The next proposition characterizes the cross-sectional variance under RE.

Proposition 5 *The cross-sectional variance of hours worked is given by*

$$\int_0^1 \left(\widehat{h}_{i,t} - \widehat{H}_t \right)^2 di = \left[\frac{1-\gamma}{\eta+\gamma} \right]^2 \sigma_s^2,$$

and is constant over the business cycle. The cross-sectional variances of output $y_{i,t}$ *and consumption* $c_{i,t}$ *are increasing in the volatility* $\sigma_{a,t-1}^2$ *of the idiosyncratic TFP.*

Proof. See Appendix. ■

Under RE, the cross-sectional variance of hours stays constant over the business cycle. This is because once the model is linearized, the news effect of changes in uncertainty is muted under RE. The cross-sectional variances of output and consumption are instead mechanically affected by $\sigma_{a,t-1}^2$ because of the change in realized volatility.

5.3 Smooth DE solution

We now solve the model under Smooth DE. We consider the case of distant memory, meaning that agents' memory recall is based on a more distant past, rather than just the immediate past. This means that the reference group is based on the information set available $J > 1$ periods ago. Bianchi et al. (2024) find that in standard models, distant memory can account for salient features of data, such as persistence and repeated boom-bust cycles. Under distant memory, a time-inconsistency problem arises due to the failure of the law of iterated expectations. Bianchi et al. (2024) address this issue by adopting the *naïveté* approach (O'Donoghue and Rabin (1999)), which we follow here. Under this approach, the agent fails to take into account that her preferences are time-inconsistent and thinks that in the future she will make choices under perfect memory recall, or RE. However, when the future arrives, the agent ends up changing behavior and be again subject to her imperfect memory recall.[11]

Let θ-superscripts and RE-superscripts denote equilibrium Smooth DE choices and choices under a RE policy function, respectively. The island i agent's Bellman equation is

$$\max_{h_{i,t}^{\theta}} \left\{ U(c_{i,t}^{\theta}, h_{i,t}^{\theta}) + \beta \mathbb{E}_{i,t}^{\theta} \left[\mathcal{V}(h_{i,t}^{\theta}, z_{i,t+1}) \right] \right\},$$

subject to $y_{i,t}^{\theta} = z_{i,t} h_{i,t-1}^{\theta}$ and $c_{i,t}^{\theta} = y_{i,t}^{\theta}$. The continuation value is given by

$$\mathcal{V}(h_{i,t-1}^{\theta}, z_{i,t}) = \max_{h_{i,t}^{RE}} \left\{ U(c_{i,t}^{RE}, h_{i,t}^{RE}) + \beta \mathbb{E}_{i,t} \left[\mathcal{V}(h_{i,t}^{RE}, z_{i,t+1}) \right] \right\},$$

subject to $y_{i,t}^{RE} = z_{i,t} h_{i,t-1}^{\theta}$ and $c_{i,t}^{RE} = y_{i,t}^{RE}$.

Similar to the RE problem, the agent optimally equates the marginal disutility of labor with its expected benefit, except that the benefit is evaluated under Smooth DE:

$$\left(h_{i,t}^{\theta}\right)^{\eta} = \mathbb{E}_{i,t}^{\theta} \left[\left(c_{i,t+1}^{RE}\right)^{-\gamma} z_{i,t+1} \right]. \tag{29}$$

Proposition 6 *The equilibrium under Smooth DE is given as follows:*

1. *Individual hours worked are given by*

$$\begin{aligned} \widehat{h}_{i,t}^{\theta} &= \varepsilon \left[\rho_A A_t + s_{i,t} \right] \\ &\quad + \frac{\widetilde{\theta}_{t,t-J} \eta}{\eta + \left(1 + \widetilde{\theta}_{t,t-J}\right) \gamma} \varepsilon \left[\rho_A \mathbb{N}_{t-J,t} \left[A_t \right] + s_{i,t} \right], \end{aligned} \tag{30}$$

[11] Bianchi et al. (2024) further argue that the naïveté approach is psychologically coherent and consistent with the underlying foundation of diagnostic beliefs as a heuristic and a mental short-cut.

where ε is given in Proposition 4 and $\mathbb{N}_{t-J,t}[A_t] \equiv A_t - \mathbb{E}_{t-J}[A_t]$ represents the news in A_t, compared to past expectation. Equilibrium output and consumption follow as

$$\widehat{y}^{\theta}_{i,t} = A_t + a_{i,t} + \widehat{h}^{\theta}_{i,t-1} = \widehat{c}^{\theta}_{i,t}. \tag{31}$$

2. The effective diagnosticity parameter $\widetilde{\theta}_{t,t-J}$ is given by

$$\widetilde{\theta}_{t,t-J} = \frac{R_{t+1|t,t-J}\theta}{1 + \left(1 - R_{t+1|t,t-J}\right)\theta}, \tag{32}$$

where

$$R_{t+1|t,t-J} = \frac{\mathbb{V}_{i,t}\left(-\gamma \widehat{c}^{RE}_{i,t+1} + A_{t+1} + a_{i,t+1}\right)}{\mathbb{V}_{i,t-J}\left(-\gamma \widehat{c}^{RE}_{i,t+1} + A_{t+1} + a_{i,t+1}\right)}. \tag{33}$$

3. Equilibrium aggregate variables are given by

$$\begin{aligned} \widehat{H}^{\theta}_t &= \varepsilon \rho_A A_t + \frac{\widetilde{\theta}_{t,t-J}\eta}{\eta + \left(1 + \widetilde{\theta}_{t,t-J}\right)\gamma} \varepsilon \rho_A \mathbb{N}_{t-J,t}[A_t] \\ \widehat{Y}^{\theta}_t &= A_t + \widehat{H}^{\theta}_{t-1} = \widehat{C}^{\theta}_t \end{aligned} \tag{34}$$

Proof. See Appendix. ■

First, consider the policy function for individual hours $h^{\theta}_{i,t}$. The first line of (30) is identical to the RE policy function. The second line captures the overreaction to news, i.e. surprises.[12] Consider, for instance, a positive surprise to an economy-wide TFP A_t. Smooth diagnostic agents are over-influenced by this surprise and become over-optimistic about the future benefit of working, and hence work more (if $\varepsilon > 0$). The coefficient on this overreaction, $\frac{\widetilde{\theta}_{t,t-J}\eta}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}$, is increasing in the effective diagnosticity $\widetilde{\theta}_{t,t-J}$. From (31) individual output and consumption also overreact when individual hours overreact. Second, the effective diagnosticity $\widetilde{\theta}_{t,t-J}$ is positively related to $R_{t+1|t-J}$, given by (33): the ratio between the current uncertainty about the marginal benefit of labor and the uncertainty perceived at period $t-J$. Third, aggregate hours, output, and consumption also feature overreaction, controlled by $\widetilde{\theta}_{t,t-J}$, to news about economy-wide shocks.

The expressions (32) and (33) in Proposition 6 suggest that an increase in uncertainty about future idiosyncratic productivity could raise $\widetilde{\theta}_{t,t-J}$ and, in turn, the overreaction to news.[13] The Proposition below indeed confirms that this is the case.

[12] Note that, for $s_{i,t}$, since it is i.i.d., the surprise is $s_{i,t}$ itself.

[13] In the current model, we must take a stance on how agents deal with time-varying volatility when forming expectations. There are two approaches to compute the conditional variance at $t-J$ in (33) that preserve normality of the Smooth DE density. The first approach consists of making an "anticipated

Proposition 7 *An increase in the volatility $\sigma_{a,t}^2$ of idosyncratic TFP raises the effective diagnosticity parameter $\widetilde{\theta}_{t,t-J}$.*

Proof. See Appendix. ■

There are two important implications of this proposition. First, *the business cycle is asymmetric*, even if the underlying shocks are symmetric. A positive TFP shock would lower uncertainty and, as result, overreaction. In contrast, a negative TFP shock would raise uncertainty and, as result, overreaction. Thus, a drop in economic activity in response to a negative shock would be sharper, while an expansion in response to a symmetric positive shock would be milder, generating asymmetric fluctuations. Second, *macroeconomic volatility is countercyclical*. During expansions uncertainty and overreaction are low while in recessions agents overreact more to economy-wide shocks.

State-dependent overreaction also implies that micro-level volatility is countercyclical:

Proposition 8 *The cross-sectional variance of hours worked is given by*

$$\int_0^1 \left(\widehat{h}_{i,t}^{\theta} - \widehat{H}_t^{\theta}\right)^2 di = \left[\frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta+(1+\widetilde{\theta}_{t,t-J})\gamma}\right]^2 \sigma_s^2,$$

and is increasing in $\widetilde{\theta}_{t,t-J}$ and, thus, in the volatility $\sigma_{a,t}^2$ of the idiosyncratic TFP. The cross-sectional variances of output $y_{i,t}$ and consumption $c_{i,t}$ are similarly increasing in the volatility $\sigma_{a,t-1}^2$ of the idiosyncratic TFP.

Proof. See Appendix. ■

As uncertainty increases, the overreactions to the predictable component of idiosyncratic TFP and the future benefit of labor, captured in the $\left[\frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta+(1+\widetilde{\theta}_{t,t-J})\gamma}\right]^2$ term, rise. Hence, an increase in uncertainty about idiosyncratic TFP raises the cross-sectional variances of individual actions.

