Diagnostic Business Cycles *

Francesco Bianchi
Johns Hopkins, CEPR & NBER

Cosmin Ilut
Duke & NBER

Hikaru Saijo
UC Santa Cruz

June 2022

Abstract

A large psychology literature argues that, due to selective memory recall, decision-makers' forecasts of their future circumstances appear overly influenced by the new information embedded in their current circumstances. We adopt the diagnostic expectations (DE) paradigm (Bordalo et al. (2018)) to capture this feature of belief formation and develop the behavioral foundations for applying DE to business cycle models, while demonstrating its empirical relevance for aggregate dynamics. First, we address (i) the theoretical challenges associated with modeling the feedback between optimal actions and agents' DE beliefs and (ii) the time-inconsistencies that arise under distant memory. Second, we show that under distant memory the interaction between actions and DE beliefs naturally generate repeated boom-bust cycles in response to a single initial shock. Finally, we propose a portable solution method to study DE in dynamic stochastic general equilibrium models and use it to estimate a quantitative DE New Keynesian model. Both endogenous states and distant memory play a critical role in successfully replicating the boom-bust cycle observed in response to a monetary policy shock.

Keywords: Diagnostic expectations, boom-bust cycles, imperfect memory recall.
JEL Codes: D91, E21, E32, E71.

*Email addresses: Bianchi francesco.bianchi@jhu.edu, Ilut cosmin.ilut@duke.edu, Saijo hsaijo@ucsc.edu.
This paper supersedes an earlier version, titled "Implications of Diagnostic Expectations: Theory and Applications". We would like to thank the editor Nicola Gennaioli and four anonymous referees, as well as Daniele d'Arienzo, Spencer Kwon, Jean-Paul L'Huillier, Peter Maxted, Sanjay Singh, Andrei Shleifer, and Stephen Terry for helpful comments and all seminar participants at the 2021 NBER Summer Institute, the spring 2022 meeting of the NBER Monetary Economics group, Emory University, UC Davis, University of Copenhagen, University of Maryland, 2021 SED annual meeting, Hitotsubashi University, Keio University, and Osaka University.

1 Introduction

A large psychology and experimental literature documents that decision-makers' forecasts of their future circumstances appear overly influenced by the surprises embedded in their current circumstances. In economics, this feature of belief formation has been captured by the diagnostic expectations (DE) paradigm, formulated recently by Bordalo et al. (2018), building on the representativeness heuristic by Kahneman and Tversky (1972). While promising in the breadth of its potential implications, so far the DE paradigm has focused on environments where the surprise is determined exogenously and perceived with respect to a reference belief in the immediate past. However, these two characteristics appear overly restrictive in applications because decisions often involve a feedback between agents' beliefs and endogenously determined economic states, and empirical evidence indicates that the selective memory recall may be based on more distant information sets.[1]

Motivated by these observations, our paper makes three main contributions. First, we develop micro-foundations to jointly address the theoretical challenges associated with modeling (i) the feedback between optimal actions and agents' DE beliefs over both exogenous and *endogenous* variables, and (ii) the time-inconsistencies in those optimal actions that arise when selective memory recall is based on a more *distant* past, rather than just the immediate past. Second, we show that under distant memory, but not under recent memory, the interaction between actions and DE beliefs naturally generate *endogenous* repeated boom-bust cycles in response to a single initial shock. Third, we build on these foundations to propose a *portable solution method* to study DE in linear recursive macroeconomic models, which can thus accommodate large-scale dynamic stochastic general equilibrium models. We leverage the tractability of our proposed method to incorporate DE into an estimated quantitative New Keynesian model of the type widely used for policy analysis. We find that DE about endogenous states and distant memory recall play a critical role in replicating the empirically documented boom-bust cycle in response to a monetary policy shock.

In the recent formulation of Gennaioli and Shleifer (2010) and Bordalo et al. (2018), under the assumption of normality of the data generating process, DE distort current rational expectations (RE) with a term that depends on the difference between current RE (the representative, or diagnostic group) and lagged RE (the reference, or comparison group). Thus, the size of distortion is proportional to the revision in RE (or the representative information). In Bordalo et al. (2018), this idea is formalized in terms of two parameters. A

[1]For example, Bordalo et al. (2020b) find that a reference belief based on the four quarters ago information set seems to account well for the empirical over-reaction observed in the surveys of professional forecasters, while Bordalo et al. (2019b) argue that the sluggishness in expected returns is best explained by a reference information set eleven quarters in the past.

parameter θ controls the severity of the distortion, while a parameter J controls the lag of those reference beliefs.

Applied theory contributions. In the first part of the paper, we use a three-period consumption-savings model to analyze the implications of DE beliefs under endogenous states and distant memory recall. We identify two theoretical challenges.

First, in an economic model, a future uncertain object (such as consumption) typically depends both on future exogenous forces (such as future income shocks) and on the agent's current endogenous actions (such as current savings). We emphasize how this latter, additional source of dependence, which we label *endogenous predictability*, raises the challenge of jointly determining forward-looking optimal actions and DE forecasts. We micro-found this joint determination as the outcome of an intra-person Nash equilibrium between a *Memory* and *Deliberation* self. We argue that this approach is consistent with recent evidence provided by Bordalo et al. (2021) showing that memory is to some extent spontaneous: the new relevant data (e.g., current income shock) and the particular hypotheses under consideration (e.g., conditional density over future consumption), jointly bring to the top of the agent's mind events that associate that relevant data with the hypothesis, without necessarily disentangling the particular sources behind this event association.

Second, when the reference point for the DE distortion depends on the distant past ($J > 1$), as opposed to the immediate past ($J = 1$), *the law of iterated expectations (LOIE) fails*. In a multi-period model, the failure of the LOIE is important because it leads to time-inconsistent choices, as a result of the change in beliefs induced by imperfect memory. To address the issue of time inconsistency, we adopt the *naïveté* approach (e.g. O'Donoghue and Rabin (1999)). 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. We find the naïveté approach psychologically coherent and more consistent with the underlying foundation of diagnostic beliefs as a heuristic and a mental short-cut than the alternative approach of *sophistication*. Under sophistication, the agent fully understands how imperfect memory recall changes her future preferences. Furthermore, the naïveté approach turns out to be computationally more tractable, since the current naive agent does not need to internalize the life-time indirect effect of the current action on the formation of future comparison groups.

The consumption-savings model yields two critical insights. First, the DE agent exhibits a *higher* marginal propensity to consume (MPC) out of temporary income shocks relative to the RE agent.[2] The current high income innovation leads the agent to make a consumption-

[2]Jappelli and Pistaferri (2010), Kueng (2018), Fagereng et al. (2020) and McDowall (2020) find that the

savings decision under an 'as if' overly optimistic view of future consumption. This perception arises because memory is overly influenced by the statistical association between periods of high income realizations and subsequent high consumption realizations. Memory is not cued to distinguish whether this association arises because high income realizations are associated with high subsequent income realizations (an effect that is in fact absent with temporary shocks) or with periods of high savings. In both cases, what memory spontaneously recalls is that high future consumption realizations typically follow ('good times' are typically followed by 'good times'). Second, when memory recall is based on the more distant past ($J > 1$), the surprise in current capital relative to past expectations emerges as a novel informational state that determines the agent's decisions. This is because the current level of capital is *not* a sufficient statistic for the comparison group in forming DE.

We showcase the importance of this novel informational state dynamics by studying an infinite-horizon extension of our setup: the permanent income hypothesis (PIH) model. Under distant memory, a single, temporary income shock can generate endogenous, repeated boom-bust cycles because past actions feed into current beliefs, that in turn inform new actions. Following a positive iid income shock, the income surprise creates optimism about the future, leading to over-consumption. This initial over-consumption is eventually followed by *disappointment* in the available amount of capital, as the comparison group evolves to reflect past good times. Once again the agent over-reacts, cutting consumption more forcefully than what she would under RE, causing a bust. Subsequently, the over-accumulation of capital leads to positive surprises and renewed optimism. As a result of this surprise her consumption recovers, and so on. The larger the lag J of the reference expectation, the longer *and* more severe the under- and over-accumulation of capital.

Methodological contribution and quantitative evaluation. In the second part of the paper we first explain how to solve linear general equilibrium models in the presence of DE (under naïveté) by using standard solution methods, such as Sims (2000). In a nutshell, the model is solved under the assumption that agents can observe the current state of the economy, but that when they form expectations they are subject to DE. In turn, DE are based on a linear transformation of a shadow RE law of motion. Our solution method is *portable, tractable* and, importantly, also allows for *general* forms of how memory recall loads on different past information sets.

We apply this solution method to incorporate DE into a quantitative New Keynesian

MPC out of temporary income shocks is puzzlingly large, even for financially un-constrained agents. Our mechanism differs from two recent related approaches that generate such high MPCs. Lian (2020) shows that (partial) sophistication is key for an agent to decide to save less today out of anticipation of future mistakes. In Ilut and Valchev (2020), agents are similarly naive as in our benchmark model, but have uncertainty over their optimal consumption functions, which endogenously leads to stable beliefs characterized by high MPC.

model (Christiano et al. (2005), Smets and Wouters (2007)). Given our particular interest in the role played by distant memory recall, we model the reference distribution in a flexible manner, as a weighted average of lagged RE expectations. We estimate the model using a Bayesian version of the impulse-response-function (IRF) matching method developed by Christiano et al. (2010), where the empirical IRFs are recovered using a local Jordà (2005) projection to a Romer and Romer (2004) monetary policy shock.

We find that the DE model reproduces the empirical IRFs to a monetary policy shock well, successfully generating a persistent and hump-shaped boom-bust cycle in consumption and other macro variables. In contrast, the RE model fails in delivering the empirical boom-bust and amplitude, indicating that DE are a critical force in the estimated model. As a result, the marginal likelihood, a Bayesian measure of fit that penalizes models with more parameters, heavily favors the estimated DE over the RE model. In addition, the DE model is also able to match remarkably well untargeted empirical responses, including the Survey of Professional Forecasters expectations on inflation and GDP growth.

Distant memory is crucial for the empirical success of the DE NK model. The estimated memory weights are centered on expectations formed five and six quarters ago. A re-estimated model imposing the constraint of recent memory ($J = 1$) finds no role for DE, with the estimated IRFs identical to the RE ones. This is because distant memory and the degree of diagnosticity θ are *complements* and interact to amplify the role of DE through the feedback between memory and actions. Counterfactual simulations imposing smaller lags J highlight that distant memory affects both the magnitude and the duration of boom-bust cycles. The longer the lag in memory, the more consequential the actions taken by agents in the meantime.

As in the consumption-savings and the PIH models, this result can be understood in light of surprises in the capital stock. An expansionary monetary policy shock stimulates consumption and investment so the stock of capital increases and the agent is positively surprised by the resources available. Spending further rises, which in turn leads to more capital stock and further positive capital surprise. This virtuous feedback loop continues until the reference expectation of capital begins to catch up to the realized capital. The agent is then less optimistic about the future and begins cutting back on spending. Eventually, she becomes *disappointed* in the level of capital relative to the reference distribution formed at the height of the boom, leading her to over-correct. Consumption is now reduced below the steady-state level, pushing down the level of aggregate demand and capital accumulation. A bust period arises, where the feedback between beliefs and actions leads to further economic declines and disappointment.

In contrast to the consumption-savings and the PIH models, where the interest rate was constant, in the New Keynesian model the consumption boom-bust must now be accompanied

by a corresponding movement in the interest rate. Perceived consumption growth under DE is linked to the perceived real interest rate. The perceived real rate is negatively related to the perception of changes in the price level and can be decomposed into (i) the one-step-ahead expectation of inflation under DE and (ii) the surprise in the current price level. We label this second term the *perceived innovation in cumulative inflation*, because the surprise in today's price level reflects the cumulative inflation between the current period and the time at which reference expectations were formed. On impact, because of an increase in utilization, inflation declines. This determines a negative surprise in the price level and a lower than usual expected future price level that is consistent with a perceived acceleration in consumption. Inflation eventually starts picking up, leading to a reduction in the negative surprises for the price level and then to positive surprises. This path determines a *reversal* in the perceived innovation in cumulative inflation, which moves into the positive territory during the bust part of the cycle, when agents find the resulting high perceptions of future price level consistent with their pessimism about future consumption.

Our paper is closely related to some recent contributions that study DE in macro models. Bordalo et al. (2019a) analyzes DE about a TFP process to account for credit cycles, Maxted (2020) builds a He and Krishnamurthy (2019) style macro-finance model featuring DE, while d'Arienzo (2020) introduces DE into a term-structure model to study bond market puzzles. L'Huillier et al. (2021) further shares a similar interest with us in introducing DE into linear, dynamic general equilibrium models. We contribute to the literature in two key ways. First, we address the conceptual challenges of modeling the role of endogenous states and distant memory recall in jointly affecting DE beliefs and optimal actions. In particular, compared to this existing work, we confront the problem of time inconsistency by providing a behavioral foundation of naïveté and sophistication and show that distant memory is necessary to generate repeated boom-bust cycles in response to a single initial shock.[3] Second, in quantitative terms, we propose and use an easily portable solution method to estimate a New Keynesian model with DE to show that the feedback between actions and distant memory is critical in replicating the boom-bust cycle we recover from the data.

2 DE with endogenous states and distant memory

Consider an environment where the only source of stochasticity is a (one-dimensional) Y_t process. Time is discrete and indexed by t. Let Y^t denote the history $\{Y_0, ..., Y_t\}$ of Y_t realizations up to, and including, time t. At any t, the agent observes Y^t. Under the true

[3] In this context, L'Huillier et al. (2021) study the role of endogenous states in driving DE beliefs, but their analysis and solution method applies only when memory is based on the immediate past. d'Arienzo (2020) explores the LOIE failure as a mechanism for a maturity increasing overreactions of expectations to news. Here we connect this failure to time-inconsistency and study it in models with endogenous states.

data generating process (DGP) Y_{t+1} is conditionally normal:

$$Y_{t+1} = \mu_Y(Y^t) + \varepsilon_{t+1}, \tag{1}$$

where ε_{t+1} are mean-zero iid normal shocks with variance $\sigma^2 > 0$. Here $\mu_Y(.)$ gives the time t conditional mean, as a function of current and past Y_t realizations.

Endogenous predictability. We now introduce the process through which selective memory recall distorts subjective forecasts in the presence of endogenous states. Let C_{t+1} be the random variable that the agent is interested in forecasting (e.g. future consumption or future wealth). Suppose that its underlying DGP is

$$C_{t+1} = \mu_{C,Y}(Y^t) + \mu_{C,K}(Y^t) + \delta\varepsilon_{t+1}. \tag{2}$$

In the expression above, $\mu_{C,Y}(.)$ captures the *exogenous predictability* in C_{t+1} that arises because of the exogenous state Y^{t+1}. Instead, $\mu_{C,K}(.)$ reflects the *endogenous* part of the law of motion of C_{t+1}. In an economic model, in which C_{t+1} is partly endogenously-determined, this component arises from actions that are optimally set as a response to Y^t. We refer to the determination through $\mu_{C,K}(.)$ as *endogenous predictability*. Finally, $|\delta| > 0$ reflects the exposure of C_{t+1} to ε_{t+1}, ensuring that the former is also non-predetermined as of time t. The two sources of predictability can be added and summarized as an overall predictability:

$$\mu_C(Y^t) \equiv \mu_{C,Y}(Y^t) + \mu_{C,K}(Y^t). \tag{3}$$

To illustrate how the two sources of predictability interact with each other, let C_{t+1} be future consumption determined by a budget constraint as $Y_{t+1} + K_t$, where Y_{t+1} is a stochastic labor income with a conditional mean $\mu_Y(Y^t) = \rho Y_t$ and K_t is accumulated savings. For the sake of the argument, suppose that K_t is optimally determined as αY_t. Then, the underlying, true density for C_{t+1} has $\delta = 1$ and conditional predictability

$$\mu_C(Y^t) = (\rho + \alpha)Y_t. \tag{4}$$

Diagnostic beliefs. We now discuss how selective memory recall may distort agent's subjective probability judgments over C_{t+1}. For this purpose, we build on Gennaioli and Shleifer (2010), Bordalo et al. (2018) who formulate a behavioral model of *diagnostic expectations* (DE). The fundamental psychological first-principle basis for this model is that due to limited and selective memory retrieval, subjective probability assessments are overweighted by event realizations that are "representative," in the sense of the Kahneman and Tversky (1972) representativeness heuristic of probabilistic judgments. This heuristic has been motivated and documented by a large psychology and experimental literature (e.g. Bordalo et al. (2018), Bordalo et al. (2020a), and Bordalo et al. (2021)).

The basic intuition brought forward by DE is that the judged probability of an otherwise

uncertain event partly reflects its 'true,' objective, frequency (the 'kernel of truth') as well as a subjective element that reflects the accessibility of that event in the agent's working memory. When new information arrives, memory selectively recalls more vividly past events that are more associated with, or representative of, the current news. In our context, we describe the DE model as the distorted density

$$h_t^\theta(\widehat{C}_{t+1}) = h(\widehat{C}_{t+1}|\mu_C(Y^t)) = \mu_C(\widehat{Y}^t)\left[\frac{h(\widehat{C}_{t+1}|\mu_C(Y^t)=\mu_C(\widehat{Y}^t))}{h(\widehat{C}_{t+1}|\mu_C(Y^t)=\mathbb{E}_{t-J}[\mu_C(Y^t)])}\right]^\theta \frac{1}{a} \tag{5}$$

where a is an integration constant that ensures that $h_t^\theta(\widehat{C}_{t+1})$ integrates to one. We use 'hats' when needed to emphasize the specific realization of any given random variables.

There are several important elements in this distorted distribution $h_t^\theta(\widehat{C}_{t+1})$, all of which appear in some form in the earlier formulations of DE (e.g. Bordalo et al. (2018)). First, $h(\widehat{C}_{t+1}|\mu_C(Y^t)=\mu_C(\widehat{Y}^t))$ is the true density in equation (2) of a future realization $\widehat{C}_{t+1}$ for a given current realization of the conditional mean, $\mu_C(\widehat{Y}^t)$. Second, $\mathbb{E}_{t-J}[\mu_C(Y^t)]$ is the *comparison group* for the random variable $\mu_C(Y^t)$. $\mathbb{E}_{t-J}[.]$ denotes the expectation operator for any arbitrary random variable conditional on $t-J$ information (i.e. conditional on Y^{t-J}) under the true law of motion for Y_t in equation (1). This comparison group gives the state prevailing if there is no news, compared to the immediate ($J=1$), or more distant past ($J>1$). Third, $\theta>0$ captures the severity of representativeness on judgments. In contrast, if $\theta=0$, memory recall is frictionless and the heuristic has no effects. Fourth, the distorted density in equation (5) applies if and only if the conditional variance $\delta^2\sigma^2>0$ in equation (2). When $\delta^2\sigma^2=0$, the conditional likelihood of observing any other scenario for $\widehat{C}_{t+1}$ than the one the agent is now fully informed on (given by $\mu_C(\widehat{Y}^t)$) 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.[4]

Bordalo et al. (2018) show how the normality assumption of $h(.)$ leads to a tractable characterization of $h_t^\theta(.)$. Specifically, compared to $h(.)$, under the DE density $h_t^\theta(.)$ the random variable C_{t+1} remains conditionally normally distributed with the same variance $\delta^2\sigma^2>0$, but a distorted conditional mean

$$\mathbb{E}_t^\theta(C_{t+1}) = \mu_C(\widehat{Y}^t) + \theta[\mu_C(\widehat{Y}^t) - \mathbb{E}_{t-J}[\mu_C(Y^t)]]. \tag{6}$$

The DE distortion captures the *over-reaction* of the conditional mean to the new information. This distortion is proportional by a factor of θ to the 'surprise' in the realized conditional

[4]This natural property also implies that if the agent is *only* interested in forecasting (or 'now-casting') a random variable like $\mu_C(Y^t)$, conditional on observing Y^t, then $\mathbb{E}_t^\theta[\mu_C(Y^t)] = \mathbb{E}_t[\mu_C(Y^t)] = \mu_C(\widehat{Y}^t)$. In the language developed in Bordalo et al. (2018), to compute $\mathbb{E}_t^\theta[\mu_C(Y^t)]$ the realization $\mu_C(\widehat{Y}^t)$ constitutes its infinitely representative state (see appendix in Bordalo et al. (2018) on Corollary 1).

mean. The surprise in the RE mean, in turn, depends on the new information arising between time $t-J$ and current t, i.e. the realized path for $Y_{t-J+1}, ...Y_t$, through the function μ_C.