Our theory has an important policy implication. As we saw above, the micro-level volatility and macroeconomic volatility are tightly linked through the state-dependent overreaction controlled by $\widetilde{\theta}_{t,t-J}$. Thus, a policy that reduces microeconomic uncertainty through, for instance, a redistributive tax policy can also be effective in stabilizing the macroeconomy. To fix ideas, consider a progressive income tax and subsidy scheme where the individual rate $\tau_{i,t}$ is increasing in the realized idiosyncratic productivity level $\tau_{i,t} = \tau a_{i,t}$, where $\tau \geq 0$ is a

utility" assumption (Kreps (1998)). In this case, agents' uncertainty depends on the volatility at the time of the forecast, disregarding the possibility of volatility changes. The second approach consists of assuming that agents take into account the possibility of volatility changes, but that memory retrieves a Normal approximation of the resulting mixture of Normal's. We adopt the first approach, as it is arguably more consistent with the naïveté assumption and the general motivation of DE as a mental heuristics.

parameter that controls the progressivity. The island resource constraint is

$$c_{i,t} + \tau_{i,t} y_{i,t} = y_{i,t},$$

so the agent pays a tax ($\tau_{i,t} > 0$) if the realized TFP shock is positive ($a_{i,t} > 0$) and receives a transfer ($\tau_{i,t} < 0$) otherwise. The scheme is budget neutral.

Proposition 9 *A higher progressivity τ is associated with a smaller increase in the effective diagnosticity parameter $\widetilde{\theta}_{t,t-J}$ when the volatility $\sigma^2_{a,t}$ of idiosyncratic TFP rises.*

Proof. See Appendix. ■

Intuitively, redistribution dampens the state-dependent overreaction by reducing cross-sectional uncertainty about the future benefit of labor. Thus, the government can stabilize the macroeconomy by using the tax policy to reduce the increase in uncertainty and overreaction. For instance, in times of low aggregate TFP and high uncertainty, the government can implement the tax policy or increase its progressivity. These interventions would dampen the overreaction and make the downturn less severe.

5.4 Calibrated example

We illustrate the quantitative potential of the Smooth DE mechanism in the context of the parsimonious RBC model presented above by examining its dynamics.

Calibration. We calibrate the model to a quarterly frequency. We set the discount factor $\beta = 0.99$, the IES $\gamma^{-1} = 0.25^{-1}$, and $\eta = 0.4$, which implies a Frisch elasticity of labor supply of 2.5.[14] For the economy-wide TFP shock, we set $\rho_A = 0.95$ and $\sigma_A = 0.7/100$. The calibration satisfies the condition $\gamma^{-1} > 1$, so labor increases in response to an increase in expected TFP.

Consider the time-varying standard deviation $\sigma_{a,t}$ of the idiosyncratic TFP shocks. Using Census micro data, Bloom et al. (2018) and Ilut et al. (2018) find that, during recessions, the dispersion of TFP shocks increases by 13% and 7%, respectively. Motivated by these findings, we assume that a *negative* innovation to economy-wide TFP larger or equal than one standard deviation is associated with a 10% increase in the standard deviation $\sigma_{a,t}$ of the idiosyncratic TFP shocks relative to the steady-state standard deviation σ_a. Conversely, a *positive* economy-wide TFP innovation of the same magnitude is associated with a 10% decrease in the standard deviation $\sigma_{a,t}$ of the idiosyncratic TFP shocks.

[14] These values of IES and Frisch elasticity allow us to generate realistic labor volatility. Our calibrated model generates the time-series standard deviation of aggregate hours worked of 1.67%. In the data, the standard deviation of total hours worked in the nonfarm business sector (1983:Q1–2019:Q4) is 1.66%.

Table 1: Internally calibrated parameters and targeted moments

Parameters		Targeted moments		
			Data	Model
σ_a	0.022	Realized absolute forecast error	0.143	0.143
σ_s	0.027	Residual uncertainty	0.41	0.41
$\widetilde{\theta}$	1.547	Skewness of aggregate hours	−0.21	−0.21

Notes: The table reports the parameters and their calibrated values as well as the targeted moments. σ_a is the steady-state standard deviation of unpredictable component of the idiosyncratic TFP shock, σ_s is the standard deviation of the predictable component of the idiosyncratic TFP shock, and $\widetilde{\theta}$ is the long-run average effective diagnosticity implied by the calibrated value of θ. The realized absolute forecast error is reported in Barrero (2022) using survey data on US managers, calculated from realized forecast errors of sales growth between t to $t+4$, with observations employment-weighted. The residual uncertainty from David et al. (2016) captures the amount of posterior uncertainty relative to the ex-ante uncertainty. The skewness of aggregate hours is calculated using total hours worked in the nonfarm business sector (1983:Q1–2019:Q4). The model moments are calculated using simulated data from the Smooth DE model.

We assume that the agent's comparison group is the expectation formed $J = 5$ periods ago. The parameter J mainly determines the persistence of overreaction and the value is consistent with Bianchi et al. (2024), who find that in an estimated structural model the memory weights center around five- and six-quarters-ago expectations.

There are three remaining parameters: the steady-state standard deviation of the unpredictable component of the idiosyncratic TFP shock σ_a, the standard deviation of the predictable component of the idiosyncratic TFP shock σ_s, and the diagnosticity parameter θ. We calibrate these parameters so that the model matches the three empirical moments summarized in Table 1.[15] While multiple model parameters jointly affect these moments, we select the moments so that each moment is informative about a parameter of interest.

The first empirical moment, the mean of realized absolute forecast errors, is from Barrero (2022) who uses survey data (Atlanta Fed/Chicago-Booth/Stanford Survey of Business Uncertainty (SBU)) on US managers. Forecast errors are computed by subtracting realized sales growth between t to $t+4$ from managers' forecasts. The model counterpart is obtained by calculating the mean absolute forecast error on the simulated distribution of the realized forecast error $\mathbb{E}^{\theta}_{i,t}\left[\widehat{y}^{RE}_{i,t+4}\right] - \widehat{y}^{\theta}_{i,t+4}$.[16] This moment is informative about the steady-state

[15] We choose the parameters so that the squared-sum of distance between the data moments and the model-implied moments is minimized.

[16] Specifically, we generate 100 replications of $T = 200$ time series with $n = 500$ islands. The number of islands roughly matches the the number of firms surveyed in the SBU data in Barrero (2022).

standard deviation of the unpredictable component of idiosyncratic TFP.

The second moment, residual uncertainty, captures the amount of posterior uncertainty relative to the ex-ante uncertainty. David et al. (2016) estimate the residual uncertainty to be 41%. The model counterpart is given by $\sigma_a^2/(\sigma_s^2+\sigma_a^2)$. This moment is useful to pin down the standard deviation σ_s of the predictable component of idiosyncratic TFP.

Finally, the third moment is the skewness of total hours worked in the nonfarm business sector (1983:Q1–2019:Q4).[17] The negative skewness (-0.21) reflects macroeconomic asymmetry: drops in hours worked are steeper than increases. In our model, a negative economy-wide TFP innovation increases uncertainty $\sigma_{a,t}^2$ and, in turn, the effective overreaction $\widetilde{\theta}_{t,t-J}$. A positive TFP innovation, in contrast, reduces overreaction. Under Smooth DE, the diagnosticity parameter θ governs the strength of this mechanism to generate asymmetry. Under RE model and the standard DE model where the overreaction is constant, there is no asymmetry, and the skewness is zero.

The model moments match the empirical moments perfectly. The calibrated σ_a and σ_s imply the predictable and unpredictable components' variances are about the same in steady state. The long-run average effective diagnosticity parameter $\widetilde{\theta}$, implied by the calibrated value of θ, is 1.54. This value is somewhat larger than Bordalo et al. (2018), Bordalo et al. (2019), and d'Arienzo (2020), which tend to estimate the standard DE diagnosticity parameter around 1, but smaller than the estimate of 1.97 in Bianchi et al. (2024).[18]

Implications for untargeted survey moments. We examine to what extent our theory can explain untargeted survey evidence on overreaction and overconfidence. We use Barrero (2022)'s survey moments as an external validation because the study shows both overreaction and overconfidence based on a single dataset (SBU). The first column of Table 2 reports the coefficient from a panel regression where managers' time t forecast of $t+4$ sales growth minus the realization is regressed on the sales growth between quarter $t-1$ to t. The coefficient is positive, meaning that managers' forecasts tend to be excessively optimistic during high growth period. The second column is the realized mean absolute forecast error, reported in Table 1, and is shown here again to facilitate comparison. The third column is the subjective mean absolute forecast error, where the hypothetical realizations are drawn from the managers' subjective probability distributions. The subjective absolute forecast error is only 16% the size of the empirical errors (fourth column), indicating overconfidence:

[17] The empirical skewness of hours increases significantly to -1.75 when we extend the sample until 2022:Q1 to include the 2020 Covid-19 recession. We use the simulated data to compute the skewness of aggregate hours worked in the model. Both simulated and actual time series are HP-filtered with $\lambda = 1600$.

[18] Like Bianchi et al. (2024), our current model features distant memory ($J > 1$). Bianchi et al. (2024) notes that existing estimates are based primarily on models where imperfect memory is assumed to be driven only by the immediate past ($J = 1$), and this assumption changes inference about the diagnosticity parameter.

Table 2: Untargeted survey moments: overreaction and overconfidence

	(1) $F_t(\Delta y_{i,t+4\|t}) - \Delta y_{i,t+4\|t}$ on $\Delta y_{i,t\|t-1}$	(2) Absolute forecast error	(3)	(4)
		Realized	Subjective	(Subjective)/(Realized)
Data	0.173	0.143	0.023	0.16
	(0.059)	(0.012)	(0.002)	
Model	0.095	0.143	0.017	0.12

Notes: The table reports the coefficient on overextrapolation and realized and subjective mean absolute forecast errors. The data moments are computed by Barrero (2022) using survey data on US managers, with observations employment-weighted and standard errors in parentheses. The first column reports the coefficient from a panel regression where managers' time t forecast of $t+4$ sales growth minus the realization is regressed on the sales growth between quarter $t-1$ to t. The second column is the realized mean absolute forecast error, calculated using realized forecast errors of sales growth between t to $t+4$. The realized mean absolute forecast error is used in the calibration as a target, but is included in this table to facilitate comparison. The third column is the subjective mean absolute forecast error, where the hypothetical realizations are drawn from managers' subjective probability distributions. The fourth column is the ratio of the subjective absolute forecast error to the realized error. The model moments are calculated using the simulated data from the Smooth DE model.

managers overestimate the precision of their forecasts.

The model moments are computed by simulating the model under Smooth DE. First, consider the overextrapolation regression coefficient. The model counterpart is the coefficient on pooled OLS where we regress the Smooth DE four-quarters-ahead forecast error, $\mathbb{E}^{\theta}_{i,t}\left[\widehat{y}^{RE}_{i,t+4}\right] - \widehat{y}^{\theta}_{i,t+4}$, on output growth, $\left[\widehat{y}^{\theta}_{i,t} - \widehat{y}^{\theta}_{i,t-1}\right]$, which proxies for news. The coefficient is positive, but smaller than in the data. The reason why the calibrated model understates this coefficient relative to the data is as follows. In our model, economy-wide shocks are persistent while island-specific shocks are i.i.d. In contrast to persistent shocks, when shocks are i.i.d., the Smooth DE forecasts are orthogonal to news, so the idiosyncratic shocks push the coefficient toward zero.[19] While we specified idiosyncratic TFP shocks to be i.i.d. for tractability, allowing for persistence would increase the overextrapolation coefficient. Thus, our model provides a conservative lower bound on the macroeconomic effects of Smooth DE.

Next, consider the mean absolute forecast errors. The subjective error (third column) is

[19]This need not be the case when agents forecast endogenous variables in models with slow-moving endogenous states, such as capital. See Bianchi et al. (2024) for details.

Table 3: Countercyclical cross-sectional standard deviation of labor growth

	(1) Data	(2) Smooth DE	(3) DE	(4) RE
(Recessions)/(Expansions)	1.16	1.12	1	1

Notes: The table reports the ratio of the cross-sectional standard deviation of labor growth during recessions to the cross-sectional standard deviation during expansions. The first column shows the ratio in the data, reported by Ilut et al. (2018), where recessions and expansions are defined as NBER recessions and NBER expansions, respectively. The second, third, and fourth columns report the model-implied ratios for the Smooth DE, DE, and RE models, respectively.

obtained first by calculating the Smooth DE variance of output growth

$$\mathbb{V}_{i,t}^{\theta}\left[\widehat{y}_{i,t+4}^{RE}\right]=\frac{\mathbb{V}_{i,t}\left[\widehat{y}_{i,t+4}^{RE}\right]}{1+\left(1-R_{t+1|t,t-J}\right)\theta}, \tag{35}$$

and then leverage the normality of the RE output growth so that the subjective absolute forecast error is given by $\sqrt{2/\pi}\left(\mathbb{V}_{i,t}^{\theta}\left[\widehat{y}_{i,t+4}^{RE}\right]\right)^{\frac{1}{2}}$. The model closely matches the subjective forecast error. The size of the absolute subjective error is 12% of the size of the realized forecast error (fourth column), in line with the survey data's 16%. According to (35), the Smooth DE variance $\mathbb{V}_{i,t}^{\theta}\left[\widehat{y}_{i,t+4}^{RE}\right]$ would be lower than the econometrician's variance $\mathbb{V}_{i,t}\left[\widehat{y}_{i,t+4}^{\theta}\right]$ due to two factors. The first factor is that, under *naïveté*, (Smooth) DE agents perceive future output to follow the RE law of motion $\widehat{y}_{i,t+4}^{RE}$ instead of the equilibrium law of motion $\widehat{y}_{i,t+4}^{\theta}$. The second factor is the Smooth DE effect (the denominator in (35)) on uncertainty, according to which a reduction of uncertainty contributes to overconfidence about the precision of expectations. To disentangle these two factors, we calculate the subjective mean absolute forecast error without the Smooth DE effect. We obtain $\sqrt{2/\pi}\left(\mathbb{V}_{i,t}\left[\widehat{y}_{i,t+4}^{RE}\right]\right)^{\frac{1}{2}}=0.053$, which is 37% of the size of the realized errors. This ratio is more than double the values recovered by the data (16%) and implied by the baseline model (12%). We conclude that the Smooth DE effect is important to account for overconfidence as observed in survey data.

Countercyclical micro and macro volatility. We now study the model's ability to generate countercyclical micro and macro volatility. First, consider the micro volatility. In Table 3, we report the ratio of the cross-sectional standard deviation of labor growth during recessions to the cross-sectional standard deviation during expansions. The first column shows this ratio from the data, as reported by Ilut et al. (2018), where recessions and expansions are defined as NBER recessions and NBER expansions, respectively. The cross-sectional dispersion is countercyclical: in recessions, it is 16% higher than during expansions.

Table 4: Countercyclical volatility of aggregate labor growth

	(1) Data	(2) Smooth DE	(3) DE	(4) RE
(Recessions)/(Expansions)	1.23	1.22	1	1

Notes: The table reports the ratio of the rolling standard deviation of aggregate labor growth during recessions to the rolling standard deviation during expansions. The first column shows the ratio in the data for the period 1983:Q1-2019:Q4, where recessions and expansions are defined as NBER recessions and NBER expansions, respectively. The second, third, and fourth columns report the model-implied ratios for the Smooth DE, DE, and RE models, respectively.

The second column reports the ratio in our model, where we define recessions and expansions as periods when there are one-standard-deviation negative and positive innovations to the economy-wide TFP, respectively. Under Smooth DE the cross-sectional standard deviation of labor growth is 12% higher during recessions than in expansions, so the model explains 75% of the empirical countercyclicality of micro volatility. In the model, a negative aggregate TFP innovation triggers an increase in idiosyncratic TFP uncertainty $\sigma_{a,t}$. As a result, the overreaction $\widetilde{\theta}_{t,t-J}$ to the predictable component $s_{i,t}$ of idiosyncratic TFP rises, so the cross-sectional dispersion of actions such as labor increases. As shown in the third and the fourth columns, the cross-sectional dispersion is constant over the business cycle under the standard DE model, where the overreaction is constant, and the RE model, where we have $\theta = 0$.

Next, consider macro volatility. In our model, in times of low TFP and high idiosyncratic uncertainty $\sigma_{a,t}$, aggregate labor responds more to economy-wide shocks because the overreaction is stronger. Table 4 examines to what extent this countercyclical macro volatility is consistent with the data. To measure time-varying volatility of aggregate hours worked in the data, similar to Ilut et al. (2018), we compute the rolling window standard deviation as

$$\sigma_{H,t} = \sqrt{\frac{1}{n_w - 1} \sum_{k=-(n_w-1)/2}^{(n_w-1)/2} \left(\Delta \ln H_{t+k} - \overline{\Delta \ln H_t}\right)^2}, \tag{36}$$

where $\Delta \ln H_{t+k}$ is the log change of total hours worked in the nonfarm business sector from a quarter $t+k-1$ to $t+k$ and $\overline{\Delta \ln H_t} \equiv (1/n_w) \sum_{k=-(n_w-1)/2}^{(n_w-1)/2} \Delta \ln H_{t+k}$. We set the window size $n_w = 3$ and consider the sample 1983:Q1-2019:Q4. The first column of Table 4 reports the measured $\sigma_{H,t}$ during NBER recessions relative to $\sigma_{H,t}$ during NBER expansions. The measured volatility of aggregate labor growth is 23% higher in recessions than in expansions. We then compute the same rolling standard deviation (36) on the simulated data from

the model. We define recessions and expansions as periods when there are larger-than-or-equal-to one-standard-deviation negative and positive innovations to the economy-wide TFP, respectively. The second column shows that in the Smooth DE model, the aggregate labor growth volatility $\sigma_{H,t}$ is 22% higher in recessions than in expansions. Thus, the Smooth DE model generates the countercyclical macro volatility that is quantitatively in line with the data, even though the volatility of economy-wide shocks is constant. DE and RE models, in contrast, do not generate such countercyclical volatility (third and fourth columns).

Perceived vs. actual increase in uncertainty. Discussions of Corollary 1 and Figure 3 indicate that, in our model, agents would overestimate the increase in uncertainty in recessions. This is because tail events become more representative when uncertainty rises. To quantify how much the agent's perceived uncertainty rises relative to the actual uncertainty, we compute the Smooth DE variance of the future marginal benefit of labor,

$$\mathbb{V}_{i,t}^{\theta}\left(-\gamma\widehat{c}_{i,t+1}^{RE}+A_{t+1}+a_{i,t+1}\right)=\frac{\mathbb{V}_{i,t}\left(-\gamma\widehat{c}_{i,t+1}^{RE}+A_{t+1}+a_{i,t+1}\right)}{1+\theta(1-R_{t+1|t,t-J})}. \tag{37}$$

We are interested in (37) because it controls the labor-supply decision in response to an uncertainty increase. We find that, under our calibration, a 10% rise in $\sigma_{a,t}$ raises the perceived uncertainty (37) by 69%. In contrast, actual uncertainty, given by $\mathbb{V}_{i,t}\left(-\gamma\widehat{c}_{i,t+1}^{\theta}+A_{t+1}+a_{i,t+1}\right)$ rises only by 19%. Thus, the perceived rise in uncertainty is more than three times larger than the actual uncertainty increase in recessions.[20]

6 Conclusions

We developed a tractable and structural bridge from the representativeness heuristic of Kahneman and Tversky (1972) to the time-series domain. We built on the formalization of representativeness by Gennaioli and Shleifer (2010) and of diagnostic expectations (DE) by Bordalo et al. (2018) to allow for what we call "smooth diagnosticity." Under Smooth DE new information is defined as the difference between the current information set and a previous information set. A critical consequence of this basic approach is that current and past uncertainty interact to determine the intensity of the DE overreaction, but also create the preconditions for novel properties such as over- and under- confidence.