Memory and endogenous predictability. We emphasize that the 'kernel of truth' $\mu_C(Y^t)$ that appears in the distorted density captures the overall statistical predictability. This representation thus treats *all* sources of predictability equally and jointly. In our example, it means that memory recall does not treat the component due to *exogenous predictability*, $\mu_{C,Y}(Y^t)$, as different from the component due to *endogenous predictability*, i.e. $\mu_{C,K}(Y^t)$.[5]

The conceptual emphasis on the overall statistical predictability also conforms with recent evidence provided by Bordalo et al. (2020a), Bordalo et al. (2021) showing that memory is in important ways spontaneous. In particular, when cued by the decision-problem at hand to think through hypotheses (i.e. conditional density) over the future random variable C_{t+1}, decision-makers are not triggered to think separately about its sources, i.e. whether high C_{t+1} is predicted separately by the $\mu_{C,Y}(.)$ or $\mu_{C,K}(.)$ component. Instead, (i) the new relevant data and (ii) the particular hypotheses under consideration, jointly bring to the top of the agent's mind events that associate that relevant data with the hypothesis, without necessarily disentangling the particular sources behind this event association.

Intuitively, and anticipating our consumption-savings application, consider the illustrative equation (4), with positive loadings ρ and α. Memory is overly influenced by the statistical association between periods of high income realizations and subsequent high consumption realizations (since here the sum $\rho + \alpha > 0$). Memory is not cued to distinguish whether this association arises because high income realizations are associated with high subsequent income realizations (separately due to $\rho > 0$) or with periods of high savings (separately due to $\alpha > 0$). In both cases, what memory spontaneously recalls is that high consumption realizations typically follow ('good times' are typically followed by 'good times').[6]

Joint endogenous determination. The distorting effects of DE discussed above, motivated by empirical and conceptual appeal, take as given the endogenously-determined predictability through $\mu_{C,K}(Y^t)$ in equation (2). When such predictability arises from actions that are in turn affected by DE beliefs, a first theoretical challenge arises - optimal actions and DE beliefs need to be jointly determined. For example, in a consumption-savings

[5]This formalization avoids arbitrary decompositions of how the agent recalls and thinks about the different parts that affect the conditional mean of C_{t+1}. For example, the agent could further decompose $\mu_{C,Y}(Y^t)$ and/or $\mu_{C,K}(Y^t)$ into specific sources of exogenous and/or endogenous predictability. This representation rules out such decompositions. As such, it is conceptually consistent with the approach in earlier work on DE where different sources of *exogenous predictability* are not treated differently. This interpretation is also at the heart of the formal result in Bordalo et al. (2018) showing that it is not possible to freely bring a predetermined variable outside the DE operator.

[6]This simple example also makes clear that for endogenous predictability to arise, a feedback between beliefs and actions is necessary: If $\alpha = 0$, there is no endogenous predictability. The Online Appendix presents an investment model that illustrates this point.

model, optimal savings is (i) endogenously affected by DE beliefs over future consumption (like $\mathbb{E}_t^\theta(C_{t+1})$ in equation (6)) and (ii) in turn affects the underlying statistical process for consumption, over which selective memory recall distorts those DE beliefs (like the conjectured response coefficient α in equation (4)).

To briefly preview our approach, we will interpret this joint determination as the outcome of a Nash equilibrium of an intra-personal game between two components of individual decision-making, *Memory* (forming DE beliefs, given actions) and *Deliberation* (finding optimal actions, given DE beliefs).[7] A fixed point between actions and beliefs will then ensure that the 'kernel of truth' component of DE beliefs (like $\mu_C(Y^t)$ in equation (6)) is consistent with the endogenous predictability implied by optimal actions.

Distant memory and LOIE. A second conceptual challenge in economic models driven by DE beliefs arises because the Law of Iterated Expectations (LOIE) fails when DE is based on distant memory. To see this issue in the environment summarized by equations (1) and (3), consider some arbitrary periods $t > J$, integer $n \geq 1$, and some comparison group $t - J$ in equation (5), where $J \geq 1$. As discussed in Bordalo et al. (2018), Corollary 1, we can establish (proof in Appendix) that

Lemma 1. *LOIE holds generically under DE, i.e.* $\mathbb{E}_t^\theta \left[\mathbb{E}_{t+1}^\theta[C_{t+1+n}]\right] = \mathbb{E}_t^\theta \left[C_{t+1+n}\right]$*, if and only if memory is based on the immediate past (i.e* $J = 1$*).*

The key term in $\mathbb{E}_t^\theta \left[\mathbb{E}_{t+1}^\theta[C_{t+1+n}]\right]$ that determines if LOIE holds is the perceived surprise

$$\mathbb{E}_t[C_{t+1+n}] - \mathbb{E}_t[\mathbb{E}_{t+1-J}[C_{t+1+n}]]. \tag{7}$$

Intuitively when memory is based on more distant past ($J > 1$), the time t expectation over the $t+1$ DE forecast of C_{t+1+n} introduces an *additional* lagged forecast (here $\mathbb{E}_{t+1-J}[C_{t+1+n}]$) which would not be otherwise included in the time t DE forecast of C_{t+1+n} itself. Indeed, the key term in equation (7) becomes generically zero if and only if $J = 1$.

We show below that the failure of the LOIE under distant memory typically leads to time-inconsistency in optimal actions. To confront this issue, we study decision outcomes as an intra-personal game between current and future selves (i.e naïveté vs. sophistication, as for example in O'Donoghue and Rabin (1999) and Laibson (1997)).

Plan. The rest of the paper is organized as follows. In Section 3 we use a simple consumption-smoothing problem to present our approach to the joint determination of DE beliefs and optimal actions in the presence of endogenous states, with recent and distant memory. In Section 4 we incorporate DE into a quantitative New Keynesian model of the

[7] We thank an anonymous referee for suggesting this interpretation.

type widely used for policy analysis, showcasing both the methodological tractability and quantitative success of our approach.

3 Joint determination of beliefs and actions under DE

The previous section illustrated the DE representation and introduced the conceptual challenges we are interested in addressing. In this section, we start with a three-period consumption-saving model to detail our approach. We solve the model under RE and then introduce DE. We first discuss the issue of jointly determining optimal actions and DE beliefs in the presence of *endogenous predictability*. We consider the case of recent ($J = 1$) and distant ($J > 1$) memory. We discuss two alternative ways to handle the time inconsistency that arises with distant memory, namely *naïveté* and *sophistication*, and opt for the former as a benchmark. We then analyze the economic implications of the *interaction* between endogenous predictability and distant memory, in a simple infinite horizon extension. In the last subsection, we micro-found the joint determination of beliefs and actions as the outcome of an intra-person Nash equilibrium between a *Memory* and *Deliberation* self.

Economic environment. An agent born at a generic time 1 inherits beliefs from J periods ago and capital K_0 from last period. In each period, the agent receives the exogenous income $Y_t = \overline{Y} + \varepsilon_t$ for $t = 1, 2, 3$, where ε_t are mean zero i.i.d. normal shocks with variance $\sigma^2 > 0$. At time 1, the agent chooses actual savings K_1 and a contingent plan for savings at time 2, K_2, so to maximize current utility and the expected discounted sum of future utilities. Here we also assume for simplicity a real interest rate $r = 0$, a discount factor $\beta = 1$, and a quadratic utility function $u(C) = bC - .5C^2$, where $b > 0$ and $C < b$ are such that utility is increasing in consumption in that region. The three-period assumption greatly simplifies this problem, as her optimal end-of-life $K_3 = 0$, since we rule out bequests motives.

3.1 Rational Expectations solution

We first solve the model under RE. At time 1, the agent solves

$$\max_{K_1, K_2} \{u(C_1) + \mathbb{E}_1 [u(C_2) + u(C_3)]\} \tag{8}$$
$$s.t.\ C_1 = Y_1 + K_0 - K_1;\ \ C_2 = Y_2 + K_1 - K_2;\ \ C_3 = Y_3 + K_2 - K_3.$$

and at time 2, the agent solves, given the inherited capital K_1,

$$\max_{K_2} \{u(C_2) + \mathbb{E}_2 [u(C_3)]\} \tag{9}$$
$$s.t.\ C_2 = Y_2 + K_1 - K_2;\ \ C_3 = Y_3 + K_2 - K_3.$$

Proposition 1. *The solution under RE is time consistent and is given by*

$$K_1 = \alpha_{K_0}^{RE} K_0 + \alpha_{\varepsilon_1}^{RE} \varepsilon_1, \quad K_2 = \alpha_{K_1}^{RE} K_1 + \alpha_{\varepsilon_2}^{RE} \varepsilon_2,$$

where $\alpha_{K_0}^{RE} = \frac{2}{3} = \alpha_{\varepsilon_1}^{RE}$ *and* $\alpha_{K_1}^{RE} = \frac{1}{2} = \alpha_{\varepsilon_2}^{RE}$.

First, under RE, agents spread their resources so to achieve a perfectly smooth consumption path (in expectation). Second, the marginal propensity to save is invariant to the origin (savings K_0 and K_1 or income shock ε_1 and ε_2) of available resources. Finally, the time consistency means that the planned savings policy $K_2(K_1, \varepsilon_2)$ under time 1 problem (8) coincides with the optimal policy $K_2(K_1, \varepsilon_2)$ under time 2 problem (9).

3.2 Diagnostic Expectations solution

We now introduce DE. We use θ, p-superscripts and θ-superscripts to denote planned choices and equilibrium choices, respectively, under the DE problem. The time 1 problem is to choose actual savings $K_1^{\theta,p}$ and a contingent plan for $K_2^{\theta,p}$ to maximize

$$\max_{K_1^{\theta,p}, K_2^{\theta,p}} \left\{ u(C_1^{\theta,p}) + \mathbb{E}_1^{\theta} \left[u(C_2^{\theta,p}) + u(C_3^{\theta,p}) \right] \right\} \tag{10}$$

$$s.t.\ C_1^{\theta,p} = Y_1 + K_0 - K_1^{\theta,p};\ \ C_2^{\theta,p} = Y_2 + K_1^{\theta,p} - K_2^{\theta,p};\ \ C_3^{\theta,p} = Y_3 + K_2^{\theta,p} - K_3.$$

The end-of-life savings K_3 is optimally set to zero. The first-order conditions are:

$$C_1^{\theta,p} = \mathbb{E}_1^{\theta} \left[C_2^{\theta,p} \right], \quad \mathbb{E}_1^{\theta} \left[C_2^{\theta,p} \right] = \mathbb{E}_1^{\theta} \left[C_3^{\theta,p} \right]$$

At time 2, conditional on K_1^{θ} and Y_2, the agent re-optimizes (and thus may exhibit time-inconsistency) over her initially planned $K_2^{\theta,p}$, by looking for a K_2^{θ} that solves

$$\max_{K_2^{\theta}} \left[u(C_2^{\theta}) + \mathbb{E}_2^{\theta} u(C_3^{\theta}) \right] \tag{11}$$

$$s.t.\ C_2^{\theta} = Y_2 + K_1^{\theta} - K_2^{\theta};\ \ C_3^{\theta} = Y_3 + K_2^{\theta} - K_3.$$

The first-order condition at time 2 is:

$$C_2^{\theta} = \mathbb{E}_2^{\theta} \left[C_3^{\theta} \right] \tag{12}$$

Before solving the model and discussing the differences between recent and distant memory, we find it useful to explain how to interpret the optimization problem under imperfect memory.

An intra-personal game interpretation. Our proposed approach to micro-found the joint determination of actions and DE beliefs in the presence of endogenous predictability takes the form of a Nash equilibrium of an intra-personal game between *Memory* (self M) and *Deliberation* (self D) part of a decision-maker.

In this formulation, the Memory self captures the psychology of selective memory recall, and builds on properties discussed in Section 2. In particular, conditional on the new, representative information cued by the observed exogenous variation (here the current income shock), Memory constructs DE beliefs over the future variable of interest for current decision-making (here of future consumption). In this construction, Memory takes as given

a relationship that associates exogenous variation to the overall (i.e., both exogenous and endogenous) predictability in that variable of interest. As introduced in Section 2, our approach thus continues to be motivated by the spontaneous property of memory documented in works like Bordalo et al. (2020a) and Bordalo et al. (2021). In turn, Deliberation finds optimal current actions (e.g. current savings), as a function of economic states (income shock and inherited capital) and of the conditional DE forecasts provided by Memory. In equilibrium, we then impose that the endogenous part of the associative relationship used by Memory is consistent with the optimal response function of Deliberation.

We defer derivations to Section 3.4. For the purposes of describing the emerging solution, we note for now that our approach allows us to formally represent an optimality condition such as in equation (12) as characterizing the K_2^θ that solves

$$\underbrace{Y_2 + K_1^\theta - K_2^\theta}_{=C_2^\theta} = \mathbb{E}_2 \underbrace{\left[Y_3 + K_2^\theta\right]}_{=\mathbb{E}_2[C_3^\theta]} + \theta \underbrace{\left[\mathbb{E}_2\left[Y_3 + K_2^\theta\right] - \mathbb{E}_{2-J}\left[Y_3 + K_2^\theta\right]\right]}_{=\text{ Perceived surprise in conditional mean}}. \tag{13}$$

Intuitively, self D optimally chooses K_2^θ to implement perceived consumption smoothing, where the RHS of the equation gives the self M's DE belief over C_3^θ and connects to the discussion around the DE representation in equation (6). In the presence of endogenous predictability the 'kernel of truth' component (i.e. $\mathbb{E}_2[Y_3 + K_2^\theta]$) is an equilibrium outcome, which by itself is consistent with the agent's (i.e. self D) own optimal action for K_2^θ.[8]

With this interpretation of the joint determination of actions and beliefs at hand, we now proceed to characterize their properties by distinguishing between recent and distant memory.

3.2.1 Recent Memory ($J = 1$)

We prove two properties when memory recall is based on the immediate past, i.e. $J = 1$.

Proposition 2. *When $J = 1$, the conditional time-2 optimal solution $K_2^\theta(K_1^\theta, \varepsilon_2)$ is identical ('time-consistent') to the time-1 optimal contingent plan $K_2^{\theta,p}(K_1^\theta, \varepsilon_2)$.*

To see this, note that while the optimal time-1 plan $K_2^{\theta,p}$ in equation (10) is set such that $\mathbb{E}_1^\theta\left[C_2^{\theta,p} - C_3^{\theta,p}\right] = 0$, the conditional optimal K_2^θ solves the time-2 perceived tradeoff $C_2^\theta - \mathbb{E}_2^\theta[C_3^\theta] = 0$. Since the LOIE holds when $J = 1$ (Lemma 1), the conditional optimal K_2^θ implements exactly the time-1 desired consumption path under $K_2^{\theta,p}$:

$$\mathbb{E}_1^\theta\left[C_2^\theta - C_3^\theta\right] = \mathbb{E}_1^\theta\left[C_2^\theta - \mathbb{E}_2^\theta C_3^\theta\right] = 0.$$

Guided by the RE solution, we conjecture an optimal policy of a similar form:

$$K_1^\theta = \alpha_{K_0}^\theta K_0 + \alpha_{\varepsilon_1}^\theta \varepsilon_1;\ K_2^\theta = \alpha_{K_1}^\theta K_1^\theta + \alpha_{\varepsilon_2}^\theta \varepsilon_2.$$

[8] The iid assumption on Y_t also simplifies here the environment so that there is no exogenous predictability.

Proposition 3. *When $J = 1$, compared to the RE policy functions, the optimal policy functions K_1^θ and K_2^θ feature the same optimal response to the endogenous state but a muted response to the current income innovation, i.e.*

$$\alpha_{K_0}^\theta = \frac{2}{3} = \alpha_{K_0}^{RE}; \quad \alpha_{K_1}^\theta = \frac{1}{2} = \alpha_{K_1}^{RE}; \quad \alpha_{\varepsilon_1}^\theta = \frac{2}{3+\theta} < \alpha_{\varepsilon_1}^{RE}; \quad \alpha_{\varepsilon_2}^\theta = \frac{1}{2+\theta} < \alpha_{\varepsilon_2}^{RE}.$$

The intuition for the muted savings response to the income shock is at the heart of the *endogenous predictability* mechanism that we emphasize. To see this, recall the $t = 2$ optimality in equation (13), which for $J = 1$ reads

$$\overline{Y} + \varepsilon_2 + K_1^\theta - K_2^\theta = \mathbb{E}_2\left[\overline{Y} + \varepsilon_3 + K_2^\theta\right] + \theta\left[\mathbb{E}_2\left[\overline{Y} + \varepsilon_3 + K_2^\theta\right] - \mathbb{E}_1\left[\overline{Y} + \varepsilon_3 + K_2^\theta\right]\right]$$

Given a current unusually high income shock ε_2, and thus level of assets K_2^θ, the agent correctly recognizes that her total future resources and consumption are likely to be higher than usual. Since income Y_3 is iid, this conditional predictability of future resources comes just from K_2^θ, which through the response $\alpha_{K_1}^\theta$ induces the *endogenous* persistence from ε_2 to C_3^θ. As described in the motivating Section 2, an agent subject to the representativeness heuristic is then *overly influenced* by her perception of the new information contained in this unusual state of high total expected resources $\overline{Y} + K_2^\theta$. Due to her imperfect memory, she recalls more vividly state realizations that are representative in light of this new information and inflates the likelihood of future resources $Y_3 + K_2^\theta$ to be high.