After formally characterizing Smooth DE and its key properties, we leveraged its insights

[20]Since we linearize our model, the increase in uncertainty affects first-order economic outcomes through the state-dependent overreaction and not through the conventional risk channel. The risk adjusted log-linearization method as in Bianchi et al. (2023b) would allow us to capture the impact of perceived increase in uncertainty while preserving tractability.

along two substantive directions. First, we embedded the Smooth DE framework in a standard signal extraction problem and showed that Smooth DE can account for recent evidence indicating that overreaction is stronger for weaker signals and for longer horizon forecasts. Second, we embedded Smooth DE in a parsimonious RBC model with time-varying uncertainty. This model can account for survey data on overreaction and overconfidence as well as three salient properties of the business cycle: (1) asymmetry, (2) countercyclical micro volatility, and (3) countercyclical macro volatility. We uncovered a novel policy implication: a redistributive policy that reduces idiosyncratic uncertainty could be beneficial for macroeconomic stabilization because it dampens the state-dependent overreaction.

References

Altig, D., J. M. Barrero, N. Bloom, S. J. Davis, B. Meyer, and N. Parker (2020). Surveying business uncertainty. *Journal of Econometrics*.

Augenblick, N., E. Lazarus, and M. Thaler (2021). Overinference from weak signals and underinference from strong signals.

Baker, S. R., N. Bloom, and S. J. Terry (2024). Using disasters to estimate the impact of uncertainty. *Review of Economic Studies 91*(2), 720–747.

Barberis, N. (2018). Psychology-based models of asset prices and trading volume. In *Handbook of behavioral economics: applications and foundations 1*, Volume 1, pp. 79–175.

Barrero, J. M. (2022). The micro and macro of managerial beliefs. *Journal of Financial Economics 143*(2), 640–667.

Basu, S. and B. Bundick (2017). Uncertainty shocks in a model of effective demand. *Econometrica 85*, 937–958.

Bianchi, F., C. Ilut, and H. Saijo (2024). Diagnostic business cycles. *The Review of Economic Studies 91*(1), 129–162.

Bianchi, F., H. Kung, and M. Tirskikh (2023a). The origins and effects of macroeconomic uncertainty. *Quantitative Economics 14*(3), 855–896.

Bianchi, F., H. Kung, and M. Tirskikh (2023b, July). The origins and effects of macroeconomic uncertainty. *Quantitative Economics 14*(3), 855–896.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bloom, N. (2014, Spring). Fluctuations in uncertainty. *Journal of Economic Perspectives 28*(2), 153–176.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018, May). Really uncertain business cycles. *Econometrica 86*(3), 1031–1065.

Bordalo, P., K. Coffman, N. Gennaioli, F. Schwerter, and A. Shleifer (2020). Memory and representativeness. *Psychological Review*. forthcoming.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *The Quarterly Journal of Economics 131*(4), 1753–1794.

Bordalo, P., N. Gennaioli, R. La Porta, M. OBrien, and A. Shleifer (2023). Long term expectations and aggregate fluctuations. *NBER Macro Annual*.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020). Overreaction in macroeconomic expectations. *American Economic Review 110*(9), 2748–82.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019). Diagnostic expectations and stock returns. *The Journal of Finance 74*(6), 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *The Journal of Finance 73*(1), 199–227.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022). Overreaction and diagnostic expectations in macroeconomics. *Journal of Economic Perspectives 36*(3), 223–44.

Bordalo, P., N. Gennaioli, A. Shleifer, and S. J. Terry (2019). Real credit cycles. Harvard, mimeo.

Born, B., Z. Enders, M. Menkhoff, G. J. Müller, and K. Niemann (2022). Firm expectations and news: Micro v macro. Technical report, mimeo.

Born, B., Z. Enders, G. J. Müller, and K. Niemann (2022). Firm expectations about production and prices: Facts, determinants, and effects. *Handbook of Economic Expectations*.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance 53*(6), 1839–1885.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam (2001). Overconfidence, arbitrage, and equilibrium asset pricing. *The Journal of Finance 56*(3), 921–965.

d'Arienzo, D. (2020). Maturity increasing overreaction and bond market puzzles. Bocconi University, Working Paper.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, misallocation, and aggregate productivity. *Quarterly Journal of Economics 131*(2), 943—1005.

De Bondt, W. F. and R. H. Thaler (1995). Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science 9*, 385–410.

Fernández-Villaverde, J., P. Guerrón-Quintana, K. Kuester, and J. F. Rubio-Ramírez (2015). Fiscal volatility shocks and economic activity. *American Economic Review 105*(11), 3352–3384.

Fernández-Villaverde, J., P. Guerrón-Quintana, J. F. Rubio-Ramírez, and M. Uribe (2011). Risk matters: The real effects of volatility shocks. *American Economic Review 101*(6), 2530–2561.

Gennaioli, N. and A. Shleifer (2010). What comes to mind. *The Quarterly Journal of Economics 125*(4), 1399–1433.

Gennaioli, N. and A. Shleifer (2018). *A crisis of beliefs: Investor psychology and financial fragility*. Princeton University Press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Ilut, C. L., M. Kehrig, and M. Schneider (2018, September). Slow to hire, quick to fire: Employment dynamics with asymmetric responses to news. *Journal of Political Economy 126*(5), 2011–2071.

Jurado, K., S. C. Ludvigson, and S. Ng (2015, March). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology 3*(3), 430–454.

Kreps, D. (1998). Anticipated utility and dynamic choice. In D. P. Jacobs, E. Kalai, and M. I. Kamien (Eds.), *Frontiers of Research in Economic Theory*. Cambridge University Press.

McKay, A. and R. Reis (2008). The brevity and violence of contractions and expansions. *Journal of Monetary Economics 55*(4), 738–751.

Morley, J. and J. Piger (2012). The asymmetric business cycle. *Review of Economics and Statistics 94*(1), 208–221.

Neftci, S. N. (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, 556–569.

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American Economic Review 89*(1), 103–124.

Sichel, D. F. (1993). Business cycle asymmetry: A deeper look. *Economic Inquiry 31*(2), 224–236.

Tversky, A. and D. Kahneman (1975). Judgment under uncertainty: Heuristics and biases. In *Utility, probability, and human decision making*, pp. 141–162.

Appendix

A Proof of Proposition 1

Expression (9) can be written as:

$$f_t^{\theta}\left(\widehat{x}_{t+1}\right) \propto \exp\left[-\frac{1}{2\sigma_{t+h|t}^2}\left(x_{t+1}-\mu_{t+1|t}\right)^2\right]\left[\frac{\exp\left[-\frac{1}{2\sigma_{t+1|t}^2}\left(x_{t+1}-\mu_{t+1|t}\right)^2\right]}{\exp\left[-\frac{1}{2\sigma_{t+1|t-J}^2}\left(x_{t+1}-\mu_{t+1|t-J}\right)^2\right]}\right]^{\theta}\frac{1}{Z}$$

Collecting the terms in the exponents, we get:

$$f_t^{\theta}\left(\widehat{x}_{t+1}\right) \propto \exp\left[-\frac{1}{2\sigma_{t+1|t}^2}\left[\left(1+\theta\right)\left(x_{t+1}-\mu_{t+1|t}\right)^2-\frac{\sigma_{t+1|t}^2}{\sigma_{t+1|t-J}^2}\theta\left(x_{t+1}-\mu_{t+1|t-J}\right)^2\right]\right]\frac{1}{Z}$$

Developing the squared terms and keeping track of the terms involving x_{t+1}, we obtain:

$$f_t^{\theta}\left(\widehat{x}_{t+1}\right) \propto \exp\left[\begin{array}{c} -\frac{1}{2\sigma_{t+1|t}^2}\left(1+\theta-\frac{\sigma_{t+1|t}^2}{\sigma_{t+1|t-J}^2}\theta\right) \\ \left[x_{t+1}^2-2x_{t+1}\left(1+\theta-\frac{\sigma_{t+1|t}^2}{\sigma_{t+1|t-J}^2}\theta\right)^{-1}\left(\mu_{t+1|t}\left(1+\theta\right)-\frac{\sigma_{t+1|t}^2}{\sigma_{t+1|t-J}^2}\theta\mu_{t+1|t-J}\right)\right] \end{array}\right]\frac{1}{Z}$$

where the remaining terms are absorbed in the constant of integration.