In other words, selective memory recall associates the current high income innovation with past times of *optimism* about total future resources. Such association does not occur here for the standard reason of exogenous income being persistent. Instead, it arises because past savings have smoothed past income innovations through time, making them endogenously persistent. Thus, due to these selective memory associations, a higher current income innovation leads the agent to currently act under an "as if" overly optimistic view of future resources. The kernel of truth is satisfied, given that it is in fact the case that current good times predict future good times, but the agent's memory overstates the strength of this relation and is not cued to discern that the relation arises because of her behavior. Given this memory-based view, the agent optimally consumes more and saves less today than the RE agent, resulting in a *high* marginal propensity to consume (MPC) and a lack of consumption smoothing as measured by an external observer.[9]

Let us turn now to the DE response to the endogenous state K_1^θ. The key economic observation here is that when $J = 1$, the economic state K_1^θ also serves as the *necessary*

[9]Our simple economic model does not feature financial constraints, usually viewed as the standard economic reason for high MPC. Thus, our model's implication speaks closer to the challenge posed to standard models by the empirical evidence on high MPC out of temporary income shocks of financially un-constrained agents (see for example evidence in Kueng (2018), Fagereng et al. (2020) and McDowall (2020)).

and sufficient conditioning information to form the comparison group $\mathbb{E}_{2-J}\left[\overline{Y}+\varepsilon_3+K_2^\theta\right]$. Therefore, the DE beliefs' over-reaction to the new information, $K_2^\theta-\mathbb{E}_{2-J}[K_2^\theta]$, only contains the current innovation ε_2 and not the endogenous state K_1^θ, that was already known when the reference expectations were formed. Thus, when $J=1$, DE affect the reaction to ε_2, but not to K_1^θ - indeed, the latter response remains identical to the RE one.

3.2.2 Distant Memory ($J>1$)

Now consider the case when memory recall is based on the more distant past, i.e. $J>1$.

Proposition 4. *When $J>1$, the conditional time-2 optimal solution for K_2^θ is not equal to the time-1 optimal contingent plan $K_2^{\theta,p}(K_1^\theta,\varepsilon_2)$ (i.e. it is not 'time-consistent').*

This result is a direct manifestation of the LOIE failure under distant memory (Lemma 1). As we show in the Online Appendix, the time inconsistency arises because of the information content of $K_1^{\theta,p}$ with respect to the capital expected at time 0 (assuming $J=2$). Between when reference expectations were formed, at time 0, and when a new decision is made, at time 2, an income shock occurred and agents reacted to the shock. As a result, capital is not what the agent expected it to be. Agents do not take into account this surprise in capital when they solve the planning problem at time 1.

Faced with this inherent time-inconsistency, we then need to model the agent's current beliefs about her future actions. Here we use insights from existing literature on time-inconsistency (e.g. the seminal work by Strotz (1955) and Pollak (1968)) that point to two different frameworks. The first approach, coined in this literature as *naïveté* in the O'Donoghue and Rabin (1999) sense, used for example in Akerlof (1991), models an agent who does not forecast her future self's behavior to be governed by the representativeness heuristic. The second approach is *sophistication* (e.g. Laibson (1997)), where the agent understands that her future action is dictated by the representativeness heuristic.

Naïveté problem. Under *naïveté* , the time 1 problem is

$$\max_{K_1^\theta}\left\{u(C_1^\theta)+\mathbb{E}_1^\theta\left[u(C_2^{RE})+u(C_3^{RE})\right]\right\} \tag{14}$$

where the agent at time 1 believes her time 2 future self will take the action K_2^{RE} so to

$$\max_{K_2^{RE}}\left[u(C_2^{RE})+\mathbb{E}_2[u(C_3^{RE})]\right]. \tag{15}$$

The RE-superscript on a time t variable signify choices that are made under an RE policy function, taking as given the state variable entering that period. From the budget constraints, the (forecasted) consumption choices are therefore

$$C_1^\theta=Y_1+K_0-K_1^\theta;\ C_2^{RE}=Y_2+K_1^\theta-K_2^{RE};\ C_3^{RE}=Y_3+K_2^{RE}-K_3^{RE},$$

where K_1^θ (and K_2^{RE}) signify the choice resulting from a DE under naïveté (and RE, respectively) policy function that solve (14) (and (15), respectively) and trivially $K_3^{RE} = 0$.

The optimal solution for K_1^θ in equation (14) solves the intertemporal tradeoff

$$C_1^\theta = \mathbb{E}_1^\theta \left[C_2^{RE} + \frac{\partial K_2^{RE}}{\partial K_1^\theta} \left(C_3^{RE} - C_2^{RE} \right) \right], \tag{16}$$

This captures the direct effects of the current choice on tomorrow's consumption and the indirect effects through the capital choice at time 2, anticipated to follow K_2^{RE} (where $\frac{\partial K_2^{RE}}{\partial K_1^\theta} = 0.5$ by Proposition 1) and the resulting consumption path.

Lemma 2. *Under naïveté, for any $J \geq 1$, the tradeoff for the optimal K_1^θ in equation (16) reduces to*

$$C_1^\theta = \mathbb{E}_1^\theta [C_2^{RE}]. \tag{17}$$

The key for this result is that the future self is expected to optimally select K_2^{RE}, which conditional on time 2 states achieves $\mathbb{E}_2[C_3^{RE}] - C_2^{RE} = 0$. Thus, $(C_3^{RE} - C_2^{RE})$ equals just the income innovation ε_3, unpredictable under $\mathbb{E}_1^\theta$ (for any $J \geq 1$). Due to this induced unpredictability, for the naive agent the indirect effect of K_1^θ as a relevant state for future conditionally optimal choices can be ignored - a form of envelope-theorem result that makes the problem particularly tractable.

While these are her beliefs at time 1, entering period 2 with the state realization K_1^θ and new information determined at time 2, her problem is once again influenced by the representativeness heuristic. Her conditionally optimal action solves

$$\max_{K_2^\theta} \left[u(C_2^\theta) + \mathbb{E}_2^\theta [u(C_3^{RE})] \right], \tag{18}$$

where $C_2^\theta = Y_2 + K_1^\theta - K_2^\theta$ and $C_3^{RE} = Y_3 + K_2^\theta - K_3^{RE}$. The optimal action implements

$$C_2^\theta = \mathbb{E}_2^\theta [C_3^{RE}]. \tag{19}$$

The behavioral interpretation of equations (14), (15) and (18) is that, at time 1, the agent maximizes assuming that after time 2 the future selves will not be subject to any memory heuristics (i.e. she will act 'fully rationally'), even though at time 2 the decision maker ends up changing behavior and is in fact subject to her imperfect memory recall.

Joint naive beliefs and actions. To characterize the resulting optimal actions and DE beliefs, we can extend the terminology introduced earlier in section 3.2 around Memory and Deliberation selves to account for the naïveté assumption. In particular, similar to the argument around equation (13), we look to impose equilibrium consistency of beliefs for the Memory self of the naive agent.

There are two properties of equilibrium consistency that arise here. First, the 'kernel of

truth' component part is

$$\mathbb{E}_2[C_3^{RE}] = \overline{Y} + K_2^{\theta,*} - \mathbb{E}_2[K_3^{RE}],$$

where $K_2^{\theta,*}$ is the optimal choice by the Deliberation self (which Memory self takes as given) and we make specific the role of perceived future optimal behavior under RE through $\mathbb{E}_2[K_3^{RE}]$. Second, the comparison group for the time-2 naive Memory self is the RE forecast made by the former (J periods ago) naive self about time-3 perceived optimal behavior. Since naive agents at any given time believe future optimal behavior to be governed by the RE policy, the comparison groups are also built under forecasts of future optimal RE behavior. Therefore, this internal consistency requires that the comparison group is built as

$$\mathbb{E}_{2-J}[\mathbb{E}_2[C_3^{RE}]] = \mathbb{E}_{2-J}[\overline{Y} + K_2^{RE} - \mathbb{E}_2[K_3^{RE}]].$$

Put together, the naive Deliberation self solves the trade-off in (19), where the naive Memory self's beliefs are (as a counterpart of the RHS of equation (13)):

$$\mathbb{E}_2^{\theta}[C_3^{RE}] = \mathbb{E}_2[C_3^{RE}] + \theta \left(\mathbb{E}_2[C_3^{RE}] - \mathbb{E}_{2-J}[C_3^{RE}] \right), \tag{20}$$

where we used the LOIE under RE to simplify the comparison group expression above.

Solution under naïveté. We focus on $J = 2$ in this three-period model. The optimality conditions (17), (19), and the RE policies of Proposition 1 produce the following solution.

Proposition 5. *When $J = 2$, the optimal time 1 and 2 policy functions under naïveté are*

$$K_1^{\theta} = -\frac{2\theta}{3(2+\theta)} \mathbb{N}_{-1,0}[K_0] + \frac{2}{3} K_0 + \frac{2}{3+\theta} \varepsilon_1,$$
$$K_2^{\theta} = -\frac{\theta}{2(2+\theta)} \mathbb{N}_{0,1}[K_1^{\theta}] + \frac{1}{2} K_1^{\theta} + \frac{1}{2+\theta} \varepsilon_2,$$

where $\mathbb{N}_{-1,0}[K_0] \equiv K_0 - \mathbb{E}_{-1}[K_0]$ and $\mathbb{N}_{0,1}\left[K_1^{\theta}\right] \equiv K_1^{\theta} - \mathbb{E}_0\left[K_1^{RE}\right]$ represent the surprises in the stock of capital with respect to the expectations formed in the past.

The difference with the $J = 1$ case of Proposition 3 is the presence of a novel informational state, given by the surprise in capital.[10] These novel state dynamics arise because the economic states K_0 and K_1^{θ} are *not* sufficient statistics anymore for the comparison group that matters for decisions at time 1 and 2, respectively. With $J = 2$, the relevant comparison groups are built on conditional expectations $\mathbb{E}_{-1}[K_0]$ and $\mathbb{E}_0[K_1^{RE}]$, respectively.

In particular, in Proposition 5 the elasticities on K_0 and K_1^{θ} continue to recover the role of capital stock as an economic state, which influences decisions as in the RE policy function. The novel informational role is captured by the elasticities on $\mathbb{N}_{-1,0}[K_0]$ and $\mathbb{N}_{0,1}[K_1^{\theta}]$. Consider, for example, an increase in K_1^{θ} caused by a positive innovation in ε_1. A higher K_1^{θ}

[10] When $J = 1$, because the LOIE holds, the time-consistent policy functions of Proposition 3 are equivalent to those derived under naïveté (or sophistication). See Proposition 7 in the Online Appendix.

than expected at time 0 under the relevant comparison group leads to a perceived positive innovation in $K_2^\theta - \mathbb{E}_0[K_2^{RE}]$. Since agents are over-influenced by this surprise, they become over-optimistic about future resources and invest less. This explains why the innovation $\mathbb{N}_{0,1}\left[K_1^\theta\right]$ enters with a negative sign in the time 2 policy function. Similar intuition explains why the innovation $\mathbb{N}_{-1,0}\left[K_0\right]$ enters with a negative sign in the time 1 policy function.

Under recent memory ($J = 1$), these news terms collapse to zero: $\mathbb{N}_{0,0}\left[K_0\right] \equiv K_0 - \mathbb{E}_0\left[K_0\right] = 0$ and $\mathbb{N}_{1,1}\left[K_1^\theta\right] \equiv K_1^\theta - \mathbb{E}_1\left[K_1^{RE}\right] = 0$. In the second relation, we have used $\mathbb{E}_1\left[K_1^{RE}\right] = K_1^\theta$ because a projection of current capital at horizon zero is always equal to the current capital, no matter the data generating process that Memory uses when forming the projection. Intuitively, when $J = 1$, the news terms disappear because the stock of capital inherited from the past is already part of the information set entering the comparison group. Instead, when $J > 1$, the agent makes decisions in the meantime, and these decisions create surprises with respect to the comparison group based on the more distant past.

Naïveté vs sophistication. We conclude this subsection by briefly considering the alternative assumption of sophistication. We provide details in the Online Appendix. For the time 2 policy function we recover the same coefficients as the naïveté case, except that $\mathbb{E}_0[K_1^\theta]$ enters into the savings rule instead of $\mathbb{E}_0[K_1^{RE}]$. This occurs because we assume that, in order to maintain belief consistency across selves, the sophisticated agents' comparison group is the expectation formed J periods ago by the former sophisticated self. In turn, at time 1 the agent would choose a different plan than what she anticipates will be her optimal time-2 *conditional* action. Thus, her optimal time 1 action aims to fix that misalignment. The sophistication counterpart of the optimality condition in equation (16) contains indirect effects of the current action on the future policy. Since the agent anticipates that she will over-consume at time 2, the consumption-smoothing motive between time 1 and 2 leads the agent to consume more at time 1 out of temporary income shock ε_1 relative to naïveté.

In extending the theoretical framework of this consumption-smoothing model to more realistic and quantitatively relevant business cycle models, we propose to focus on the *naïveté* approach. The key reason is is that the sophistication approach's required hyper-rationality runs counter to the motivation of modeling agents' beliefs about their future circumstances as influenced by a heuristic. Indeed, the latter is usually viewed as a cognitive, mental shortcut that allows agents to make judgments quickly and efficiently (Tversky and Kahneman (1975) and Kahneman (2011)).[11] The naïveté approach is arguably psychologically more coherent and consistent with the underlying foundation of diagnostic beliefs as a heuristic reflecting a

[11]Part of this hyper-rationality is that in infinite-horizon models the current sophisticated agent would internalize the life-time indirect effect of current savings as a future information state, i.e. how current savings affect the formation of comparison groups that will matter in the future selective memory recall of the past.

memory representation affected by imprecise, selective, and less than fully rational recall.[12]

Moreover, computationally the naïveté approach is also significantly more tractable, a property that we explain and leverage in the rest of the paper. Therefore, we present the naïveté approach as a 'portable extension of existing models' (as advocated by Rabin (2013)) that tractably incorporates the psychology foundation of the representativeness heuristic and the role of imperfect memory recall in standard business cycle models.

3.3 Distant memory and boom-bust cycles

Our analysis shows that the interaction between endogenous economic states and distant memory introduces novel informational states that affects the model's propagation mechanism. In this subsection we use the simplest infinite-horizon extension of the three-period model to showcase this altered propagation.

In particular, we study the permanent income hypothesis (PIH) model under DE with naïveté. We continue to assume quadratic utility and iid income shocks. Households can save buying capital at the price $q = (1+r)^{-1}$, where $r > 0$ is the exogenous real interest rate and the discount factor is $\beta = (1+r)^{-1}$. The time t budget constraint is then:

$$K_t = (1+r)(K_{t-1} + \overline{Y} + \varepsilon_t - C_t).$$

As before, we first solve the model under RE, where the FOC is:

$$C_t^{RE} = \mathbb{E}_t(C_{t+1}^{RE}).$$

In the Appendix, we conjecture and verify the RE consumption policy function:

$$C_t^{RE} = \frac{r}{1+r}\left(K_{t-1}^{RE} + \varepsilon_t\right) + \overline{Y} \tag{21}$$

and the resulting RE saving decision

$$K_t^{RE} = K_{t-1}^{RE} + \varepsilon_t \tag{22}$$

Thus, under RE, capital is a random walk and shocks have a permanent effect on savings.

Similar to Subsection 3.2.2 under DE and naïveté the agent's problem at time t is:

$$\max_{K_t^\theta} \left\{ u(C_t^\theta) + \mathbb{E}_t^\theta \left[\mathcal{V}(K_t^\theta) \right] \right\}, \tag{23}$$

where the continuation utility $\mathcal{V}(.)$ reflects the maintained assumption that the agent at time t believes her future selves from time $t+1$ on will act under RE. Thus the agent expects that the time $t+1$ self will take the action K_{t+1}^{RE} so as

$$\mathcal{V}(K_t^\theta) = \max_{K_{t+1}^{RE}} \left[u(C_{t+1}^{RE}) + \mathbb{E}_{t+1} \left[\mathcal{V}(K_{t+1}^{RE}) \right] \right]. \tag{24}$$

[12] In terms of literature, the naïveté approach is also consistent with how previous contributions such as Bordalo et al. (2019a) and Maxted (2020) dealt with the failure of LOIE in exogenous processes.

Proposition 6. *The optimal savings policy function under DE and naïveté is*

$$K_t^\theta = K_{t-1}^\theta - \frac{r\theta}{1+r\,(1+\theta)}\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right) + \frac{1+r}{1+r\,(1+\theta)}\varepsilon_t.$$

where the news term $\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right) \equiv K_{t-1}^\theta - \mathbb{E}_{t-J}[K_{t-1}^{RE}]$.

In the Appendix we show that this solution arises from the FOC of the problem in (23)

$$C_t^\theta = \mathbb{E}_t^\theta(C_{t+1}^{RE}),$$

where the agent anticipates future behavior according to the $t+1$ RE policy (see equation (21)), similar to the result detailed in Lemma 2. As in the three period model (see Proposition 5), the DE solution presents an additional informational state variable $\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right)$. This state variable is relevant to the extent that it induces optimism or pessimism about the future as a result of a discrepancy between the resources currently available and those anticipated based on the agent's imperfect memory. Furthermore, this additional state variable is activated only to the extent that memory is based on the distant ($J > 1$) past.