Define: $R_{t+1|t,t-J} \equiv \frac{\sigma^2_{t+1|t}}{\sigma^2_{t+1|t-J}}$. If $R_{t+1|t,t-J} > (1+\theta)/\theta$, the expression above corresponds to the kernel of a normal with mean:

$$\begin{aligned}\mathbb{E}_t^\theta(x_{t+1}) &= \left(1+\theta-\frac{\sigma^2_{t+1|t}}{\sigma^2_{t+1|t-J}}\theta\right)^{-1}\left[\mu_{t+1|t}(1+\theta)-\frac{\sigma^2_{t+1|t}}{\sigma^2_{t+1|t-J}}\theta\mu_{t+1|t-J}\right] \\ &= \left[\mu_{t+1|t}+\frac{R_{t+1|t,t-J}\theta}{1+\left(1-R_{t+1|t,t-J}\right)\theta}\left(\mu_{t+1|t}-\mu_{t+1|t-J}\right)\right]\end{aligned}$$

and variance:

$$\begin{aligned}\mathbb{V}_t^\theta(x_{t+1}) &= \sigma^2_{t+1|t}\left(1+\theta-\frac{\sigma^2_{t+1|t}}{\sigma^2_{t+1|t-J}}\theta\right)^{-1} \\ &= \frac{\sigma^2_{t+1|t}}{1+\left(1-R_{t+1|t,t-J}\right)\theta}.\end{aligned}$$

This gives us the result stated in Proposition 2.

B Upper bound on DE distortion

Suppose that we are interested in imposing an upper bound on the Smooth DE distortion. Imposing such upper bound on the overreaction in the mean guarantees that both distortions remain finite and non-decreasing as the ratio $R_{t+h|t,t-J}$ goes to infinity. Thus, we propose the following approach.

Let $\overline{\overline{\theta}}$ be the desired upper bound of effective overreaction in conditional mean. By effective overreaction we refer to the object defined in equation (14). As a first step, we exploit the fact that the size of the distortion is increasing in $R_{t+h|t,t-J}$ to find the threshold value $\overline{R}$, such that, for a given θ, for each $R_{t+h|t,t-J} > \overline{R}$, the overreaction in the mean would be larger than $\overline{\overline{\theta}}$:

$$\frac{\overline{R}\theta}{1+\theta\left(1-\overline{R}\right)} = \overline{\overline{\theta}}$$

It follows that the upper threshold in terms of $R_{t+h|t,t-J}$ is

$$\overline{R} = \frac{\overline{\overline{\theta}}}{1+\overline{\overline{\theta}}}\frac{1+\theta}{\theta}$$

Whenever $R_{t+h|t,t-J} > \overline{R}$, we thus replace θ with θ_R, the value of θ such that the overreaction in the mean is equal to $\overline{\overline{\theta}}$. Thus, we solve:

$$\frac{R_{t+h|t,t-J}\theta_R}{1+\theta_R\left(1-R_{t+h|t,t-J}\right)} = \overline{\overline{\theta}}$$

and obtain:

$$\theta_R = \frac{\overline{\overline{\theta}}}{R_{t+h|t,t-J} - \overline{\overline{\theta}}\left(1-R_{t+h|t,t-J}\right)}$$

Plugging in θ_R in the formulas for the overreaction in mean and variance, we obtain:

$$\begin{aligned}
\mathbb{E}_t^{\theta}\left(x_{t+h}\right) &= \mu_{t+h|t} + \theta_R \frac{R_{t+h|t,t-J}}{1+\theta_R\left(1-R_{t+h|t,t-J}\right)}\left(\mu_{t+h|t} - \mu_{t+h|t-J}\right) \\
\mathbb{E}_t^{\theta}\left(x_{t+h}\right) &= \mu_{t+h|t} + \overline{\overline{\theta}}\left(\mu_{t+h|t} - \mu_{t+h|t-J}\right)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}_t^{\theta}\left(x_{t+h}\right) &= \frac{1}{1+\theta_R\left(1-R_{t+h|t,t-J}\right)}\sigma^2_{t+h|t} \\
\mathbb{V}_t^{\theta}\left(x_{t+h}\right) &= \left[1+\overline{\overline{\theta}}\left(1-\frac{1}{R_{t+h|t,t-J}}\right)\right]\sigma^2_{t+h|t}
\end{aligned}$$

Note that while the overreaction in the mean remains constant once $R_{t+h|t,t-J} > \overline{R}$, the overreaction in the variance keeps growing as relative uncertainty increases, but it converges to a finite value:

$$\lim_{R_{t+h|t,t-J}\to\infty} \mathbb{V}_t^{\theta}\left(x_{t+h}\right) = \left[1+\overline{\overline{\theta}}\right]\sigma^2_{t+h|t}$$

C Overreaction to new information and Smooth DE

To further understand Smooth DE, we rewrite the distorted conditional mean and variance in Proposition 1 as a function of the revised information, as follows.

Corollary 2 *(A revision representation). The Smooth DE density of Proposition 1 can be*

represented as distorting the RE revisions in conditional mean and variance, as follows:

$$\underbrace{\mathbb{E}_t^{\theta}\left(x_{t+h}\right)-\mu_{t+h|t-J}}_{\textit{Smooth DE revision}}=\underbrace{\left(\mu_{t+h|t}-\mu_{t+h|t-J}\right)}_{\textit{RE revision}}\underbrace{(1+\theta)}_{\textit{BGS effect}}\underbrace{\left[1+\theta\left(1-R_{t+h|t,t-J}\right)\right]^{-1}}_{\textit{Smooth DE effect}}$$

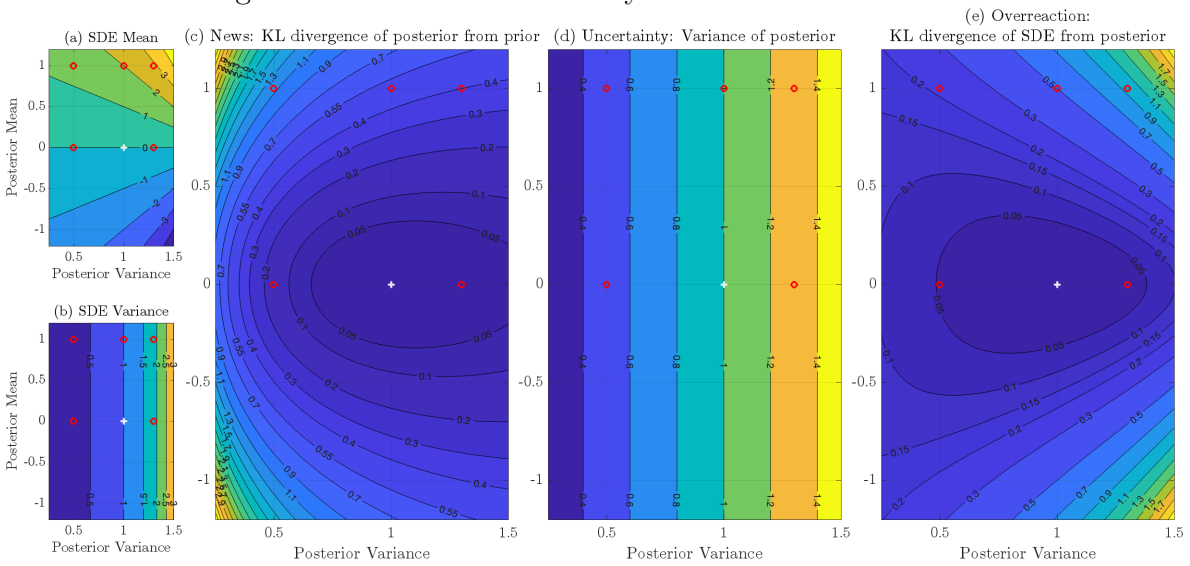
$$\underbrace{\frac{\mathbb{V}_t^{\theta}\left(x_{t+h}\right)}{\sigma^2_{t+h|t-J}}}_{\textit{Smooth DE revision}}=\underbrace{\frac{\sigma^2_{t+h|t}}{\sigma^2_{t+h|t-J}}}_{\textit{RE revision}}\underbrace{\left[1+\theta\left(1-R_{t+h|t,t-J}\right)\right]^{-1}}_{\textit{Smooth DE effect}}$$

This representation indicates how the revision in conditional moments under Smooth DE can be decomposed as having three parts: (1) the RE revision, (2) an overreaction effect from representativeness as assumed in the standard BGS implementation of DE, and (3) a separate and novel effect stemming from Smooth DE.

D News and uncertainty effects for the Normal density

In Figure 4, we study how news and uncertainty effects determine the overreaction of the Smooth DE density relative to the true current distribution. To facilitate comparison with Figure 3, we mark the coordinates corresponding to the reference mean and variance with white pluses, and the coordinates corresponding to the current means and variances of each scenario in Figure 3 with red circles. Panels (a) and (b) visually illustrate how the Smooth DE mean and variance change as the posterior mean and variance change and confirm our results in Corollary 1: Panel (a) shows larger overreactions for higher current variances and Panel (b) highlights overconfidence for current variances lower than the reference variance and underconfidence for current variances higher than the reference variance. Panel (c) shows the news effect, measured in terms of the KL divergence of the current distribution from the reference distribution. Interestingly, the KL divergence is large when the current variance is low and there is a large shift in the mean. The news effect is large in that case because we are moving probability masses of tail events under the reference distribution. In contrast, the uncertainty effect is large when the current variance is high (Panel (d)). Panel (e) shows that the uncertainty effect of higher current variance dominates the news effect, so the overreaction of the Smooth DE density (measured in terms of the KL divergence of the Smooth DE density to current density) is largest when there is a large shift in the mean and an increase in variance.

Figure 4: News and uncertainty in a Normal distribution



Notes: The figure displays how, as we vary the posterior mean and variance, the following objects change: (a) the Smooth DE mean, (b) the Smooth DE variance, (c) the news component, measured using the KL divergence of the posterior distribution from the prior distribution, (d) the uncertainty component, measured using the posterior variance, and (e) the overall overreaction, measured using the KL divergence of the Smooth DE distribution from the posterior distribution. We mark the coordinates corresponding to the prior mean and variance with white pluses, and the coordinates corresponding to the posterior means and variances of each scenario in Figure 3 with red circles.