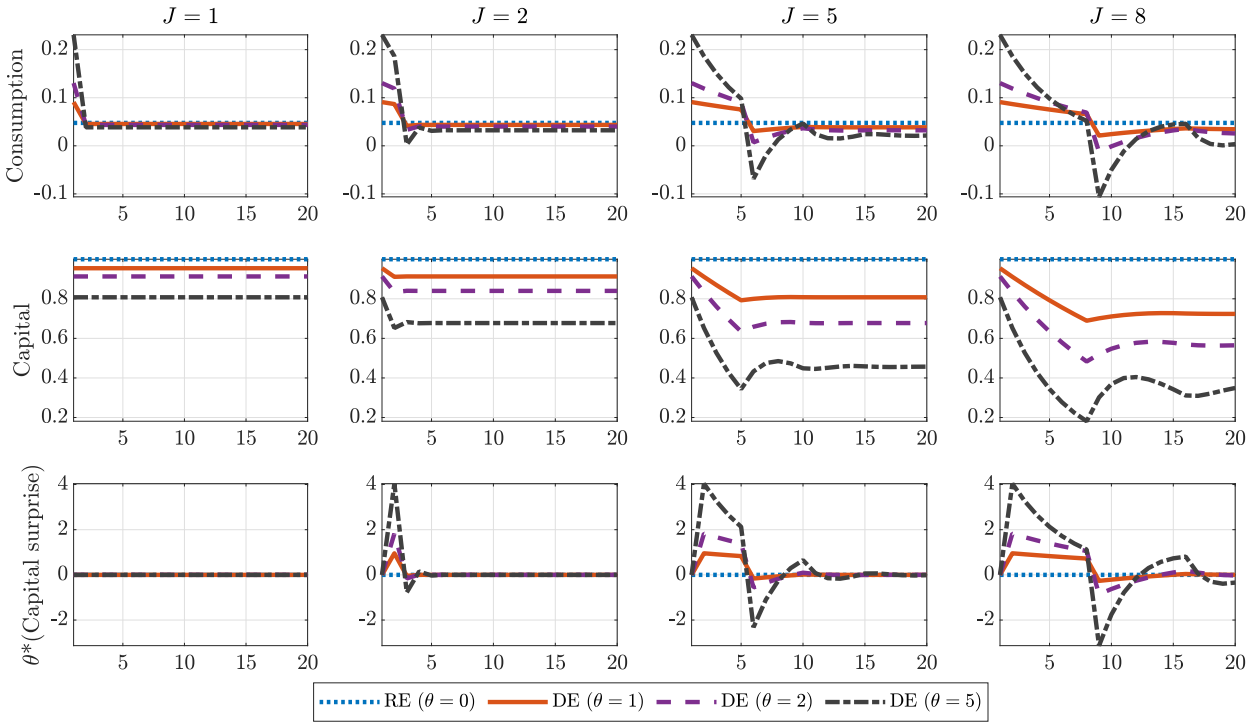
Consistent with the discussion in Subsection 3.2.2 (in particular around equation (20)), the comparison group is based on the RE solution. This is how the agent perceives capital should have evolved based on the information available at time $t-J$. Thus, using the law of motion for capital under RE in equation (22) we have $\mathbb{E}_{t-J}[K_{t-1}^{RE}] = K_{t-J}^\theta$, where K_{t-J}^θ is the capital in place at $t-J$. The optimal policy in Proposition 6 becomes:

$$K_t^\theta = K_{t-1}^\theta - \frac{r\theta}{1+r\,(1+\theta)}\left[K_{t-1}^\theta - K_{t-J}^\theta\right] + \frac{1+r}{1+r\,(1+\theta)}\varepsilon_t. \tag{25}$$

The solution above elucidates that *both* the severity of the DE distortion captured by $\theta > 0$ and the lag in the reference distribution J matter to determine the behavior of consumption and capital in response to a transitory shock. Figure 1 illustrates how these two margins interact with each other in a model in which DE apply to both exogenous and endogenous variables. Specifically, we consider a unitary income shock that dissipates after one period. The three rows report the consumption response, the capital response, and the response of the surprise in capital scaled by the DE parameter θ (i.e., $\theta\left[K_{t-1}^\theta - K_{t-J}^\theta\right]$), respectively. Across columns, we vary the lag of the reference distribution, J, while the different lines in each panel are obtained by varying the severity of the DE distortion, θ. For each variable, the panels are on the same scale to facilitate the comparison.

Endogenous booms and busts. Under DE, the initial capital response to a transitory shock ε_t is smaller than under RE (evident from equation (25)). The initial capital response decreases with θ, but it does not vary with J. This is because the shock is always unpredictable, no matter when the reference expectations were formed. In other words, the information content of the initial shock does not vary with respect to J. The under-accumulation of

Figure 1: Permanent Income Hypothesis model



Notes: The figure reports the impulse responses to a unitary iid shock for the Permanent Income Hypothesis model. The three rows report the responses of consumption, capital, and the surprise in capital scaled by the DE parameter θ (i.e., $\theta\left[K^{\theta}_{t-1} - K^{\theta}_{t-J}\right]$), respectively. Across columns, we vary the lag of the reference distribution, J, while the different lines in each panel are obtained by varying the severity of the DE distortion, θ. For each variable, the panels are on the same scale to facilitate the comparison.

capital translates in an over-reaction in consumption. Under recent memory ($J = 1$, first column), the path reverts to the one followed under RE after one period. Analytically, with $J = 1$ the news state in equation (25) is irrelevant. This is consistent with the fact that when $J = 1$ there are no further perceived surprises, as shown in the third row. Of course, both consumption and capital are permanently lower than under RE because of the initial under-accumulation of capital, but the impulse response is now flat, as under RE.

Moving from left to right, we can appreciate the importance of allowing for distant memory. While the initial response is not affected, the persistence of the initial over-reaction is controlled by J. More importantly, after $J + 1$ periods the agent is surprised again. However, this time the surprise is the result of the interaction between Deliberation and Memory. Consider the case $J = 2$ (second column). In period 3, the memory formed at time 1 becomes relevant. Now the agent is disappointed in the current level of capital. This is because Memory recalls the projection based on the stock of capital that was available at time 1, the reference value for period 3. Capital is lower than expected because in the

meantime the DE agent has consumed too much. The response of the DE agent at time 3 is to cut consumption more than what an RE agent would do if confronted with the same level of capital. As J increases (third and fourth column), not only it takes longer for the agent to reverse her behavior, but the correction also increases in magnitude. Thus, J does not only affect the lag in the correction, but also its *amplitude*.

As J increases, an additional, crucial feature of distant memory becomes more visible: A single, initial shock can endogenously induce repeated boom-bust cycles. As J increases, kinks and inflection points occur with lags and magnitudes that depend on J. Consider the case of $J = 8$ (fourth column). As before, a first kink occurs in period $J + 1 = 9$, when the agent reacts to a disappointing level of savings by cutting consumption more than what she would have done under RE. As the agent keeps accumulating capital and the reference level of capital progressively declines, consumption recovers. The third row highlights that the agent eventually becomes *positively* surprised by the amount of capital that she has at her disposal. This is the result of her own actions in response to the perceived low stock of capital. She is accumulating capital, while the reference value is constantly declining. As a result, capital and consumption start slowing down, generating an inflection point. Eventually, a second kink occurs (in period $2J + 1 = 17$). Now the reference level of capital is increasing once again, as a result of the agent past behavior. Thus, the agent is disappointed by her current level of capital compared to what she was expecting based on her past behavior.

Overall, distant memory creates rich interactions between the actions taken by the agent and her memory. The extent of the DE distortion captured by the parameter θ and the lag of the reference distribution as captured by J interact to create repeated boom-bust cycles. A large DE distortion implies that agents react forcefully to perceived surprises in the amount of resources. This behavior creates the conditions for future surprises. Kinks in the response of the economy occur every J periods, but inflection points can occur in between these kinks as past actions induce changes in beliefs. Finally, if memory were based on an average of multiple lags of past expectations, as in our quantitative model of Section 4, instead of on a single lag J, the kinks would appear smoother and more similar to turning points.

3.4 Memory and deliberation equilibrium

Before moving to the methodology and quantitative model of Section 4, we re-connect to Subsection 3.2, where we have introduced the interpretation of the agent solving the DE problem in the presence of endogenous predictability as a form of a Nash equilibrium between a Memory self and a Deliberation self. We now detail this interpretation in the context of the second period problem (11) and the resulting condition (13).

Consider first self M. Given the new information Y_2 and the inherited capital K_1^θ, self M

is characterized by the following conditional normal density over C_3^θ

$$C_3^\theta \sim N(\mu_{C_3}^\theta(Y_2, K_1^\theta), \sigma_{C_3}^2), \tag{26}$$

where $\mu_{C_3}^\theta$ and $\sigma_{C_3}^2$ are the perceived conditional mean and variance, respectively. This conditional density captures the spontaneous self M's reaction to the new information, when cued about the hypothesis involved in forecasting C_3^θ. Part of our equilibrium consistency requirement is that this reaction function will conform to the 'kernel of truth' component of DE formation. As such, the arguments for $\mu_{C_3}^\theta(.)$ already anticipate that consistency (we keep the state space simpler here by assuming $J = 1$), but at this stage this memory reaction is taken as given.

Deliberation, given memory formation. We now formalize how self D accounts for the effect of her choice K_2^θ on expected C_3^θ, when we insist that perceptions over C_3^θ are controlled by self M and thus given by the process in (26). For this purpose, suppose first that self D's action is given by a particular choice $\widetilde{K}_2^\theta$. Taking as given the primitive memory process in equation (26) and for any given $\widetilde{K}_2^\theta$, the perception of self M's can be recasted as

$$C_3^\theta \sim N(m_{C_3}^\theta + \widetilde{K}_2^\theta, \sigma_{C_3}^2). \tag{27}$$

To connect with the concept of 'residual uncertainty' introduced in Section 2, we can view $m_{C_3}^\theta$ as the memory distortion over the residual, conditionally stochastic part of C_3^θ.[13] This distortion can thus be further interpreted as self M's response, or adjustment, in forecasting C_3^θ, as a function of its primitive belief $\mu_{C_3}^\theta(.)$ and the number $\widetilde{K}_2^\theta$. The function is

$$m_{C_3}^\theta(\mu_{C_3}^\theta(Y_2, K_1^\theta), \widetilde{K}_2^\theta) = \mu_{C_3}^\theta(Y_2, K_1^\theta) - \widetilde{K}_2^\theta. \tag{28}$$

The basic intuition here is that self D views K_2^θ as the component of C_3^θ that she can affect, by reasoning that one extra unit of savings today must mean one more unit of available C_3^θ tomorrow. Instead, she takes as given the $m_{C_3}^\theta$ part of expected C_3^θ as a memory distortion that self D cannot manipulate.

Turning to the self D's time-2 problem in equation (11), she now takes as given a value for $m_{C_3}^\theta$ in self M's density in (27), to optimize

$$\max_{K_2^\theta} \left[u(Y_2 + K_1^\theta - K_2^\theta) + \int u[\sigma_{C_3}\xi + m_{C_3}^\theta + K_2^\theta]\phi(\xi)\, d\xi \right],$$

where $\phi(\xi)$ denotes the standard normal density. The first-order condition for the optimal

[13] By the time $t = 3$ budget constraint the residual uncertainty here is formally over Y_3. Thus, any shock realization Y_3 that achieves any given realization $\widehat{C}_3^\theta$ is now perceived as $\widehat{Y}_3 = \widehat{C}_3^\theta - \widetilde{K}_2^\theta$. Equation (27) captures an as if memory's perceived distribution for $Y_3 \sim N(m_{C_3}^\theta, \sigma_{C_3}^2)$. This formulation is not saying that memory suddenly thinks in isolation over each individual component of C_3^θ. For example, if $C_3^\theta = \widetilde{K}_2^\theta + \sum_{i=1}^N Y_{3,i}$, i.e. if there would be multiple sources of stochasticity indexed by i, then the sum $\sum_{i=1}^N Y_{3,i} \sim N(m_{C_3}^\theta, \sigma_{C_3}^2)$.

choice of K_2^θ is then given by

$$Y_2 + K_1^\theta - K_2^\theta = \int [\sigma_{C_3}\xi + m_{C_3}^\theta + K_2^\theta]\phi(\xi)\, d\xi$$

Integrating with respect to ξ, and further using equation (28), this condition gives self D's optimal response for K_2^θ as

$$K_2^\theta(Y_2, K_1^\theta, \mu_{C_3}^\theta(Y_2, K_1^\theta)) = Y_2 + K_1^\theta - \mu_{C_3}^\theta(Y_2, K_1^\theta), \tag{29}$$

a function of the economic states and the given memory mapping $\mu_{C_3}^\theta(.)$. Economically, the resulting optimal choice for K_2^θ implements the time-2 perceived consumption smoothing

$$C_2^\theta = Y_2 + K_1^\theta - K_2^\theta(Y_2, K_1^\theta, \mu_{C_3}^\theta(Y_2, K_1^\theta)) = \mu_{C_3}^\theta(Y_2, K_1^\theta). \tag{30}$$

Consistency of beliefs. Turning to the memory mapping $\mu_{C_3}^\theta(.)$, DE put discipline on its formation by imposing the 'kernel of truth' *consistency* of beliefs, as follows:

Definition 1. *The DE equilibrium self M's density (26) over C_3^θ is*

$$\mu_{C_3}^{\theta,*}(Y_2, K_1^\theta) = \mu_{C_3}(Y_2, K_1^\theta) + \theta\left(\mu_{C_3}(Y_2, K_1^\theta) - \mathbb{E}_1[\mu_{C_3}(Y_2, K_1^\theta)]\right) \tag{31}$$

where consistency of beliefs in the 'kernel of truth' component requires that

$$\mu_{C_3}(Y_2, K_1^\theta) = \mathbb{E}_2[C_3^\theta] = \overline{Y} + K_2^{\theta,*}(Y_2, K_1^\theta), \tag{32}$$

and the perceived conditional uncertainty $\sigma_{C_3}^2$ over C_3^θ to equal $\mathbb{V}_2[C_3] = \sigma^2 > 0$.

In turn, by equation (29) the equilibrium action under DE for self D satisfies

$$K_2^{\theta,*}(Y_2, K_1^\theta) = Y_2 + K_1^\theta - \mu_{C_3}^{\theta,*}(Y_2, K_1^\theta). \tag{33}$$

Put together, the resulting perceived consumption smoothing

$$C_2^{\theta,*} = Y_2 + K_1^\theta - K_2^{\theta,*}(Y_2, K_1^\theta) = \mu_{C_3}^{\theta,*}(Y_2, K_1^\theta). \tag{34}$$

underlines our early equation (13). In this formal sense we have interpreted throughout a *single* first-order condition like equation (13) as the equilibrium outcome of a game between Memory and Deliberation, where we impose consistency of beliefs on the 'kernel of truth' component of DE.

4 A quantitative DE New Keynesian model

We leverage the previous qualitative insights to incorporate DE into a *quantitative New Keynesian model* of the type widely used for policy analysis. We emphasize the critical role played by endogenous predictability and distant memory recall in this new class of models. Methodologically, we formally rely on the naïveté approach to model beliefs, as argued earlier. This allows us to develop a tractable and recursive *solution method* to characterize equilibrium

laws of motion when agents act under DE. We estimate our model and show that it replicates the empirical boom-bust cycle in response to a monetary policy shock.

In deriving our theoretical results, we made use of the tractability arising in a model with Gaussian shocks where perceived tradeoffs are linear, thus maintaining conditional normality. However, in more general cases, the non-linear version of the model will not lead to a conditionally normal distribution. Indeed, in the class of models that we analyze in this section, it is the solution of the *log-linearized* model that has this property. In this setting, we exploit the convenient formulation of the representativeness heuristic based on the density h^θ in equation (5) by applying it on the log-linearized perceived tradeoffs. Our primitive approach, in line with what proposed in Bordalo et al. (2018), consists of emphasizing the role of the representativeness heuristic in distorting the perceptions of the marginal tradeoffs. We find the direct modeling of perceptions of linearized marginal tradeoffs as distorted by the density h^θ appealing because: (a) in linearized models these perceptions guide actual (marginally driven) decisions, and (b) in standard Gaussian environments these tradeoffs can be tractably characterized, a feature that we leverage throughout the paper. In our final remarks, we briefly discuss directions for future research, including how to allow for non-linearities while preserving tractability.

4.1 The model

The model features monopolistic competition in the labor and goods market, subject to adjustment costs in setting nominal prices. Consumption-investment decisions are influenced by real rigidities, in the form of habit formation and investment adjustment costs, and monetary policy follows a Taylor rule.

Household. The representative household chooses capital K_t^θ, investment I_t^θ, capital utilization rate u_t^θ, bonds B_t^θ, consumption C_t^θ, labor $N_{h,t}^\theta$ and nominal wage $W_{h,t}^\theta$ to solve

$$\max_{K_t^\theta, I_t^\theta, u_t^\theta, B_t^\theta, C_t^\theta, N_{h,t}^\theta, W_{h,t}^\theta} \left[\ln(C_t^\theta - b\overline{C}_{t-1}^\theta) - \frac{(N_{h,t}^\theta)^{1+\eta}}{1+\eta} + \beta \mathbb{E}_t^\theta \mathcal{V}(\mathcal{S}_{t+1}^\theta) \right] \tag{35}$$

subject to the budget constraint

$$\begin{aligned} &P_t^\theta C_t^\theta + P_t^\theta I_t^\theta + P_t^{B,\theta} B_t^\theta + (\varphi_w/2) \left(W_{h,t}^\theta / W_{h,t-1}^\theta - \gamma \Pi \right)^2 W_t^\theta \\ &= B_{t-1}^\theta + P_t^\theta R_t^{k,\theta} u_t^\theta K_{t-1}^\theta + W_{h,t}^\theta N_{h,t}^\theta + \int_0^1 D_{i,t}^\theta di - P_t^\theta a(u_t^\theta) K_{t-1}^\theta. \end{aligned}$$

where P_t^θ is the price level, $R_t^{k,\theta}$ is the capital rental rate, and $\int_0^1 D_{i,t}^\theta di$ is the combined current nominal profits from intermediate firms, given below in the firms' profit maximization problem. $P_t^{B,\theta}$ is the price of bond that pays 1 unit of consumption at $t+1$ so $P_t^{B,\theta} = 1/R_t^\theta$, where R_t^θ is the gross nominal interest rate. We allow for a capital utilization rate u_t^θ choice, subject to a resource cost specified as $a(u_t^\theta) = R^k(1+\tau)^{-1}\left((u_t^\theta)^{1+\tau} - 1\right)$. We explain the

continuation value $\mathcal{V}(.)$ in detail below.

Each household is monopolistically competitive in its labor supply. A perfectly competitive labor packer combines household labor and sells the composite labor N_t^θ to intermediate firms, described below, using the CES technology $N_t^\theta = \left[\int_0^1 (N_{h,t}^\theta)^{\frac{1}{\lambda_n}} dh\right]^{\lambda_n}$, where λ_n controls the steady-state wage markup. The packer's cost minimization leads to a standard demand curve taken by the household as an additional constraint in solving equation (35), namely $N_{h,t}^\theta = N_t^\theta \left(W_{h,t}^\theta / W_t^\theta\right)^{-\lambda_n/(\lambda_n - 1)}$, where W_t^θ is the aggregate wage level.

As we detail below, our approach handles large state space models, which allows us to incorporate DE into a NK model with *nominal and real frictions* that are typical of such quantitative business cycle models (see e.g., Christiano et al. (2005) and Smets and Wouters (2007)). In particular, the budget constraint above describes how nominal wages are subject to an adjustment cost (as in Kim (2000)), governed by the parameter φ_w, where γ is the rate of deterministic technological progress and Π is the steady-state inflation rate. On the preference side, note that in equation (35) we allow for habit formation, where $\overline{C}_{t-1}^\theta$ is the average aggregate consumption in the previous period and b is the external habit parameter.

Finally, the optimization in equation (35) is further subject to the physical capital law of motion, which features a standard quadratic investment adjustment cost

$$K_t^\theta = (1-\delta)K_{t-1}^\theta + \left\{1 - (\kappa/2)\left((I_t^\theta / I_{t-1}^\theta) - \gamma\right)^2\right\} I_t^\theta,$$

where δ is the depreciation rate and κ is the adjustment cost parameter.

As explained in Section 3, in this naïveté approach, in evaluating the continuation value $\mathcal{V}(.)$ in equation (35), the household assumes that her and other agents' future conditional preferences and resulting conditionally optimal actions will be taken under perfect memory (or RE), given values of the states entering next period, collected in the vector $\mathcal{S}_{t+1}^\theta$. To construct that continuation value we thus set up a 'shadow' economy (indexed by RE) where the household problem is solved under perfect memory, conditional on inherited states:

$$\mathcal{V}(\mathcal{S}_t^\theta) = \max_{K_t^{RE}, I_t^{RE}, u_t^{RE}, B_t^{RE}, C_t^{RE}, N_{h,t}^{RE}, W_{h,t}^{RE}} \left[\ln(C_t^{RE} - b\overline{C}_{t-1}^\theta) - (1+\eta)^{-1} N_{h,t}^{RE} + \beta \mathbb{E}_t \mathcal{V}(\mathcal{S}_{t+1}^{RE})\right],$$

subject to the budget constraint

$$\begin{aligned} &P_t^{RE} C_t^{RE} + P_t^{RE} I_t^{RE} + P_t^{B,RE} B_t^{RE} + (\varphi_w/2)\left(W_{h,t}^{RE}/W_{h,t-1}^\theta - \gamma\Pi\right)^2 W_t^{RE} \\ &= B_{t-1}^\theta + P_t^{RE} u_t^{RE} R_t^{k,RE} K_{t-1}^\theta + W_{h,t}^{RE} N_{h,t}^{RE} + \int_0^1 D_{i,t}^{RE} di - P_t^{RE} a(u_t^{RE}) K_{t-1}^\theta. \end{aligned}$$

The law of motion for capital is given by

$$K_t^{RE} = (1-\delta)K_{t-1}^\theta + \left\{1 - (\kappa/2)\left(I_t^{RE}/I_{t-1}^\theta - \gamma\right)^2\right\} I_t^{RE},$$

while the labor demand curve is simply $N_{h,t}^{RE} = N_t^{RE}\left(W_{h,t}^{RE}/W_t^{RE}\right)^{-\lambda_n/(\lambda_n-1)}$.

Firms. The final output is produced by a perfectly competitive representative firm that combines a continuum of intermediate goods $Y_{i,t}^{\theta}$ using the CES technology:

$$Y_t^{\theta} = \left[\int_0^1 (Y_{i,t}^{\theta})^{\frac{1}{\lambda_f}} di\right]^{\lambda_f},$$

where λ_f controls the steady-state markup. Intermediate goods firms' production function is $Y_{i,t}^{\theta} = (u_{i,t}^{\theta} K_{i,t}^{\theta})^{\alpha} (\gamma^t N_{i,t}^{\theta})^{1-\alpha}$, where $K_{i,t}^{\theta}$ and $N_{i,t}^{\theta}$ are the capital and labor employed by firm i. From the cost minimization problem, the real marginal cost is given by

$$MC_t^{\theta} = \frac{(R_t^{k,\theta})^{\alpha} (W_t^{\theta}/P_t^{\theta})^{1-\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha} u_{i,t}^{\theta} (\gamma^t)^{1-\alpha}}.$$

As with households, intermediate firms also face an adjustment cost (à-la Rotemberg (1982)) in changing their nominal price. Their problem is to choose $P_{i,t}^{\theta}$ to maximize

$$(C_t^{\theta} - bC_{t-1}^{\theta})^{-1} D_{i,t}^{\theta}/P_t^{\theta} + \beta \mathbb{E}_t^{\theta} \mathcal{V}_f(P_{i,t}^{\theta}), \tag{36}$$

where $D_{i,t}^{\theta} = \left(P_{i,t}^{\theta} Y_{i,t}^{\theta} - P_t^{\theta} MC_t^{\theta} Y_{i,t}^{\theta} - (\varphi_p/2)\left(P_{i,t}^{\theta}/P_{i,t-1}^{\theta} - \Pi\right)^2 P_t^{\theta} Y_t^{\theta}\right)$ and φ_p is the price adjustment cost parameter. $\mathcal{V}_f(.)$ is the continuation value

$$\mathcal{V}_f(P_{i,t-1}^{\theta}) = \max_{P_{i,t}^{RE}} \left[(C_t^{RE} - bC_{t-1}^{\theta})^{-1} D_{i,t}^{RE}/P_t^{\theta} + \beta \mathbb{E}_t \mathcal{V}_f(P_{i,t}^{RE})\right],$$

where $D_{i,t}^{RE} = \left(P_{i,t}^{RE} Y_{i,t}^{RE} - P_t^{RE} MC_t^{RE} Y_{i,t}^{RE} - (\varphi_p/2)\left(P_{i,t}^{RE}/P_{i,t-1}^{\theta} - \Pi\right)^2 P_t^{RE} Y_t^{RE}\right)$. Thus, in equation (36), firms' instantaneous payoff is given by current real profits and the continuation value is given by the discounted sum of real profits $V_f(P_{i,t}^{\theta})$. Under naïveté, in computing continuation value, agents assume that firms inherit the chosen price $P_{i,t}^{\theta}$ (which is relevant for the adjustment cost) but future prices are set according to RE.

Market clearing and monetary policy. The resource constraint is given by

$$C_t^{\theta} + I_t^{\theta} + (\varphi_p/2)\left(\Pi_t^{\theta} - \Pi\right)^2 Y_t^{\theta} + (\varphi_w/2)\left(\Pi_{w,t}^{\theta} - \gamma\Pi\right)^2 W_t^{\theta}/P_t^{\theta} + a(u_t^{\theta}) K_{t-1}^{\theta} = Y_t^{\theta},$$

where $\Pi_{w,t}^{\theta} \equiv W_t^{\theta}/W_{t-1}^{\theta}$ is nominal wage inflation. The central bank follows a Taylor rule:

$$R_t/R^{\theta} = \left(R_{t-1}^{\theta}/R\right)^{\rho_R} \left\{\left(\widetilde{\Pi}_t^{\theta}/\Pi\right)^{\phi_{\pi}} \left(Y_t^{G,\theta}/(\gamma Y_{t-1}^{G,\theta})\right)^{\phi_Y}\right\}^{1-\rho_R} exp(\varepsilon_t), \qquad \varepsilon_t \sim N(0, \sigma_R^2),$$

where $\widetilde{\Pi}_t^{\theta} \equiv 0.25 \sum_{s=0}^{3} \Pi_{t-s}^{\theta}$ is annual inflation, ε_t is the iid monetary policy shock, and GDP is defined as $Y_t^{G,\theta} \equiv C_t^{\theta} + I_t^{\theta}$. We provide the equilibrium conditions in Appendix C.

4.2 Solution method

Our solution method exploits the fact that naive DE agents expect future actions to be taken under the RE policy function. Below we outline our solution method. We provide additional details and formulas in Online Appendix D.

1. The first step of the solution algorithm consists of obtaining the shadow RE law of motion used by agents to form DE. We start from a linear RE system

$$\mathbf{\Gamma}_0 \mathbf{x}_t^{RE} = \mathbf{\Gamma}_1 \mathbf{x}_{t-1}^{RE} + \mathbf{\Psi} \varepsilon_t + \mathbf{\Pi} \eta_t^{RE}, \tag{37}$$

where $\mathbf{x}_t^{RE}$, ε_t and η_t^{RE} are vectors of endogenous variables, shocks, and expectation errors, respectively. This RE system is simply the RE version of the economy, with linear equilibrium conditions where DE ($\mathbb{E}_t^\theta$) is replaced with RE ($\mathbb{E}_t$).

A recursive law of motion can be obtained, using for example Sims (2000),

$$\mathbf{x}_t^{RE} = \mathbf{T}^{RE} \mathbf{x}_{t-1}^{RE} + \mathbf{R}^{RE} \varepsilon_t.$$

2. Consider a linear DE system

$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta \mathbb{E}_t^\theta [\mathbf{y}_{t+1}^{RE}] + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t, \tag{38}$$

where we provide expressions for $\mathbf{\Gamma}_0^\theta$, $\mathbf{\Gamma}_2^\theta$, $\mathbf{\Gamma}_1^\theta$ and $\mathbf{\Psi}^\theta$ in the Appendix. Relative to the RE system (37), which implicitly defines expectations in $\mathbf{x}_t^{RE}$ by using expectation errors η_t^{RE}, the DE system (38) explicitly accommodates DE ($\mathbb{E}_t^\theta[\mathbf{y}_{t+1}^{RE}]$).

We can substitute the $\mathbb{E}_t^\theta[\mathbf{y}_{t+1}^{RE}]$ in the DE system (38) as

$$\mathbb{E}_t^\theta [\mathbf{y}_{t+1}^{RE}] = \mathbb{E}_t [\mathbf{y}_{t+1}^{RE}] + \theta (\mathbb{E}_t [\mathbf{y}_{t+1}^{RE}] - \mathbb{E}_t^r [\mathbf{y}_{t+1}^{RE}]), \tag{39}$$

where $\mathbb{E}_t^r[\mathbf{y}_{t+1}^{RE}]$ denotes the comparison group, or the *reference* distribution, characterizing the representativeness heuristic. Our method allows for a general form of memory recall and thus of this comparison group. In particular, as we further explain below, we model this reference distribution in a flexible, yet parsimonious manner, as a weighted average of lagged RE expectations:

$$\mathbb{E}_t^r [\mathbf{y}_{t+1}^{RE}] = \sum_{j=1}^{J} \alpha_j \mathbb{E}_{t-j} [\mathbf{y}_{t+1}^{RE}], \tag{40}$$

where $\{\alpha_j\}_{j=1}^J$ are weight parameters on lagged expectations (and thus $\sum_{j=1}^J \alpha_j = 1$). Let $\mathbf{y}_t^{RE} = \mathbf{M} \mathbf{x}_t^{RE}$, where $\mathbf{M}$ is a selection matrix that selects variables from a vector $\mathbf{x}_t^{RE}$. Given the DE beliefs characterized by (39) and (40), the system (38) then becomes

$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta \left[(1+\theta) \mathbf{M} \mathbf{T}^{RE} \mathbf{x}_t^\theta - \sum_{j=1}^{J} \theta \alpha_j \mathbf{M} \left(\mathbf{T}^{RE} \right)^{j+1} \mathbf{x}_{t-j}^\theta \right] + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t, \tag{41}$$

where $\mathbf{T}^{RE}$ is the auto-regressive component of the RE solution. The expression (41) also clarifies that agents form DE based on state variables inherited from the DE economy, but assuming that in the future the economy follows the RE law of motion.

3. Inverting matrices and rewriting (41) more compactly gives the DE law of motion:

$$\mathbf{z}_t^\theta = \mathbf{T}^\theta \mathbf{z}_{t-1}^\theta + \mathbf{R}^\theta \varepsilon_t, \tag{42}$$

where we provide expressions for $\mathbf{T}^\theta$ and $\mathbf{R}^\theta$ in Appendix D and $\mathbf{z}_t^\theta$ is a vector that includes not only $\mathbf{x}_t^\theta$ but also its lags. Finally, we check that all variables over which we take DE present residual uncertainty using the formula we provide in Appendix D.

The key advantages of our solution method are thus its *portability and tractability*: a researcher can transform a standard linear dynamic equilibrium model (37) and compute the DE law of motion (42) with a few additional lines of code.

4.3 Estimation

Our aim is to demonstrate that DE matter in practical and policy-relevant settings. We choose the estimation method that aligns with this goal. The starting point of our analysis is a local projection estimation of impulse responses to a monetary policy shock using U.S. quarterly macroeconomic data over the sample period 1969Q1–2006Q4.[14] Specifically, we estimate the following regressions:

$$x_{t+h} = c^h + \tau_1^h t + \tau_2^h t^2 + \sum_{l=1}^{L} A_l^h x_{t-l} + \sum_{i=0}^{I} B_i^h e_{t-i} + u_{t+h}, \quad h = 0, \ldots, H$$

where x_t is the variable of interest and e_t is the Romer and Romer (2004) monetary policy shock, extended by Coibion et al. (2017). The coefficients of interest are $\{B_0^h\}_{h=0}^H$. We set $L = 4$ and $I = 0$ and compute the impulse response for $H = 32$ horizons.

We estimate the model parameters using the Bayesian version of the impulse-response-matching method, developed by Christiano et al. (2010). In this method, the likelihood depends on how closely the model matches the empirical response to a shock. The likelihood is then combined with priors on the model parameters.[15] In our empirical analysis below, we target the impulse responses of four variables: log real per capita consumption, log per capita hours worked, log GDP deflator inflation, and the log Federal Funds rate (FFR). We then also use the implied responses of four other variables, namely log real per capita investment, log real per capita GDP, SPF expected inflation, and SPF GDP growth expectations, as 'untargeted' moments that serve as external validation.[16]

[14] We do not include the period after 2007Q1 to avoid complications arising from the zero lower bound.

[15] We provide a detailed description of the estimation method in the Online Appendix E.

[16] To obtain real per capita GDP we divide real GDP by total population. Real per capita consumption is measured by the sum of personal consumption expenditure on nondurables and services divided by total population. Real per capita investment is the sum of gross private domestic investment and personal consumption expenditure on durables divided by total population. Per capita hours worked is the total hours in nonfarm business sector divided by total population.

We fix several parameters before the estimation. The deterministic growth rate γ and the steady-state inflation rate Π are set to 1.004 and 1.01, respectively, which imply a steady-state annual output growth rate of 1.6% and the annualized inflation rate of 4%. The capital share α, the discount factor β and the depreciation rate δ are set to 0.3, 0.99 and 0.025, respectively. We set λ_f and λ_n to 1.1, which imply steady-state price and wage markups of 10%. For parameters that are common in the New Keynesian literature, we center our priors around conventional values.

For the diagnostic parameter θ, we choose a conservative prior that puts significant weight on the RE case ($\theta = 0$) but also encompasses the estimates found in Bordalo et al. (2018) and Bordalo et al. (2019b) ($\theta \approx 1$). Specifically, we chose a Normal distribution with mean 0 and standard deviation 0.2, but we truncate this prior above $\theta \geq 0$ so as to be consistent with the theoretical restriction that the diagnosticity parameter has to be (weakly) positive. Note that the prior mode of this truncated Normal distribution is 0.[17] As explained above (see equation (40)), we allow for flexible reference expectations in memory recall and thus the comparison group is a weighted average of lagged expectations. To estimate the weights $\{\alpha_j\}_{j=1}^{J}$ on past memory, we consider a parsimonious parameterization. We set $J = 32$ and estimate the mean μ and the standard deviation σ of a Beta distribution. We then rescale and discretize the implied $Beta(\mu, \sigma^2)$ distribution to span the discrete interval [0,32] and obtain the weights $\widetilde{\alpha}_j$. We then apply the transformation $\alpha_j = \widetilde{\alpha}_j / (\sum_{j=1}^{J} \widetilde{\alpha}_j)$ so that $\{\alpha_j\}_{j=1}^{J}$ sum to one. We report the priors and all estimated parameters in Table 1 in the Online Appendix, while below we focus on the key parameters that control the effects of DE.

4.4 Results

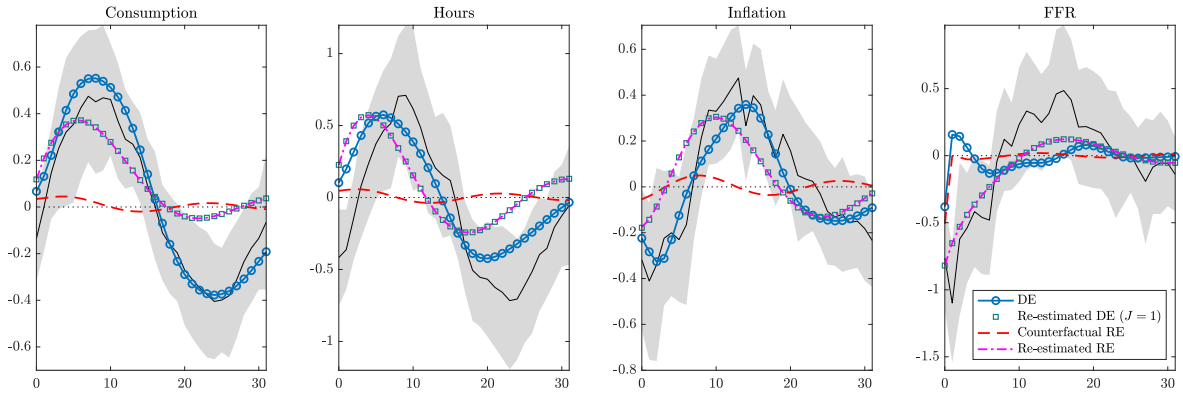
Figure 2 presents the local projection impulse responses (black solid lines) to a one-standard-deviation expansionary monetary policy shock along with the 90% confidence bands. In response to a reduction in the FFR, real variables such as hours and consumption all increase in a hump-shaped manner, peaking around 10 quarters after the initial shock. These variables then undershoot below the steady states and reach their trough around 5 to 6 years after the shock, followed by a gradual recovery.[18] Inflation builds up slower and tends to peak at the end of the boom, followed by a slow return to the steady state.

The DE New Keynesian model (blue lines with circles) reproduces the empirical impulse response functions (IRF) well, successfully generating the boom-bust cycle observed in the data. We use three alternative models to argue that the DE distortion *and* distant memory provide the key mechanism for this successful fit.

[17] Online Appendix F shows that results are similar if we center the prior for θ around 1.

[18] McKay and Wieland (2021) find a similar boom-bust pattern in their estimated responses to a monetary policy shock.

Figure 2: Impulse responses to a monetary policy shock: Fit for targeted responses



Notes: This figure reports impulse responses for targeted variables. The black lines are the mean responses from the local projection and the shaded areas are the 90% confidence bands. The blue circled lines are IRFs from the baseline model with DE, allowing for distant memory. The green squares are IRFs from the DE model imposing that recall is based only on recent memory. In this case, $J = 1$ by assumption. The red dashed lines are IRFs from the counterfactual RE model where we set $\theta = 0$ while holding fixed other estimated parameters. The magenta dashed lines are IRFs from the re-estimated RE model. The consumption and hours responses are in percentage deviations from steady states while inflation and the FFR are in annual percentage points.

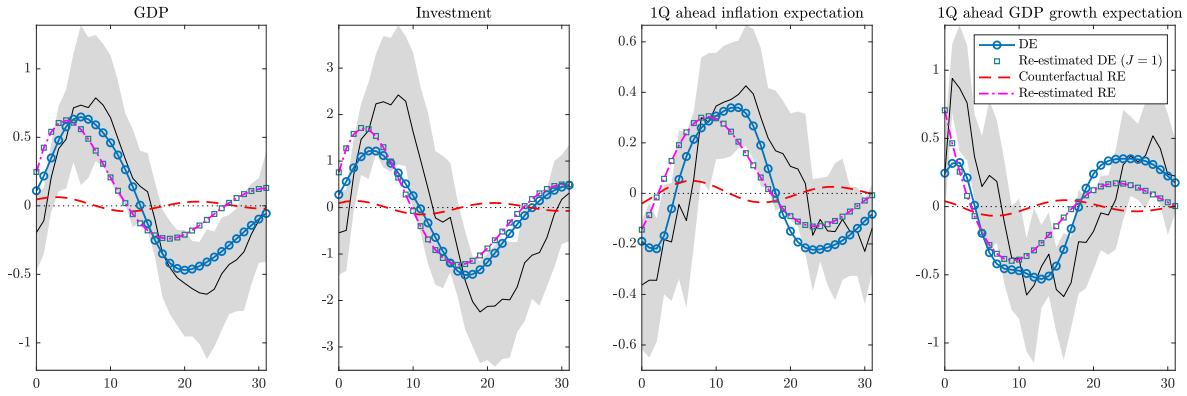
Alternative models. First, we evaluate a *counterfactual* RE model, where we set the diagnostic parameter $\theta = 0$, while holding fixed all other estimated parameters. This counterfactual model (red dashed lines) generates transitory and negligible responses, indicating that much of our success is due to the DE mechanism. Second, we re-estimate the model under RE, i.e., imposing the constraint that $\theta = 0$. The re-estimated RE model (magenta dashed lines) fails in delivering the empirical boom-bust dynamics and the amplitude of the IRFs. As a result, the marginal likelihood, a Bayesian measure of fit that penalizes models with more parameters, is $(-345 - (-369) =)$ 24 log points higher in the DE model.