E Signal extraction under Smooth DE

An important class of models emphasizing changes in subjective uncertainty belongs to the large literature on Bayesian learning. In what follows, we show that Smooth DE can be easily extended to this class of models.

Let us start with some more general notation, which connects to the one used in Definition 1. Consider the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ introduced in Section 3.1. But now allow for noisy information. In particular, let the information available up to time t be represented by another filtration $(\mathcal{G}_t)_{t\geq 0}$, generated by an observed process s_t. Then let $f(x_{t+h}|\mathcal{G}_t)$ be the conditional density function of X_{t+h} based on the information available in $\mathcal{G}_t$ from the imperfect observations.

Like in Definitions 1 and 2, the representativeness of an event would then become

$$rep\left(\widehat{x}_{t+h}|\mathcal{G}_t, \mathcal{G}_{t-J}\right) \equiv \frac{f\left(\widehat{x}_{t+h}|\mathcal{G}_t\right)}{f\left(\widehat{x}_{t+h}|\mathcal{G}_{t-J}\right)} \tag{38}$$

while the conditional density distorted by representativeness is

$$f^{\theta}\left(\widehat{x}_{t+h}|\mathcal{G}_t,\mathcal{G}_{t-J}\right)=f\left(\widehat{x}_{t+h}|\mathcal{G}_t\right)\left[rep\left(\widehat{x}_{t+h}|\mathcal{G}_t,\mathcal{G}_{t-J}\right)\right]^{\theta}Z^{-1} \tag{39}$$

where Z is a constant of integration and the parameter $\theta \geq 0$.

Smooth Diagnostic Kalman filter. To derive closed form solutions, we focus on a standard Gaussian case of noisy information and maintain $J = 1$. In particular, consider a standard state-space representation. The observation equation is:

$$s_t = x_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

and the state transition equation for the unobserved x_t is

$$x_t = \rho x_{t-1} + u_t, \qquad u_t \sim N(0, \sigma_u^2)$$

The Kalman Filter gives the Bayesian forecast, and its derivation is standard. The one-step-ahead prediction from the period $t-1$ estimate $\tilde{x}_{t-1|t-1}$ and its associated error variance $\Sigma_{t-1|t-1}$ are given by

$$\tilde{x}_{t|t-1} = \rho\tilde{x}_{t-1|t-1}; \Sigma_{t|t-1} = \rho^2\Sigma_{t-1|t-1} + \sigma_u^2.$$

Then, the estimates are updated according to

$$\tilde{x}_{t|t} = \tilde{x}_{t|t-1} + K_t(s_t - \tilde{x}_{t|t-1}), \qquad K_t = \frac{\Sigma_{t|t-1}}{\Sigma_{t|t-1} + \sigma_\varepsilon^2},$$

where K_t is the Kalman gain and the updating rule for the variance is

$$\Sigma_{t|t} = \left[\frac{\sigma_\varepsilon^2}{\Sigma_{t|t-1} + \sigma_\varepsilon^2}\right]\Sigma_{t|t-1}. \tag{40}$$

In this environment, we now derive the version of Kalman filter used by agents subject to Smooth DE. This connects and extends the diagnostic Kalman Filter derived within the standard BGS formulation in earlier work like Bordalo et al. (2019) and Bordalo et al. (2020).

Let $f(x_t|\mathcal{G}_t)$ be the probability density of the rational, or Bayesian, period t estimate of the current underlying state x_t based on the Kalman Filter derived above. Intuitively, by equation (38), a state x_t is more representative if it becomes more likely relative to the $t-1$ forecast. As in our discussion of equation (9), the key feature with respect to the original BGS formulation is to condition on the whole past information set, and as a result, to take

into account the associated *uncertainty*.

Let the ratio of current to prior estimation uncertainty under RE be denoted as

$$R_{t|t,t-1} \equiv \Sigma_{t|t}/\Sigma_{t|t-1}.$$

We derive the following result.

Proposition 10 *(Smooth DE Kalman Filter.) The density* $f^{\theta}\left(\widehat{x}_t|\mathcal{G}_t, \mathcal{G}_{t-1}\right)$ *in equation (39) has a Normal distribution with mean*

$$\mathbb{E}_t^{\theta}\left(x_t\right) = \tilde{x}_{t|t} + \frac{R_{t|t,t-1}\theta}{1+\left(1-R_{t|t,t-1}\right)\theta}\left(\tilde{x}_{t|t} - \tilde{x}_{t|t-1}\right), \tag{41}$$

and variance

$$\mathbb{V}_t^{\theta}\left(x_t\right) = \frac{\Sigma_{t|t}}{1+\left(1-R_{t|t,t-1}\right)\theta} \tag{42}$$

Proof. Re-writing the expression (39):

$$f_t^{\theta}\left(x_t\right) \propto \exp\left[-\frac{1}{2\Sigma_{t|t}}\left(x_t - \tilde{x}_{t|t}\right)^2\right]\left[\frac{\exp\left[-\frac{1}{2\Sigma_{t|t}}\left(x_t - \tilde{x}_{t|t}\right)^2\right]}{\exp\left[-\frac{1}{2\Sigma_{t|t-1}}\left(x_t - \tilde{x}_{t|t-1}\right)^2\right]}\right]^{\theta}\frac{1}{Z}$$

Collecting the terms in the exponents, we get:

$$f_t^{\theta}\left(x_t\right) \propto \exp\left[-\frac{1}{2\Sigma_{t|t}}\left[\left(1+\theta\right)\left(x_t - \tilde{x}_{t|t}\right)^2 - \frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\left(x_t - \tilde{x}_{t|t-1}\right)^2\right]\right]\frac{1}{Z}$$

Developing the squared terms and keeping track of the terms involving x_t, we obtain:

$$f_t^{\theta}\left(x_t\right) \propto \exp\left[\begin{array}{c} -\frac{1}{2\Sigma_{t|t}}\left(1+\theta-\frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\right) \\ \left[x_t^2 - 2x_t\left(1+\theta-\frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\right)^{-1}\left(\left(1+\theta\right)\tilde{x}_{t|t} - \frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\tilde{x}_{t|t-1}\right)\right] \end{array}\right]\frac{1}{Z}$$

where the remaining terms are absorbed in the constant of integration. The one above is

the kernel of a normal with mean:

$$\begin{aligned}\mathbb{E}_t^\theta(x_t) &= \left(1+\theta-\frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\right)^{-1}\left((1+\theta)\,\tilde{x}_{t|t}-\frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\tilde{x}_{t|t-1}\right)\\ &= \tilde{x}_{t|t}+\frac{R_{t|t,t-1}\theta}{1+\left(1-R_{t|t,t-1}\right)\theta}\left(\tilde{x}_{t|t}-\tilde{x}_{t|t-1}\right)\\ &= \tilde{x}_{t|t-1}+\left(1+\frac{R_{t|t,t-1}\theta}{1+\left(1-R_{t|t,t-1}\right)\theta}\right)\left(\tilde{x}_{t|t}-\tilde{x}_{t|t-1}\right)\\ &= \tilde{x}_{t|t-1}+\left(1+\frac{R_{t|t,t-1}\theta}{1+\left(1-R_{t|t,t-1}\right)\theta}\right)K_t(s_t-\tilde{x}_{t|t-1}),\end{aligned}$$

where $R_{t|t,t-1}\equiv\Sigma_{t|t}/\Sigma_{t|t-1}$ and in the fourth line we used (E), and variance:

$$\begin{aligned}\mathbb{V}_t^\theta(x_t) &= \Sigma_{t|t}\left(1+\theta-\frac{\Sigma_{t|t}}{\Sigma_{t|t-1}}\theta\right)^{-1}\\ &= \frac{\Sigma_{t|t}}{1+\left(1-R_{t|t,t-1}\right)\theta}.\end{aligned}$$

This gives us the result stated in Proposition 10. ■

Like in our earlier general discussion, we observe overreaction of the conditional mean when $\theta>0$ and the new information does not fully resolve uncertainty, i.e. when $\sigma_\varepsilon^2>0$. Furthermore, similarly to the earlier AR(1) example, this environment is also characterized by a conditional reduction in uncertainty, and therefore by overconfidence. Indeed, as long as σ_ε^2 is finite, by equation (40) estimation uncertainty decreases over time, as the new signal is at least partly informative. It follows that the ratio $R_{t|t,t-1}<1,\forall t$ and that given equation (42), subjective uncertainty is lower than Bayesian estimation uncertainty, i.e. $\mathbb{V}_t^\theta(x_t)<\Sigma_{t|t}$.