In the third and final exercise, we re-estimate the DE model (i.e., we allow for $\theta \geq 0$) but *impose* the constraint that $J = 1$. Strikingly, we then estimate a value of $\theta = 0$. As a result, in this alternative model that constrains memory recall to be based entirely on the immediate past, the IRFs (green squares) coincides with those of the re-estimated RE model.[19] This exercise showcases how in our model distant memory and DE distortion θ are *complements*, since they interact and jointly magnify the role of DE.

Untargeted moments. The model also matches remarkably well the IRFs that were

[19]It is then not surprising that the marginal likelihood of the re-estimated RE model beats that of the re-estimated DE model with $J = 1$ by $(-369 - (-371) =)$ 2 log points, because the DE model with $J = 1$ has an additional parameter θ that is not estimated to be significant.

Figure 3: Impulse responses to a monetary policy shock: Fit for untargeted responses



Notes: This figure reports impulse responses for untargeted variables. The black lines are the mean responses from the local projection and the shaded areas are the 90% confidence bands. The blue circled lines are IRFs from the baseline model with DE, allowing for distant memory. The green squares are IRFs from the DE model imposing that recall is based only on recent memory. In this case, $J = 1$ by assumption. The red dashed lines are IRFs from the counterfactual RE model where we set $\theta = 0$ while holding fixed other estimated parameters. The magenta dashed lines are IRFs from the re-estimated RE model. The responses of GDP and investment are in percentage deviations from the steady states while the inflation and output growth expectations are in annual percentage points.

not targeted in the estimation. The first two panels of Figure 3 report the responses of GDP and investment to the monetary policy shock. The model delivers a good fit. The right two panels of Figure 3 report the impulse response of expected inflation and expected output growth.[20] The model generates expectations that are very much in line with those observed in the data, even though we did not target those expectations in our estimation exercise.[21] Figure 3 also shows that the re-estimated RE version does a worse job in accounting for these untargeted moments. Formally, we find that the root-mean-square error (RMSE) for the DE model is 0.52, while for the re-estimated model RE it is 0.64. Crucially, focusing only on the untargeted survey moments (inflation and output growth expectations), the RMSE for the DE model is 0.25, while for the re-estimated RE model it is 16% larger, at 0.29.[22]

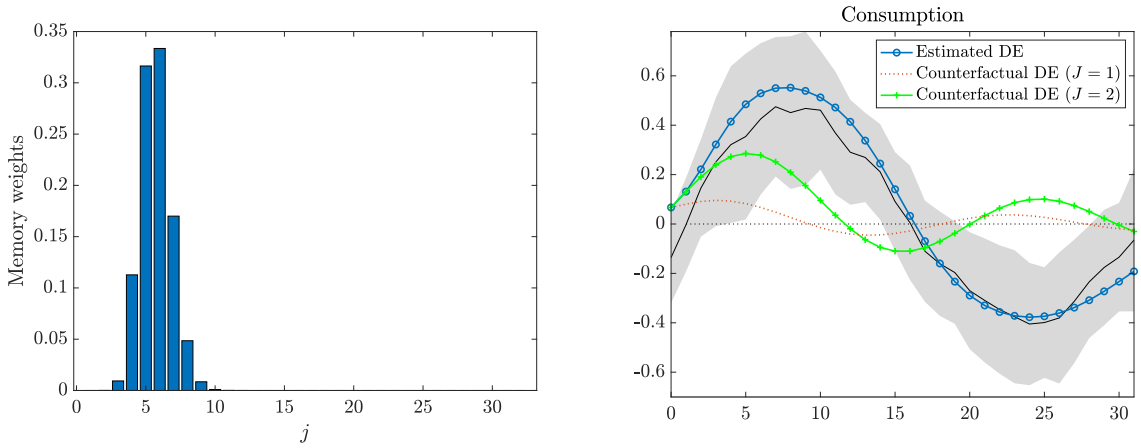
[20] We measure inflation and output growth expectations using the median of the SPF survey responses of one-quarter-ahead inflation and output growth expectations, respectively. We assume that the model implied inflation and output growth expectations coincide with what a DE agent in the model would predict ($\mathbb{E}_t^{\theta}[\widehat{\pi}_{t+1}^{RE}]$ and $\mathbb{E}_t^{\theta}[\Delta \widehat{Y}_{t+1}^{G,RE}]$).

[21] In the Online Appendix F, we estimate the model targeting inflation and output growth expectations and show that it can generate boom-bust cycles in macro variables.

[22] We compute $RMSE = \sqrt{\sum_{i=1}^{N}\sum_{t=1}^{T}(IRF_{data,t}^{i} - IRF_{model,t}^{i})^2/T}$, where $IRF_{data,t}^{i}$ and $IRF_{model,t}^{i}$ indicate the local projection IRF and model IRF, respectively, for (i) GDP, investment and expected inflation and output growth (all untargeted moments) or for (ii) expected inflation and output growth (untargeted survey moments only).

Figure 4: Estimated selective memory

(a) Estimated memory weights α_j (b) Empirical role of distant memory



Notes: The left panel reports the estimated memory weights α_j. The right panel reports the consumption IRF in the estimated DE model (blue circled line) and the counterfactual model where only recent memory matters (orange dotted line) and when only two-period-ago memory matters (green line with plus signs).

DE parameters. We estimate $\theta = 1.97$ for the parameter controlling the severity of the DE distortion. This value is in the same order of magnitude of previous estimates (for example, Bordalo et al. (2018), Bordalo et al. (2019a), d'Arienzo (2020), L'Huillier et al. (2021), which tend to estimate $\theta \approx 1$), even if larger. However, we note that the existing estimates are based primarily on models where imperfect memory is assumed to be driven only by the immediate past, an assumption that we show fundamentally changes inference in our structural model (per our discussion of the DE version imposing $J = 1$).

The mean and standard deviation of the Beta distribution that controls the weights $\alpha_j's$ attached to each of the $J = 32$ lagged expectations entering the comparison group are 0.17 and 0.03, respectively. As shown in the left panel of Figure 4, these estimates imply that the weights are centered on the expectations formed six quarters ago, with positive weights assigned to expectations formed between three and ten quarters ago. To further examine the importance of distant memory and complement our previous discussion on alternative models, the right panel of Figure 4 shows how the impulse response for consumption changes as we vary the lag for the reference distribution. We consider the counterfactual case in which only recent memory matters ($J = 1$) or when only two-period-ago expectations matter ($J = 2$). Other parameters are fixed at the benchmark estimates. Reducing the lag impacts the frequency and the amplitude of the boom-bust cycles. As we discuss below, when J increases the effects of past misperceptions accumulate, leading to larger fluctuations.

4.5 Mechanism

We have emphasized throughout this paper that the interaction of endogenous states and distant memory affects equilibrium actions and DE beliefs. In this class of NK models, these effects occur both on the real and nominal side of the economy, which in equilibrium are jointly determined. To describe the overall mechanism in this model, we first focus on consumption dynamics by leveraging the qualitative insights on repeated boom-busts of the PIH model of Subsection 3.3. In contrast to the PIH model, where the real interest rate was constant, a consumption boom-bust is now accompanied by a corresponding movement in the perceived real interest rate. Thus, our second line of argument is to describe the nominal side, and in particular the novel and critical role played by perceptions over inflation.

4.5.1 Capital surprise as an endogenous informational state

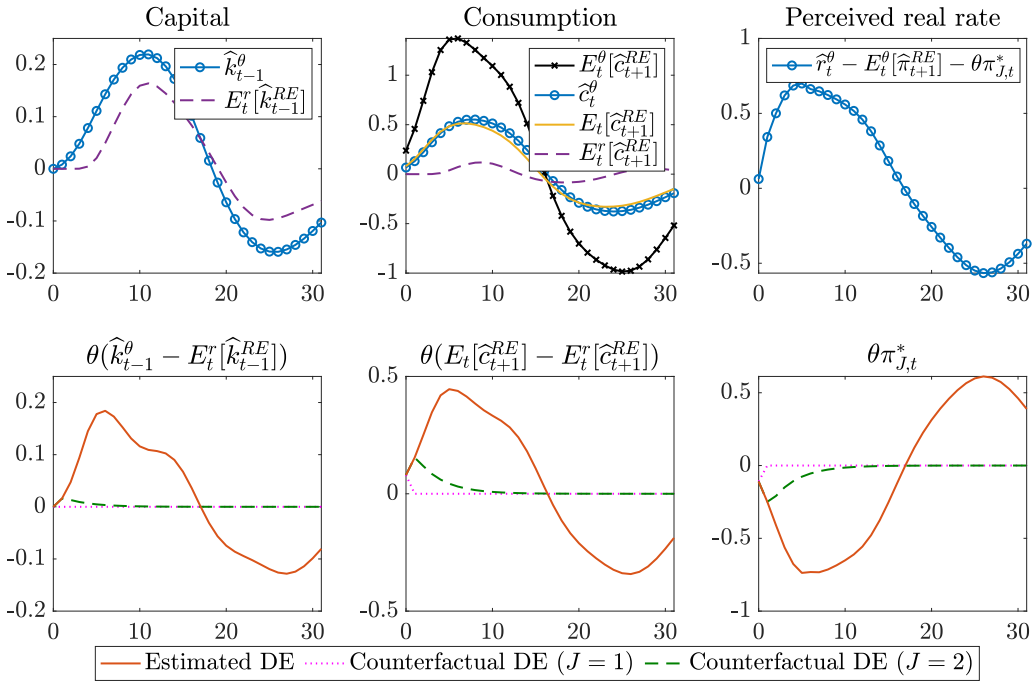
We have described in detail in the PIH model how the interaction of endogenous predictability and distant memory delivers a novel, informational state, capturing surprises over the endogenous state of capital. A similar force characterizes this richer NK model. In particular, resembling Proposition 6, the surprise $\mathbb{N}^k_{J,t-1} \equiv \widehat{k}^{\theta}_{t-1} - \mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}]$ emerges as an endogenous informational state. Here $\mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}]$ is the reference expectation for $\widehat{k}_{t-1}$, where we use lowercase letters with hats to denote variables in log-deviations from the steady states.[23]

Figure 5 shows the path of equilibrium capital $\widehat{k}^{\theta}_{t-1}$, its reference expectation $\mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}]$ (in top left panel), and the resulting equilibrium surprise $\mathbb{N}^k_{J,t-1}$ (solid line in bottom left panel). An expansionary monetary policy shock stimulates consumption and investment so capital $\widehat{k}^{\theta}_{t-1}$ increases. The reference distribution $\mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}]$ moves slowly, so the agent is positively surprised by the resources available. Due to these positive surprises, consumption and investment further rise, which in turn leads to more capital stock and further positive capital surprises. This virtuous feedback loop continues until the reference expectation $\mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}]$ of capital begins to catch up to the realized $\widehat{k}^{\theta}_{t-1}$. The agent is then less optimistic about the future and begins cutting back on consumption and investment.

Eventually, the economy enters a bust phase when the agent becomes *disappointed* in the level of capital relative to the reference distribution formed at the height of the boom. The capital surprise $\mathbb{N}^k_{J,t-1}$ thus turns from positive to negative, leading now the agent to over-correct. Consumption is now reduced below the steady-state level, pushing down the level of aggregate demand and capital accumulation. A bust period arises, where the feedback between perceived pessimism leads to further economic declines and disappointment in the agent's perceptions of her resources (relative to her more optimistic forecast formed several

[23] By equation (40), this reference expectation is a weighted average of projections over capital at $t-1$, conditional on $t-j$ information. When $j=1$, this projection is simply the observed realized state $\widehat{k}^{\theta}_{t-1}$, while for $j>1$ the projection makes $t-j$ conditional forecasts over the uncertain $\widehat{k}_{t-1}$ using the RE law of motion.

Figure 5: Capital, consumption, and perceived real rate



Notes: The figure reports the response to the monetary policy shock for capital, consumption, and the perceived real rate (top panels). We also report responses of the capital and consumption surprises and the perceived innovation in cumulative inflation, scaled by θ (bottom panels).

periods ago). Thus, similar to the illustrations in the PIH model, in this quantitative model an endogenous boom-bust in the informational state $\mathbb{N}^k_{J,t-1}$ emerges, triggered by the sole realization of an iid shock (here the monetary policy shock).

As emphasized in Section 3, when memory recall is based only on immediate past (i.e., $J = 1$), there is no perceived surprise over the endogenous state $\widehat{k}^{\theta}_{t-1}$, since its realization entirely informs the comparison group relevant for the DE beliefs. Indeed, in a counterfactual case where $J = 1$, the reference expectation $\mathbb{E}^r_t[\widehat{k}^{RE}_{t-1}] = \widehat{k}^{\theta}_{t-1}$ and thus the perceived surprise $\mathbb{N}^k_{J,t-1} = 0$ at all times, as shown by the flat magenta dotted line in the bottom left panel of Figure 5. The same panel also shows that as soon as $J = 2$ the surprise $\mathbb{N}^k_{J,t-1}$ is activated (the green dashed line). However, in the counterfactual DE model where $J = 2$, the duration and magnitude of the surprise are both very small. In contrast, consistent with the discussion around the PIH model (see also Figure 1), the more distant memory estimated in our model increases both the duration and the magnitude of the boom-bust cycles in the surprise $\mathbb{N}^k_{J,t-1}$.

4.5.2 Joint real and nominal dynamics

We now analyze the key mechanism behind the joint determination of the real and the nominal side, and in particular the novel and critical role played by inflation dynamics. For this purpose, we focus on the Euler equation for bonds, momentarily ignoring habit formation:

$$\frac{\left(C_t^\theta\right)^{-1}}{P_t^\theta} = \beta R_t^\theta \mathbb{E}_t^\theta \left[\frac{\left(C_{t+1}^{RE}\right)^{-1}}{P_{t+1}^{RE}}\right],$$

that, expressing in terms of deviations from the steady state and rearranging, becomes

$$\begin{aligned} \mathbb{E}_t^\theta[\widehat{c}_{t+1}^{RE}] - \widehat{c}_t^\theta &= \widehat{r}_t^\theta - \mathbb{E}_t^\theta[\widehat{\pi}_{t+1}^{RE}] - \theta \underbrace{\left[p_t^\theta - \mathbb{E}_t^r\left[p_t^{RE}\right]\right]}_{\text{Surprise in price level}} \\ &= \underbrace{\widehat{r}_t^\theta - \mathbb{E}_t^\theta[\widehat{\pi}_{t+1}^{RE}] - \theta\pi_{J,t}^*}_{\text{Perceived real rate}}, \end{aligned} \tag{43}$$

where

$$\pi_{J,t}^* \equiv \underbrace{\textstyle\sum_{j=1}^J \alpha_j \left(\widehat{\pi}_{t-j+1,t}^\theta - \mathbb{E}_{t-j}[\widehat{\pi}_{t-j+1,t}^{RE}]\right)}_{\text{Perceived innovation in cumulative inflation}} = p_t^\theta - \mathbb{E}_t^r\left[p_t^{RE}\right],$$

and each term $\widehat{\pi}_{t-J+1,t} = \widehat{\pi}_{t-J} + \widehat{\pi}_{t-J+1} + ...\widehat{\pi}_t = p_t - p_{t-J} - \pi$ denotes the *cumulative* inflation between $t-J$ and t. For further reference, we denote that surprise, or *perceived innovation in cumulative inflation*, as the equilibrium object $\pi_{J,t}^*$.[24]

Thus, in the Euler equation (43) the perceived consumption growth (on the LHS) equals the perceived real rate (on the RHS), where both equilibrium objects are jointly formed under beliefs driven by DE. We analyze these two objects in turn.

Consider first expected consumption growth under DE, given by:

$$\mathbb{E}_t^\theta[\widehat{c}_{t+1}^{RE}] - \widehat{c}_t^\theta = \mathbb{E}_t[\widehat{c}_{t+1}^{RE}] + \quad \theta \underbrace{\left(\mathbb{E}_t[\widehat{c}_{t+1}^{RE}] - \mathbb{E}_t^r[\widehat{c}_{t+1}^{RE}]\right)}_{\text{Surprise in expected consumption}} \quad - \widehat{c}_t^\theta \tag{44}$$

where $\mathbb{E}_t^r[\widehat{c}_{t+1}^{RE}] = \sum_{j=1}^J \alpha_j \mathbb{E}_{t-j}\widehat{c}_{t+1}^{RE}$, by equation (40). The top middle panel of Figure 5 plots the elements entering equation (44). During the boom (bust) phase $\widehat{c}_t^\theta$ and $\mathbb{E}_t[\widehat{c}_{t+1}^{RE}]$ rise (fall) by a similar amount, with the reference expectation $\mathbb{E}_t^r[\widehat{c}_{t+1}^{RE}]$ moves sluggishly. In turn, the DE beliefs $\mathbb{E}_t^\theta[\widehat{c}_{t+1}^{RE}]$ overreact by a factor of θ to the surprise $(\mathbb{E}_t[\widehat{c}_{t+1}^{RE}] - \mathbb{E}_t^r[\widehat{c}_{t+1}^{RE}])$ in expected consumption. This surprise, plotted in the bottom middle panel of Figure 5, is an endogenous equilibrium object. The top and bottom middle panels thus show that this surprise is a key driver of the expected consumption growth under DE.

As made transparent by our consumption-smoothing model, the surprise in expected

[24] In the special case of $J = 1$, per our earlier analytical results, equilibrium variables under the RE law of motion respond to endogenous states in the same way as they do under the DE law of economy, making the equilibrium perceived innovation in cumulative inflation take the simpler but equivalent form $\pi_{1,t}^* = \widehat{\pi}_t^\theta - \mathbb{E}_{t-1}\widehat{\pi}_t^\theta$. This form recovers the nominal price surprise object that distorts consumption smoothing in the NK model of L'Huillier et al. (2021) who focus their analysis entirely on the $J = 1$ case.

consumption is generally a function of the surprise in both (i) the exogenous innovation and (ii) the endogenous states. It is only when $J = 1$ that the latter does not matter, since then the time t exogenous shock is the only change in the information set from the immediate past $t-1$ to current t. The magenta dotted line in Figure 5 confirms that in a counterfactual with $J = 1$ the surprise in expected consumption moves only at the time of the exogenous shock.

Instead, when memory is based on more distant past, the perceived surprises embedded in the realized path of the endogenous states matter for the DE overreaction to $\mathbb{E}_t[\widehat{c}^{RE}_{t+1}]$ in equation (44). In fact, the middle bottom panel of Figure 5 shows that the path of the surprise in expected consumption tracks closely the path of the surprise in capital, $\mathbb{N}^k_{J,t-1}$ (plotted in the bottom left panel). Intuitively, like in the consumption-smoothing model of Section 3, a positive (negative) surprise $\mathbb{N}^k_{J,t-1}$ makes the agent overly optimistic (pessimistic) about future resources.[25] The endogenous boom-bust in the information state $\mathbb{N}^k_{J,t-1}$ is thus reflected in periods of endogenous optimism and pessimism over future consumption.[26] The counterfactual of $J = 2$ (green dashed line) indicates again that further memory lags, like in our estimated model, amplify both the duration and magnitude of the boom-bust dynamics.

We now turn to the formation of DE beliefs over nominal prices and their role in affecting the perceived real rate. The top right panel of Figure 5 shows how the boom and bust in expected consumption growth $\mathbb{E}^\theta_t[\widehat{c}^{RE}_{t+1}] - \widehat{c}^\theta_t$ discussed above is mirrored by a corresponding rise and fall in the perceived real interest rate, $\widehat{r}^\theta_t - \mathbb{E}^\theta_t[\widehat{\pi}^{RE}_{t+1}] - \theta\pi^*_{J,t}$. By equation (43), DE affect this rate through two channels. The first is DE over future inflation, $\mathbb{E}^\theta_t[\widehat{\pi}^{RE}_{t+1}]$. The second channel is the perceived surprise in the price level, or $\pi^*_{J,t}$. Intuitively, holding constant $\mathbb{E}^\theta_t[\widehat{c}^{RE}_{t+1}]$ and $\left(\widehat{r}^\theta_t - \mathbb{E}^\theta_t[\widehat{\pi}^{RE}_{t+1}]\right)$, a higher innovation $\pi^*_{J,t}$ makes the perceived expected future price relatively high, thus lowering the incentives to postpone consumption.

The bottom right panel of Figure 5 shows how this second channel, operating through the surprise $\pi^*_{J,t}$, drives most of the variation in the perceived real rate. To understand the equilibrium path of $\pi^*_{J,t}$, recall from Figure 2 that on impact, due to an increase in utilization, inflation $\widehat{\pi}^\theta_t$ declines, which determines a negative surprise in the price level. As shown in Figure 2, $\widehat{\pi}^\theta_t$ then starts to gradually recover and eventually rises above steady state during the economic boom, as in the data. This rise first leads to a recovery in the surprise $\pi^*_{J,t}$ back to steady state from below. Importantly, as inflation accelerates at the end of the boom,

[25] While this intuition is similar to the consumption-smoothing model where capital was the only relevant endogenous state, in this rich NK model, due to its nominal and real frictions, the set of relevant endogenous states that affect $\mathbb{E}_t\left[\widehat{c}^{RE}_{t+1}\right]$ is larger than just $\widehat{k}_{t-1}$. However, the close proportionality between the path of surprises in $\mathbb{E}_t\left[\widehat{c}^{RE}_{t+1}\right]$ and $\widehat{k}_{t-1}$ indicates that in equilibrium the former is primarily influenced by the latter.

[26] The Online Appendix shows that the perceived increase in consumption more than compensates for the habit stock. In other words, not only agents expect consumption to be higher in the future, but they also expect it to grow with respect to the habit stock, lowering the marginal utility. Furthermore, we show that even without consumption habit our model is able to generate boom-bust cycles.

selective memory recall starts to increasingly weigh the high price level states, leading to positive surprises in $\pi^*_{J,t}$, as indicated in Figure 5. This path determines a *reversal* in the perceived innovation in cumulative inflation, which now moves into the positive territory during the bust part of the cycle.

DE beliefs determine misperceptions on the real and nominal side of the economy that are consistent with each other. Indeed, the bottom row of Figure 5 shows how the cycle of optimism/pessimism over future consumption tracks in equilibrium the cycle of negative/positive perceptions of prices. In particular, in the boom phase, the over-reaction to negative surprises in inflation lead to a high perceived real rate that is consistent with a perceived acceleration in consumption arising from over-reactions to capital surprises. As the reference distribution for both capital and prices slowly adjusts to their corresponding realized path, a reversal occurs. The economic boom endogenously creates the conditions for a bust, where negative surprises over capital and perceived deceleration in future consumption are consistent with high perceptions of future price levels and a low perceived real rate.

The same bottom row of Figure 5 shows systematically across its three panels the importance of our estimated distant memory process. Distant memory creates larger revisions in expectations, leading to larger surprises, and larger belief distortions. This explains why the parameter J does not only affect the frequency of the boom-bust cycle but also the amplitude. When $J = 1$, there are no surprises in the endogenous states, and the over-reaction in expected consumption and price level arises only at the time of the shock. When $J = 2$, the dynamics are still small and short-lived. Instead, under a more distant memory, agents expectations are constantly revised and missperceptions build. In our model, past decisions affect current expectations and generate new distortions that feed into current decisions, creating endogenous waves of optimism and pessimism - a form of Minsky (1977) moments.

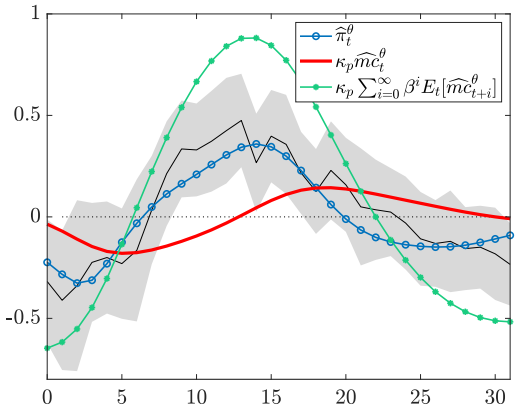
The success of the DE model also stems from its ability to accurately match inflation dynamics. As in the data, inflation movements seem relatively small with respect to fluctuations in real activity. To study this disconnect, consider the relation between inflation and marginal costs (see Online Appendix F for the derivation):

$$\widehat{\pi}_t^\theta = \kappa_p \sum_{i=1}^{\infty} \beta^i \mathbb{E}_t^\theta[\widehat{mc}_{t+i}^{RE}] + \kappa_p \widehat{mc}_t^\theta.$$

The above expression makes clear that inflation depends on the DE of future marginal costs for a given starting value of current marginal costs. To understand the effect of the distorted beliefs about future marginal costs, consider the following counterfactual measure of inflation:

$$\widehat{\pi}_t^{CF} = \kappa_p \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t[\widehat{mc}_{t+i}^\theta].$$

Figure 6: Inflation, marginal costs, and New Keynesian Phillips Curve



Notes: The figure reports the response to the monetary policy shock for inflation, marginal costs (scaled by κ_p), and a counterfactual measure of inflation built assuming that agents can correctly foresee the future path of marginal costs.

The expression above captures the path of inflation that an econometrician who can accurately predict the *realized* path of marginal costs would compute.[27]

Figure 6 reports the path for marginal costs (scaled by κ_p), inflation, and for the counterfactual measure of inflation. On impact, DE inflation drops because an increase in utilization lowers marginal costs. DE inflation keeps declining for a couple of quarters, as in the data, because agents keep revising their expectations about future marginal costs. The counterfactual measure of inflation also drops on impact, but it immediately starts increasing because there are no further revisions in expectations after the first period. As marginal costs start increasing, DE inflation starts recovering, and so does the counterfactual measure of inflation.

Importantly, the counterfactual measure of inflation shows much larger fluctuations than actual inflation. This is because the actual path of marginal costs is more persistent than what is perceived by our agents. Thus, for a given movement in marginal costs, DE lead to *under-reaction* of inflation because agents expect a relatively fast return of marginal costs to the steady state. An external observer who were endowed with the path of real activity and marginal costs would conclude that the Phillips curve is quite flat. Indeed, the estimated value for the slope of Phillips curve, κ_p, is significantly larger for the DE model, at 0.0502, compared to 0.0337 in the re-estimated RE model. While the RE model needs to appeal to a flat Phillips curve to try to reconcile the dynamics in real activity and inflation, the DE model is able to reconcile them based on the distorted expected path for marginal costs.

We conclude this subsection by discussing one additional point. In this economy there is

[27]Notice that this measure of inflation does not coincide with the shadow RE inflation, because the expected path of marginal costs is based on the DE economy.

positive co-movement between the key real aggregate variables (consumption, investment, hours). The economic channel is typical to the New Keynesian models. Intuitively, following the expansionary monetary policy shock, the demand for goods (consumption and investment) is stimulated. In this demand-driven economy, equilibrium is largely restored through a higher capacity utilization, which not only directly increases the supply of goods, but also leads to a larger labor productivity and thus stimulates firms' labor demand.

5 Conclusions

In this paper, we build on the DE paradigm proposed by Bordalo et al. (2018) to analyze the qualitative and quantitative implications of the joint determination of DE beliefs and optimal actions in the presence of (i) endogenous states and (ii) distant memory recall. In the first part of the paper, we use a three-period consumption-savings model as a laboratory to provide behavioral micro-foundations for our analysis that we argue are psychologically and model-coherent. We then extend the model to the infinite horizon to show that under distant memory the interaction between actions and DE beliefs naturally generate repeated boom-bust cycles in response to a single initial shock. In the second part of the paper, we develop a portable solution method that can be used to enrich standard general equilibrium models with DE. We incorporate DE into a quantitative New Keynesian model of the type widely used for policy analysis. We uncover a critical and novel role played by endogenous states and distant memory recall, which allows the DE model to replicate the empirical boom-bust cycle dynamics in response to a monetary policy shock.

There are two main avenues for future research. First, deriving and studying optimal monetary policy under different behavioral assumptions regarding agents' expectations would have important policy implications and further expand the practical relevance of DE. Second, it will be interesting to allow for non-linearities, such as changes in policy makers' behavior, stochastic volatility, and occasionally binding constraints. The methods developed in this paper can be extended to accommodate these cases by leveraging the *conditional* log-normality of the equilibrium distributions, as in the work of Dew-Becker (2014) and Bianchi et al. (2022). These extensions will allow us to incorporate asset pricing, breaks in the transmission mechanisms of the shocks, and changes in volatility of the macroeconomy in general equilibrium models featuring DE.

References

Akerlof, George A, "Procrastination and obedience," *American Economic Review*, 1991, *81* (2), 1–19.

Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Lindé, "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, April 2011, *14* (2), 225–247.

Bianchi, Francesco, Howard Kung, and Mikhail Tirskikh, "The origins and effects of macroeconomic uncertainty," Technical Report, National Bureau of Economic Research 2022.

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer, "Memory and probability," 2021. Harvard, Working Paper.

_ , Katherine Coffman, Nicola Gennaioli, Frederik Schwerter, and Andrei Shleifer, "Memory and representativeness," *Psychological Review*, 2020. forthcoming.

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

_ , _ , _ , and Stephen J Terry, "Real credit cycles," 2019. Harvard, Working Paper.

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

_ , _ , Yueran Ma, and Andrei Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–82.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 2005, *113* (1), 1–45.

Christiano, Lawrence J., Mathias Trabandt, and Karel Walentin, "Involuntary Unemployment and the Business Cycle," 2010. NBER Working Paper.

_ , _ , and Karl Walentin, "DSGE Models for Monetary Policy Analysis," in Benjamin M. Friedman and Michael Woodford, eds., *Handbook of Monetary Economics*, Vol. 3A, Elsevier, 2011, chapter 7, pp. 285–367.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silva, "Innocent Bystanders? Monetary Policy and Inequality," *Journal of Monetary Economics*, 2017, *88*, 70–89.

d'Arienzo, Daniele, "Maturity increasing overreaction and bond market puzzles," 2020. Bocconi University, Working Paper.

Dew-Becker, Ian, "Bond pricing with a time-varying price of risk in an estimated medium-scale Bayesian DSGE model," *Journal of Money, Credit and Banking*, 2014, *46* (5), 837–888.

Fagereng, Andreas, Martin Blomhoff Holm, and Gisle James James Natvik, "MPC heterogeneity and household balance sheets," *American Economic Journal: Macroeconomics*, 2020. forthcoming.

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

Geweke, John, "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 1999, *18*, 1–126.

He, Zhiguo and Arvind Krishnamurthy, "A macroeconomic framework for quantifying systemic risk," *American Economic Journal: Macroeconomics*, 2019, *11* (4), 1–37.

Ilut, Cosmin L and Rosen Valchev, "Economic agents as imperfect problem solvers," 2020. NBER WP 27820.

Inoue, Atsushi and Mototsugu Shintani, "Quasi-Bayesian Model Selection," *Quantitative Economics*, 2018, *9*, 1265–1297.

Jappelli, Tullio and Luigi Pistaferri, "The consumption response to income changes," *Annual Review of Economics*, 2010, *2*, 479–506.

Jordà, Òscar, "Estimation and inference of impulse responses by local projections," *American economic review*, 2005, *95* (1), 161–182.

Kahneman, Daniel, *Thinking, fast and slow*, Macmillan, 2011.

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

Kim, Jinill, "Constructing and estimating a realistic optimizing model of monetary policy," *Journal of Monetary Economics*, 2000, *45* (2), 329–359.

Kueng, Lorenz, "Excess sensitivity of high-income consumers," *The Quarterly Journal of Economics*, 2018, *133* (4), 1693–1751.

Laibson, David, "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–478.

L'Huillier, Jean-Paul, Sanjay R Singh, and Donghoon Yoo, "Incorporating diagnostic expectations into the New Keynesian framework," 2021. UC Davis, mimeo.

Lian, Chen, "Mistakes in future consumption, high MPCs now," 2020. MIT, Working Paper.

Maxted, Peter, "A macro-finance model with sentiment," 2020. Harvard, mimeo.

McDowall, Robert A, "Consumption behavior across the distribution of liquid assets," 2020. Working paper, NYU.

McKay, Alisdair and Johannes F. Wieland, "Lumpy durable consumption demand and the limited ammunition of monetary policy," *Econometrica*, 2021. forthcoming.

Minsky, Hyman P, "The financial instability hypothesis: An interpretation of Keynes and an alternative to "standard" theory," *Challenge*, 1977, *20* (1), 20–27.

Newey, Whitney K. and Kenneth D. West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, May 1987, *55* (3), 703–708.

O'Donoghue, Ted and Matthew Rabin, "Doing it now or later," *American Economic Review*, 1999, *89* (1), 103–124.

Pollak, Robert A, "Consistent planning," *The Review of Economic Studies*, 1968, *35* (2), 201–208.

Rabin, Matthew, "An approach to incorporating psychology into economics," *American Economic Review*, 2013, *103* (3), 617–22.

Romer, Christina D. and David H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 1055–1084.

Rotemberg, Julio J, "Monopolistic price adjustment and aggregate output," *The Review of Economic Studies*, 1982, *49* (4), 517–531.

Sims, Christopher A., "Solving linear rational expectations models," *Computational Economics*, 2000, *20* (1–2), 1–20.

Smets, F. and R. Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 2007, *97* (3), 586–606.

Strotz, Robert Henry, "Myopia and inconsistency in dynamic utility maximization," *The Review of Economic Studies*, 1955, *23* (3), 165–180.

Tversky, Amos and Daniel Kahneman, "Judgment under uncertainty: Heuristics and Biases," in "Utility, probability, and human decision making" 1975, pp. 141–162.

Online Appendix for "Diagnostic Business Cycles"

Francesco Bianchi, Cosmin Ilut and Hikaru Saijo

A An Investment Model and Endogenous Predictability

In this Appendix we study a two-period investment model that illustrates the point we made in footnote 6 regarding the endogenous predictability.

We start with the problem under RE.

$$\begin{aligned} & \max_{I_1} D_1 + \frac{1}{1+r}\mathbb{E}\left[D_2\right] \\ D_1 &= A_1 K_1^{\nu} - I_1 \\ D_2 &= A_2 K_2^{\nu} + (1-\delta) K_2 \\ K_2 &= (1-\delta) K_1 + I_1 \end{aligned}$$

where $r > 0$, $0 < \delta < 1$, $0 < \nu < 1$, and K_1 are taken as given. A_t is an i.i.d. Normal process. We can replace the constraints in the object function and derive the FOC:

$$\frac{1}{1+r}\mathbb{E}\left[\nu A_2 \left((1-\delta) K_1 + I_1^*\right)^{\nu-1} + (1-\delta)\right] = 1$$

We obtain:

$$I_1^* = \left[\frac{R+\delta}{\nu \mathbb{E}\left[A_2\right]}\right]^{\frac{1}{\nu-1}} - (1-\delta) K_1$$

Thus, A_1 is irrelevant as long as dividends are allowed to be negative.

In turn, under DE:

$$\begin{aligned} & \max_{I_1} D_1 + \frac{1}{1+r}\mathbb{E}^{\theta}\left[D_2\right] \\ D_1 &= A_1 K_1^{\nu} - I_1 \\ D_2 &= A_2 K_2^{\nu} + (1-\delta) K_2 \\ K_2 &= (1-\delta) K_1 + I_1 \end{aligned}$$

We can replace the constraints and compute the FOC:

$$\frac{1}{1+r}\mathbb{E}^{\theta}\left[\nu A_2 \left((1-\delta) K_1 + I_1^*\right)^{\nu-1} + (1-\delta)\right] = 1$$

In equilibrium, under the assumption of the model, the optimal investment choice is not stochastic. Thus, despite the non-linearity of the problem, normality is preserved. Since the product of a normal times a constant is still a normal, we get:

$$(1+\theta)\,\nu \left((1-\delta) K_1 + I_1^*\right)^{\nu-1} \mathbb{E}\left[A_2\right] - \theta\nu \left((1-\delta) K_1 + I_1^*\right)^{\nu-1} \mathbb{E}_{-1}\left[A_2\right] = (1+r) - (1-\delta)$$

Given that TFP is i.i.d., we have $\mathbb{E}[A_2] = \mathbb{E}_{-1}[A_2]$:

$$\nu\left((1-\delta)K_1 + I_1\right)^{\nu-1}\mathbb{E}[A_2] = (1+r) - (1-\delta)$$

Then

$$I_1^* = \left[\frac{r+\delta}{\nu\mathbb{E}[A_2]}\right]^{\frac{1}{\nu-1}} - (1-\delta)K_1$$

Thus, the solution is identical to the RE solution. This is because there is no revision in expectations coming from what happens at time 1.