F Proofs for the business cycle model

F.1 Proof of Proposition 4

First, consider the equilibrium individual policy functions. To characterize dynamics we use a log-linear approximation of decision rules around the steady state. We take logs of the optimality condition with respect to hours in (28) and constraints (26) and (27):

$$\begin{aligned}\eta\widehat{h}_{i,t} &= \mathbb{E}_{i,t}\left[-\gamma\widehat{c}_{i,t+1}+\widehat{z}_{i,t+1}\right],\\ \widehat{y}_{i,t} &= \widehat{z}_{i,t}+\widehat{h}_{i,t-1}=\widehat{c}_{i,t}.\end{aligned}$$

Substitute the constraints into the labor supply condition:

$$\begin{aligned}
\eta\widehat{h}_{i,t} &= \mathbb{E}_{i,t}\left[-\gamma\widehat{c}_{i,t+1} + \widehat{z}_{i,t+1}\right] \\
&= \mathbb{E}_{i,t}\left[-\gamma\left(\widehat{z}_{i,t+1} + \widehat{h}_{i,t}\right) + \widehat{z}_{i,t+1}\right] \\
\widehat{h}_{i,t} &= \frac{1-\gamma}{\eta+\gamma}\mathbb{E}_{i,t}\left[\widehat{z}_{i,t+1}\right] \\
&= \frac{1-\gamma}{\eta+\gamma}\left[\rho_A A_t + \widetilde{a}_{i,t+1|t}\right] \\
&= \frac{1-\gamma}{\eta+\gamma}\left[\rho_A A_t + s_{i,t}\right].
\end{aligned}$$

Equating the coefficients we obtain the equilibrium elasticities.

Next, consider aggregate variables. Note we have

$$\int_0^1 s_{i,t}di = 0, \quad \int_0^1 \widehat{z}_{i,t}di = A_t + \int_0^1 a_{i,t}di = A_t,$$

by law of large numbers. Then

$$\begin{aligned}
\widehat{H}_t &= \int_0^1 \widehat{h}_{i,t}di = \varepsilon\rho_A A_t + \varepsilon\int_0^1 s_{i,t}di \\
&= \varepsilon\rho_A A_t \\
\widehat{Y}_t &= \int_0^1 \widehat{y}_{i,t}di = \int_0^1 \widehat{z}_{i,t}di + \int_0^1 \widehat{h}_{i,t-1}di \\
&= A_t + \widehat{H}_{t-1} \\
&= \widehat{C}_t.
\end{aligned}$$

F.2 Proof of Proposition 5

Note we have

$$\int_0^1 s_{i,t}^2 di = \sigma_s^2, \quad \int_0^1 u_{a,i,t}^2 di = \sigma_{a,t-1}^2, \quad \int_0^1 a_{i,t}^2 di = \int_0^1 (s_{i,t-1} + u_{a,i,t})^2 di = \sigma_s^2 + \sigma_{a,t-1}^2.$$

Then

$$\begin{aligned}
\int_0^1 \left(\widehat{h}_{i,t} - \widehat{H}_t\right)^2 di &= \int_0^1 (\varepsilon s_{i,t})^2 di = (\varepsilon)^2 \int_0^1 s_{i,t}^2 di \\
&= (\varepsilon)^2 \sigma_s^2,
\end{aligned}$$

which is constant. Next consider the cross-sectional variance of output:

$$\begin{aligned}\int_0^1 \left(\widehat{y}_{i,t} - \widehat{Y}_t\right)^2 di &= \int_0^1 \left(\left(A_t + a_{i,t} + \widehat{h}_{i,t-1}\right) - \left(A_t + \widehat{H}_{t-1}\right)\right)^2 di \\ &= \int_0^1 \left(a_{i,t} + \varepsilon s_{i,t-1}\right)^2 di \\ &= \int_0^1 \left(s_{i,t-1} + u_{a,i,t} + \varepsilon s_{i,t-1}\right)^2 di \\ &= (1+\varepsilon)^2 \int_0^1 s_{i,t-1}^2 di + \int_0^1 u_{a,i,t}^2 di \\ &= (1+\varepsilon)^2 \sigma_s^2 + \sigma_{a,t-1}^2,\end{aligned}$$

which is increasing in $\sigma_{a,t-1}^2$. It follows that the cross-sectional variance of consumption:

$$\int_0^1 \left(\widehat{c}_{i,t} - \widehat{C}_t\right)^2 di = (1+\varepsilon)^2 \sigma_s^2 + \sigma_{a,t-1}^2,$$

is increasing in $\sigma_{a,t-1}^2$ as well.

F.3 Proof of Proposition 6

First, consider the equilibrium individual policy functions. As in the RE solution, to characterize dynamics we use a log-linear approximation of decision rules around the steady state. We take logs of the optimality condition with respect to hours in (29) and constraints (26) and (27):

$$\begin{aligned}\eta \widehat{h}_{i,t}^{\theta} &= \mathbb{E}_{i,t}^{\theta}\left[-\gamma \widehat{c}_{i,t+1}^{RE} + \widehat{z}_{i,t+1}\right] \\ \widehat{y}_{i,t}^{\theta} &= \widehat{z}_{i,t} + \widehat{h}_{i,t-1}^{\theta} = \widehat{c}_{i,t}^{\theta}.\end{aligned}$$

Substitute the constraints into the labor supply condition:

$$\begin{aligned}
\eta\widehat{h}_{i,t}^{\theta} &= (1+\widetilde{\theta}_{t,t-J})\mathbb{E}_{i,t}\left[-\gamma\widehat{c}_{i,t+1}^{RE}+\widehat{z}_{i,t+1}\right]-\widetilde{\theta}_{t,t-J}\mathbb{E}_{i,t-J}\left[-\gamma\widehat{c}_{i,t+1}^{RE}+\widehat{z}_{i,t+1}\right]\\
\eta\widehat{h}_{i,t}^{\theta} &= (1+\widetilde{\theta}_{t,t-J})\mathbb{E}_{i,t}\left[-\gamma\left(\widehat{z}_{i,t+1}+\widehat{h}_{i,t}^{\theta}\right)+\widehat{z}_{i,t+1}\right]\\
&\quad-\widetilde{\theta}_{t,t-J}\mathbb{E}_{i,t-J}\left[-\gamma\left(\widehat{z}_{i,t+1}+\widehat{h}_{i,t}^{RE}\right)+\widehat{z}_{i,t+1}\right]\\
\left[\eta+\left(1+\widetilde{\theta}_{t,t-J}\right)\gamma\right]\widehat{h}_{i,t}^{\theta} &= (1+\widetilde{\theta}_{t,t-J})(1-\gamma)\,\mathbb{E}_{i,t}\left[\widehat{z}_{i,t+1}\right]-\widetilde{\theta}_{t,t-J}(1-\gamma)\,\mathbb{E}_{i,t-J}\left[\widehat{z}_{i,t+1}\right]+\widetilde{\theta}_{t,t-J}\gamma\mathbb{E}_{i,t-J}\left[\widehat{h}_{i,t}^{RE}\right]\\
\left[\eta+\left(1+\widetilde{\theta}_{t,t-J}\right)\gamma\right]\widehat{h}_{i,t}^{\theta} &= (1+\widetilde{\theta}_{t,t-J})(1-\gamma)\left[\rho_A A_t+s_{i,t}\right]\\
&\quad-\widetilde{\theta}_{t,t-J}(1-\gamma)\rho_A^{J+1}A_{t-J}+\widetilde{\theta}_{t,t-J}\gamma\left[\varepsilon\rho_A^{J+1}A_{t-J}\right]\\
\widehat{h}_{i,t}^{\theta} &= \frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta+\left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\rho_A A_t+\frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta+\left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}s_{i,t}\\
&\quad-\frac{\widetilde{\theta}_{t,t-J}\eta}{\eta+\left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\left[\frac{1-\gamma}{\eta+\gamma}\right]\rho_A^{J+1}A_{t-J},
\end{aligned}$$

where the effective diagnosticity parameter $\widetilde{\theta}_{t,t-J}$ is given by (32). Equating the coefficients we obtain the equilibrium elasticities. As in the RE economy, we obtain equilibrium aggregate variables by simply aggregating the individual policy functions.

F.4 Proof of Proposition 7

Consider $R_{t+1|t,t-J}$:

$$\begin{aligned}
R_{t+1|t,t-J} &= \frac{\mathbb{V}_{i,t}\left(-\gamma\widehat{c}_{i,t+1}^{RE}+\widehat{z}_{i,t+1}\right)}{\mathbb{V}_{i,t-J}\left(-\gamma\widehat{c}_{i,t+1}^{RE}+\widehat{z}_{i,t+1}\right)}\\
&= \frac{\mathbb{V}_{i,t}\left(-\gamma\left(\widehat{z}_{i,t+1}+\widehat{h}_{i,t}^{\theta}\right)+\widehat{z}_{i,t+1}\right)}{\mathbb{V}_{i,t-J}\left(-\gamma\left(\widehat{z}_{i,t+1}+\widehat{h}_{i,t}^{RE}\right)+\widehat{z}_{i,t+1}\right)}\\
&= \frac{\mathbb{V}_{i,t}\left((1-\gamma)\,\widehat{z}_{i,t+1}-\gamma\widehat{h}_{i,t}^{\theta}\right)}{\mathbb{V}_{i,t-J}\left((1-\gamma)\,\widehat{z}_{i,t+1}-\gamma\widehat{h}_{i,t}^{RE}\right)},
\end{aligned}$$

where the numerator is

$$\mathbb{V}_{i,t}\left((1-\gamma)\left[\rho_A A_t+u_{A,t+1}+s_{i,t}+u_{a,i,t+1}\right]-\gamma\widehat{h}_{i,t}^{\theta}\right)=(1-\gamma)^2\left(\sigma_A^2+\sigma_{a,t}^2\right),$$

which is increasing in $\sigma_{a,t}^2$. Thus $R_{t+1|t,t-J}$ and in turn $\widetilde{\theta}_{t,t-J}$ are increasing in $\sigma_{a,t}^2$.

F.5 Proof of Proposition 8

First consider the cross-sectional variance of hours. Defining $\varepsilon_{s,t}^{\theta} \equiv \frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta+(1+\widetilde{\theta}_{t,t-J})\gamma}$, we have

$$\begin{aligned}\int_0^1 \left(\widehat{h}_{i,t}^{\theta} - \widehat{H}_t^{\theta}\right)^2 di &= \int_0^1 \left(\varepsilon_{s,t}^{\theta} s_{i,t}\right)^2 di = \left(\varepsilon_{s,t}^{\theta}\right)^2 \int_0^1 s_{i,t}^2 di \\ &= \left[\frac{\left(1+\widetilde{\theta}_{t,t-J}\right)(1-\gamma)}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\right]^2 \sigma_s^2,\end{aligned}$$

which is increasing in $\widetilde{\theta}_{t,t-J}$. Next consider the cross-sectional variance of output:

$$\begin{aligned}&\int_0^1 \left(\widehat{y}_{i,t}^{\theta} - \widehat{Y}_t^{\theta}\right)^2 di \\ &= \int_0^1 \left(\left(A_t + a_{i,t} + \widehat{h}_{i,t-1}^{\theta}\right) - \left(A_t + \widehat{H}_{t-1}^{\theta}\right)\right)^2 di \\ &= \int_0^1 \left(s_{i,t-1} + u_{a,i,t} + \varepsilon_{s,t-1}^{\theta} s_{i,t-1}\right)^2 di \\ &= \left(1 + \left[\frac{\left(1+\widetilde{\theta}_{t,t-J}\right)(1-\gamma)}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\right]\right)^2 \sigma_s^2 + \sigma_{a,t-1}^2\end{aligned}$$

which is increasing in $\widetilde{\theta}_{t,t-J}$ and $\sigma_{a,t-1}^2$. It follows that the cross-sectional variance of consumption:

$$\int_0^1 \left(\widehat{c}_{i,t}^{\theta} - \widehat{C}_t^{\theta}\right)^2 di \left(1 + \left[\frac{\left(1+\widetilde{\theta}_{t,t-J}\right)(1-\gamma)}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\right]\right)^2 \sigma_s^2 + \sigma_{a,t-1}^2$$

is increasing in $\widetilde{\theta}_{t,t-J}$ and $\sigma_{a,t-1}^2$ as well.

F.6 Proof of Proposition 9

First, we solve for the log-linearized RE decision rules under the tax policy. The optimality conditions are

$$\begin{aligned}&\eta\widehat{h}_{i,t} = \mathbb{E}_{i,t}\left[-\gamma\widehat{c}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\ &\widehat{y}_{i,t} = \widehat{z}_{i,t} + \widehat{h}_{i,t-1} \\ &\widehat{c}_{i,t} + \tau a_{i,t} = \widehat{y}_{i,t}.\end{aligned}$$

Substitute the constraints into labor supply conditions:

$$\begin{aligned}
\eta\widehat{h}_{i,t} &= \mathbb{E}_{i,t}\left[-\gamma\widehat{c}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
&= \mathbb{E}_{i,t}\left[-\gamma\left(A_{t+1} + (1-\tau)a_{i,t+1} + \widehat{h}_{i,t}\right) + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
\widehat{h}_{i,t} &= \frac{1-\gamma}{\eta+\gamma}\mathbb{E}_{i,t}\left[A_{i,t+1}\right] + \frac{1-\gamma}{\eta+\gamma}(1-\tau)\mathbb{E}_{i,t}\left[a_{i,t+1}\right] \\
&= \frac{1-\gamma}{\eta+\gamma}\left[\rho_A A_t + (1-\tau)\widetilde{a}_{i,t+1|t}\right] \\
&= \frac{1-\gamma}{\eta+\gamma}\left[\rho_A A_t + (1-\tau)s_{i,t}\right]
\end{aligned}$$

Equilibrium output and consumption follow immediately as

$$\widehat{y}_{i,t} = \widehat{z}_{i,t} + \widehat{h}_{i,t-1} = A_t + a_{i,t} + \widehat{h}_{i,t-1}, \tag{43}$$

$$\widehat{c}_{i,t} = \widehat{y}_{i,t} - \tau a_{i,t} = A_t + (1-\tau)a_{i,t} + \widehat{h}_{i,t-1}. \tag{44}$$

Next, consider the log-linearized SDE decision rules under the tax policy. To characterize dynamics we use a log-linear approximation of decision rules around the steady state. The optimality conditions are

$$\begin{aligned}
&\eta\widehat{h}^{\theta}_{i,t} = \mathbb{E}^{\theta}_{i,t}\left[-\gamma\widehat{c}^{RE}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right], \\
&\widehat{y}^{\theta}_{i,t} = \widehat{z}_{i,t} + \widehat{h}^{\theta}_{i,t-1}, \\
&\widehat{c}^{\theta}_{i,t} + \tau a_{i,t} = \widehat{y}^{\theta}_{i,t}.
\end{aligned}$$

Substitute the constraints into the labor supply condition:

$$\begin{aligned}
\eta\widehat{h}^{\theta}_{i,t} &= (1+\widetilde{\theta}_{t,t-J})\mathbb{E}_{i,t}\left[-\gamma\widehat{c}^{RE}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right] - \widetilde{\theta}_{t,t-J}\mathbb{E}_{i,t-J}\left[-\gamma\widehat{c}^{RE}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
\eta\widehat{h}^{\theta}_{i,t} &= (1+\widetilde{\theta}_{t,t-J})\mathbb{E}_{i,t}\left[-\gamma\left(A_{t+1} + (1-\tau)a_{i,t+1} + \widehat{h}^{\theta}_{i,t}\right) + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
&\quad - \widetilde{\theta}_{t,t-J}\mathbb{E}_{i,t-J}\left[-\gamma\left(A_{t+1} + (1-\tau)a_{i,t+1} + \widehat{h}^{RE}_{i,t}\right) + A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
\left[\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma\right]\widehat{h}^{\theta}_{i,t} &= (1+\widetilde{\theta}_{t,t-J})(1-\gamma)\,\mathbb{E}_{i,t}\left[A_{t+1} + (1-\tau)a_{i,t+1}\right] \\
&\quad - \widetilde{\theta}_{t,t-J}(1-\gamma)\,\mathbb{E}_{i,t-J}\left[A_{t+1} + (1-\tau)a_{i,t+1}\right] + \widetilde{\theta}_{t,t-J}\gamma\mathbb{E}_{i,t-J}\left[\widehat{h}^{RE}_{i,t}\right] \\
\left[\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma\right]\widehat{h}^{\theta}_{i,t} &= (1+\widetilde{\theta}_{t,t-J})(1-\gamma)\left[\rho_A A_t + (1-\tau)s_{i,t}\right] \\
&\quad - \widetilde{\theta}_{t,t-J}(1-\gamma)\,\rho_A^{J+1}A_{t-J} + \widetilde{\theta}_{t,t-J}\gamma\varepsilon\rho_A^{J+1}A_{t-J} \\
\widehat{h}^{\theta}_{i,t} &= \frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\rho_A A_t + \frac{(1+\widetilde{\theta}_{t,t-J})(1-\gamma)}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}(1-\tau)s_{i,t} \\
&\quad - \frac{\widetilde{\theta}_{t,t-J}\eta}{\eta + \left(1+\widetilde{\theta}_{t,t-J}\right)\gamma}\left[\frac{1-\gamma}{\eta+\gamma}\right]\rho_A^{J+1}A_{t-J},
\end{aligned}$$

where the effective diagnosticity parameter $\widetilde{\theta}_{t,t-J}$ is given by (32).

Consider $R_{t+1|t,t-J}$:

$$\begin{aligned}
R_{t+1|t,t-J} &= \frac{\mathbb{V}_{i,t}\left(-\gamma \widehat{c}^{RE}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right)}{\mathbb{V}_{i,t-J}\left(-\gamma \widehat{c}^{RE}_{i,t+1} + A_{t+1} + (1-\tau)a_{i,t+1}\right)} \\
&= \frac{\mathbb{V}_{i,t}\left(-\gamma\left(A_{t+1} + (1-\tau)a_{i,t+1} + \widehat{h}^{\theta}_{i,t}\right) + A_{t+1} + (1-\tau)a_{i,t+1}\right)}{\mathbb{V}_{i,t-J}\left(-\gamma\left(A_{t+1} + (1-\tau)a_{i,t+1} + \widehat{h}^{RE}_{i,t}\right) + A_{t+1} + (1-\tau)a_{i,t+1}\right)} \\
&= \frac{\mathbb{V}_{i,t}\left((1-\gamma)\left(A_{t+1} + (1-\tau)a_{i,t+1}\right) - \gamma \widehat{h}^{\theta}_{i,t}\right)}{\mathbb{V}_{i,t-J}\left((1-\gamma)\left(A_{t+1} + (1-\tau)a_{i,t+1}\right) - \gamma \widehat{h}^{RE}_{i,t}\right)}
\end{aligned}$$

where the numerator is

$$\begin{aligned}
&\mathbb{V}_{i,t}\left((1-\gamma)\left[\rho_A A_t + u_{A,t+1} + (1-\tau)s_{i,t} + (1-\tau)u_{a,i,t+1}\right] - \gamma \widehat{h}^{\theta}_{i,t}\right) \\
&= (1-\gamma)^2\left(\sigma^2_A + (1-\tau)^2\sigma^2_{a,t}\right),
\end{aligned}$$

which is increasing in $\sigma^2_{a,t}$ but also a change in $\sigma^2_{a,t}$ have a smaller impact when the progressivity τ is higher. Thus a higher τ is associated with a smaller increase in $R_{t+1|t,t-J}$ and $\widetilde{\theta}_{t,t-J}$.