B Omitted Proofs

B.1 Proof of Lemma 1

Proof.

$$\begin{aligned}
\mathbb{E}_t^\theta\left[\mathbb{E}_{t+1}^\theta[C_{t+1+n}]\right] &= \mathbb{E}_t^\theta\left[\mathbb{E}_{t+1}[C_{t+1+n}] + \theta\left(\mathbb{E}_{t+1}[C_{t+1+n}] - \mathbb{E}_{t+1-J}[C_{t+1+n}]\right)\right] \\
&= \mathbb{E}_t\left[\mathbb{E}_{t+1}[C_{t+1+n}] + \theta\left(\mathbb{E}_{t+1}[C_{t+1+n}] - \mathbb{E}_{t+1-J}[C_{t+1+n}]\right)\right] \\
&+ \theta\left\{\mathbb{E}_t\left[\mathbb{E}_{t+1}[C_{t+1+n}] + \theta\left(\mathbb{E}_{t+1}[C_{t+1+n}] - \mathbb{E}_{t+1-J}[C_{t+1+n}]\right)\right]\right. \\
&\left.-\mathbb{E}_{t-J}\left[\mathbb{E}_{t+1}[C_{t+1+n}] + \theta\left(\mathbb{E}_{t+1}[C_{t+1+n}] - \mathbb{E}_{t+1-J}[C_{t+1+n}]\right)\right]\right\} \\
&= \mathbb{E}_t[C_{t+1+n}] + \theta\left(\mathbb{E}_t[C_{t+1+n}] - \mathbb{E}_{t-J}[C_{t+1+n}]\right) \\
&+ \theta(1+\theta)\left(\mathbb{E}_t[C_{t+1+n}] - \mathbb{E}_t\mathbb{E}_{t+1-J}[C_{t+1+n}]\right)
\end{aligned}$$

The term $\left(\mathbb{E}_t[C_{t+1+n}] - \mathbb{E}_t\mathbb{E}_{t+1-J}[C_{t+1+n}]\right)$ in the last line is generically zero if and only if $J = 1$. Thus, $\mathbb{E}_t^\theta\left[\mathbb{E}_{t+1}^\theta[C_{t+1+n}]\right] = \mathbb{E}_t^\theta[C_{t+1+n}]$ if and only if $J = 1$. □

B.2 Proof of Proposition 1

Proof. The first order conditions at time 1 are:

$$\begin{aligned}
C_1 &= \mathbb{E}_1[C_2] \\
\mathbb{E}_1[C_2] &= \mathbb{E}_1[C_3]
\end{aligned}$$

or, equivalently

$$\begin{aligned}
Y_1 + K_0 - K_1 &= \mathbb{E}_1[Y_2 + K_1 - K_2] \\
\mathbb{E}_1[Y_2 + K_1 - K_2] &= \mathbb{E}_1[Y_3 + K_2]
\end{aligned}$$

The solution at time 1 and 2 can be obtained with backward induction or a guess-and-verify approach. We opt for the guess-and-verify approach because since it is easy to generalize for the infinite horizon case. We then guess that the solution assumes the following form:

$$K_1 = \alpha_1^{RE}\left(K_0 + \varepsilon_1\right); K_2 = \alpha_2^{RE}\left(K_1 + \varepsilon_2\right)$$

We then have:

$$K_0 + \varepsilon_1(1-\alpha_1^{RE}) = \mathbb{E}_1\left[\varepsilon_2 + \alpha_1^{RE}\left(K_0+\varepsilon_1\right) - \alpha_2^{RE}\left(\alpha_1^{RE}\left(K_0+\varepsilon_1\right)+\varepsilon_2\right)\right]$$
$$\mathbb{E}_1\left[\varepsilon_2 + \alpha_1^{RE}\left(K_0+\varepsilon_1\right) - \alpha_2^{RE}\left(\alpha_1^{RE}\left(K_0+\varepsilon_1\right)+\varepsilon_2\right)\right] = \mathbb{E}_1\left[\varepsilon_3 + \alpha_2^{RE}\left(\alpha_1^{RE}\left(K_0+\varepsilon_1\right)+\varepsilon_2\right)\right]$$

Equating coefficients, we get

$$\alpha_1^{RE} = \frac{2}{3}, \quad \alpha_2^{RE} = \frac{1}{2}$$

It is immediate to verify that the solution is time-consistent. The agent at time 2 solves the problem (9). The first order condition at time 2 is:

$$C_2 = \mathbb{E}_2\left[C_3\right],$$

or, equivalently

$$\varepsilon_2 + K_1 - K_2 = \mathbb{E}_2\left[\varepsilon_3 + K_2\right].$$

We obtain:

$$K_2 = \frac{1}{2}\left[K_1 + \varepsilon_2\right] = \alpha_2^{RE}\left[K_1+\varepsilon_2\right].$$

□

B.3 Proof of Proposition 2 and 3

Proof. For the time 1 problem, we conjecture the planned policy:

$$K_1^{\theta,p} = \alpha_{K_0}^{\theta,p} K_0 + \alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1;\ K_2^{\theta,p} = \alpha_{K_1}^{\theta,p} K_1^{\theta,p} + \alpha_{\varepsilon_2}^{\theta,p}\varepsilon_2.$$

We have two first-order conditions:

$$Y_1 + K_0 - K_1^{\theta,p} = \mathbb{E}_1^{\theta}\left[Y_2 + K_1^{\theta,p} - K_2^{\theta,p}\right] \tag{45}$$
$$\mathbb{E}_1^{\theta}\left[Y_2 + K_1^{\theta,p} - K_2^{\theta,p}\right] = \mathbb{E}_1^{\theta}\left[Y_3 + K_2^{\theta,p}\right] \tag{46}$$

We first solve for the planned policy for period 2 by plugging in the conjecture into (46):

$$\mathbb{E}_1^{\theta}\left[K_1^{\theta,p} - \left(\alpha_{K_1}^{\theta,p} K_1^{\theta,p} + \alpha_{\varepsilon_2}^{\theta,p}\varepsilon_2\right)\right] = \mathbb{E}_1^{\theta}\left[\alpha_{K_1}^{\theta,p} K_1^{\theta,p} + \alpha_{\varepsilon_2}^{\theta,p}\varepsilon_2\right]$$
$$\mathbb{E}_1^{\theta}\left[\left(1-2\alpha_{K_1}^{\theta,p}\right) K_1^{\theta,p} - 2\alpha_{\varepsilon_2}^{\theta,p}\varepsilon_2\right] = 0$$

Then:

$$(1+\theta)\left(1-2\alpha_{K_1}^{\theta,p}\right) K_1^{\theta,p} - \theta\mathbb{E}_0\left[\left(1-2\alpha_{K_1}^{\theta,p}\right) K_1^{\theta,p}\right] = 0$$
$$\left(1-2\alpha_{K_1}^{\theta,p}\right)\left[(1+\theta)\,\alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1 + \alpha_{K_0}^{\theta,p} K_0\right] = 0$$

Hence we have:

$$\alpha_{K_1}^{\theta,p} = \frac{1}{2}$$

Endowed with the contingent plan for time 2, we can then solve for the time 1 problem:

$$Y_1 + K_0 - K_1^{\theta,p} = \mathbb{E}_1^\theta \left[Y_2 + K_1^{\theta,p} - K_2^{\theta,p} \right]$$
$$\varepsilon_1 + K_0 - K_1^{\theta,p} = (1+\theta)\, \mathbb{E}_1 \left[K_1^{\theta,p} - \frac{1}{2} K_1^{\theta,p} \right] - \theta \mathbb{E}_0 \left[K_1^{\theta,p} - \frac{1}{2} K_1^{\theta,p} \right]$$
$$\varepsilon_1 + K_0 - K_1^{\theta,p} = (1+\theta)\, \frac{1}{2} \left[\alpha_{K_0}^{\theta,p} K_0 + \alpha_{\varepsilon_1}^{\theta,p} \varepsilon_1 \right] - \theta \frac{1}{2} \alpha_{K_0}^{\theta,p} K_0$$

We get:

$$K_1^{\theta,p} = -(1+\theta)\, \frac{1}{2} \left[\alpha_{K_0}^{\theta,p} K_0 + \alpha_{\varepsilon_1}^{\theta,p} \varepsilon_1 \right] + \theta \frac{1}{2} \alpha_{K_0}^{\theta,p} K_0 + \varepsilon_1 + K_0$$
$$K_1^{\theta,p} = -\frac{1}{2} \left[(1+\theta)\, \alpha_{\varepsilon_1}^{\theta,p} - 2 \right] \varepsilon_1 + \left[1 - \frac{1}{2} \alpha_{K_0}^{\theta,p} \right] K_0$$

Matching coefficients:

$$\alpha_{K_0}^{\theta,p} = \frac{2}{3}, \quad \alpha_{\varepsilon_1}^{\theta,p} = \frac{2}{3+\theta}$$

Note that when $J = 1$, there is no contingent plan formed at time 1 on how to react to ε_2, given that ε_2 does not impact utility at time 1. It is then immediate to verify that the plan is time consistent. The agent at time 2 inherits the capital K_1^θ and solves (11). The first order condition at time 2 is:

$$C_2^\theta = \mathbb{E}_2^\theta \left[C_3^\theta \right].$$

We conjecture the solution $K_2^\theta(K_1^\theta, \varepsilon_2) = \alpha_{K_1}^\theta K_1^\theta + \alpha_{\varepsilon_2}^\theta \varepsilon_2$. Then:

$$Y_2 + K_1^\theta - K_2^\theta(K_1^\theta, \varepsilon_2) = \mathbb{E}_2 \left[Y_3 + K_2^\theta(K_1^\theta, \varepsilon_2) \right] + \theta \left[\mathbb{E}_2 \left(Y_3 + K_2^\theta(K_1^\theta, \varepsilon_2) \right) - \mathbb{E}_1 \left(Y_3 + K_2^\theta(K_1^\theta, \varepsilon_2) \right) \right]$$
$$\varepsilon_2 + K_1^\theta - K_2^\theta(K_1^\theta, \varepsilon_2) = \alpha_{K_1}^\theta K_1^\theta + \alpha_{\varepsilon_2}^\theta \varepsilon_2 + \theta \alpha_{\varepsilon_2}^\theta \varepsilon_2$$

We have:

$$K_2^\theta(K_1^\theta, \varepsilon_2) = (1 - \alpha_{K_1}^\theta) K_1^\theta + (1 - \alpha_{\varepsilon_2}^\theta - \theta \alpha_{\varepsilon_2}^\theta) \varepsilon_2$$

Matching coefficients, we obtain:

$$\alpha_{K_1}^\theta = \frac{1}{2} = \alpha_{K_1}^{\theta,p}, \quad \alpha_{\varepsilon_2}^\theta = \frac{1}{2+\theta}$$

□

B.4 Proof of Proposition 4

Proof. We first solve the planning problem at time 1. We conjecture the solution:

$$K_1^{\theta,p} = \alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p} \mathbb{N}_{-1,0}\left[K_0\right] + \alpha_{K_0}^{\theta,p} K_0 + \alpha_{\varepsilon_1}^{\theta,p} \varepsilon_1; \; K_2^{\theta,p} = \alpha_{K_1}^{\theta,p} K_1^{\theta,p} + \alpha_{\varepsilon_2}^{\theta,p} \varepsilon_2.$$

where $\mathbb{N}_{-1,0}\left[K_0\right] \equiv K_0 - \mathbb{E}_{-1}\left[K_0\right]$ represents the surprise in the stock of capital with respect to the expectations formed in the past. Note that $\mathbb{E}_{-1}\left[\mathbb{N}_{-1,0}\left[K_0\right]\right] = 0$. We have two first-order

conditions:

$$Y_1 + K_0 - K_1^{\theta,p} = \mathbb{E}_1^{\theta}\left[Y_2 + K_1^{\theta,p} - K_2^{\theta,p}\right] \tag{47}$$

$$\mathbb{E}_1^{\theta}\left[Y_2 + K_1^{\theta,p} - K_2^{\theta,p}\right] = \mathbb{E}_1^{\theta}\left[Y_3 + K_2^{\theta,p}\right] \tag{48}$$

We first solve for the planned policy for period 2 by plugging in the conjecture into (48):

$$\mathbb{E}_1^{\theta}\left[K_1^{\theta,p} - K_2^{\theta,p}\right] = \mathbb{E}_1^{\theta}\left[K_2^{\theta,p}\right]$$

$$\mathbb{E}_1^{\theta}\left[\left(1 - 2\alpha_{K_1}^{\theta,p}\right)K_1^{\theta,p} - 2\alpha_{\varepsilon_2}^{\theta,p}\varepsilon_2\right] = 0$$

Then:

$$(1+\theta)\left(1 - 2\alpha_{K_1}^{\theta,p}\right)K_1^{\theta,p} - \theta\mathbb{E}_{-1}\left[\left(1 - 2\alpha_{K_1}^{\theta,p}\right)K_1^{\theta,p}\right] = 0$$

$$\left(1 - 2\alpha_{K_1}^{\theta,p}\right)\left[(1+\theta)\left(\alpha_{K_0}^{\theta,p}K_0 + \alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1\right) - \theta\mathbb{E}_{-1}\alpha_{K_0}^{\theta,p}K_0\right] = 0$$

Hence we have:

$$\alpha_{K_1}^{\theta,p} = \frac{1}{2}$$

Note that the *planned* solution for time 2 is identical to the case in which $J = 1$. This is because distant memory does not affect how the agent evaluates the trade-off between consumption at time 2 and consumption at time 3 from the point of view of time 1.

Endowed with the contingent plan for time 2, we can then solve for the time 1 problem:

$$\varepsilon_1 + K_0 - K_1^{\theta,p} = (1+\theta)\,\mathbb{E}_1\left[K_1^{\theta,p} - \frac{1}{2}K_1^{\theta,p}\right] - \theta\mathbb{E}_{-1}\left[K_1^{\theta,p} - \frac{1}{2}K_1^{\theta,p}\right]$$

$$\varepsilon_1 + K_0 - K_1^{\theta,p} = (1+\theta)\,\frac{1}{2}\left[\alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p}\mathbb{N}_{-1,0}\left[K_0\right] + \alpha_{K_0}^{\theta,p}K_0 + \alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1\right] - \theta\frac{1}{2}\alpha_{K_0}^{\theta,p}\mathbb{E}_{-1}\left[K_0\right]$$

$$\varepsilon_1 + K_0 - K_1^{\theta,p} = (1+\theta)\,\frac{1}{2}\left[\alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p}\mathbb{N}_{-1,0}\left[K_0\right] + \alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1\right] + \frac{1}{2}\alpha_{K_0}^{\theta,p}K_0 + \theta\frac{1}{2}\alpha_{K_0}^{\theta,p}\mathbb{N}_{-1,0}\left[K_0\right]$$

$$\varepsilon_1 + K_0 - K_1^{\theta,p} = \frac{1}{2}\left[(1+\theta)\,\alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p} + \theta\alpha_{K_0}^{\theta,p}\right]\mathbb{N}_{-1,0}\left[K_0\right] + (1+\theta)\,\frac{1}{2}\alpha_{\varepsilon_1}^{\theta,p}\varepsilon_1 + \frac{1}{2}\alpha_{K_0}^{\theta,p}K_0$$

Then:

$$K_1^{\theta,p} = -\frac{1}{2}\left[(1+\theta)\,\alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p} + \theta\alpha_{K_0}^{\theta,p}\right]\mathbb{N}_{-1,0}\left[K_0\right] + \left[1 - \frac{1}{2}\alpha_{K_0}^{\theta,p}\right]K_0 - \frac{1}{2}\left[(1+\theta)\,\alpha_{\varepsilon_1}^{\theta,p} - 2\right]\varepsilon_1$$

Hence we have

$$\alpha_{K_0}^{\theta,p} = \frac{2}{3}, \quad \alpha_{\varepsilon_1}^{\theta,p} = \frac{2}{3+\theta}$$

and

$$\alpha_{\mathbb{N}_{-1,0}[K_0]}^{\theta,p} = -\frac{\theta}{3+\theta}\alpha_{K_0}^{\theta,p} = -\frac{2\theta}{3\,(3+\theta)}$$

Note that even when $J > 1$, there is no contingent plan formed at time 1 on how to react to ε_2, given that this does not impact utility at time 1. However, with respect to the case of recent memory, now we have an additional state variable that depends on the news

component of the inherited capital, $\mathbb{N}_{-1,0}\left[K_0\right]$. For a given level of inherited capital K_0, the larger the surprise, the lower the amount saved at time 1. This distortion increases with θ.

We now verify that the plan made at time 1 for time 2 is time inconsistent. The agent at time 2 inherits the capital K_1^θ and solves (11). The first order condition at time 2 is:

$$C_2^\theta = \mathbb{E}_2^\theta\left[C_3^\theta\right].$$

We conjecture the solution $K_2^\theta = \alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2$. Then:

$$\begin{aligned}
Y_2 + K_1^\theta - K_2^\theta &= \mathbb{E}_2\left[Y_3 + K_2^\theta\right] + \theta\left[\mathbb{E}_2\left[Y_3 + K_2^\theta\right] - \mathbb{E}_0\left[Y_3 + K_2^\theta\right]\right] \\
\varepsilon_2 + K_1^\theta - K_2^\theta &= \mathbb{E}_2\left[\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2\right] \\
&\quad + \theta\begin{bmatrix} \mathbb{E}_2\left[\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2\right] \\ -\mathbb{E}_0\left[\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2\right] \end{bmatrix}
\end{aligned}$$

We get:

$$\begin{aligned}
\varepsilon_2 + K_1^\theta - K_2^\theta &= \alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^{\theta,p}\right]}\mathbb{N}_{0,1}\left[K_1^{\theta,p}\right] + \alpha^\theta_{K_1}K_1^{\theta,p} + \alpha^\theta_{\varepsilon_2}\varepsilon_2 \\
&\quad + \theta\left[\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^{\theta,p}\right]}\mathbb{N}_{0,1}\left[K_1^{\theta,p}\right] + \alpha^\theta_{K_1}K_1^{\theta,p} + \alpha^\theta_{\varepsilon_2}\varepsilon_2 - \mathbb{E}_0\alpha^\theta_{K_1}K_1^{\theta,p}\right]
\end{aligned}$$

Then:

$$\varepsilon_2 + K_1^\theta - K_2^\theta = \left[(1+\theta)\,\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]} + \theta\alpha^\theta_{K_1}\right]\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + (1+\theta)\,\alpha^\theta_{\varepsilon_2}\varepsilon_2$$

Rearrange:

$$K_2^\theta = -\left[(1+\theta)\,\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]} + \theta\alpha^\theta_{K_1}\right]\mathbb{N}_{0,1}\left[K_1^\theta\right] + \left(1-\alpha^\theta_{K_1}\right)K_1^\theta - \left[(1+\theta)\,\alpha^\theta_{\varepsilon_2} - 1\right]\varepsilon_2$$

Matching coefficients, we obtain:

$$\alpha^\theta_{K_1} = \frac{1}{2}, \quad \alpha^\theta_{\varepsilon_2} = \frac{1}{2+\theta}$$

and

$$\alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]} = -\frac{\theta}{2+\theta}\alpha^\theta_{K_1} = -\frac{\theta}{2\,(2+\theta)}.$$

The revised time 2 policy can then be rewritten as

$$K_2^\theta = \frac{\theta}{2(2+\theta)}\mathbb{E}_0\left[K_1^\theta\right] + \frac{1}{2+\theta}K_1^\theta + \frac{1}{2+\theta}\varepsilon_2,$$

and so the coefficient on K_1^θ is not equal to that of the time 1 plan (which is 0.5). □

The time inconsistency arises because of the information content of $K_1^{\theta,p}$ with respect to the capital expected at time zero. Between when reference expectations were formed, at time 0, and when a new decision is made, at time 2, an income shock occurred and agents reacted

to the shock. As a result, capital is not what the agent expected it to be. Agents do not take into account this surprise in capital when they solve the planning problem at time 1.

B.5 Proof of Proposition 5

Proof. To obtain the time 1 policy function, we consider the conjecture

$$K_1^\theta = \alpha^\theta_{\mathbb{N}_{-1,0}[K_0]}\mathbb{N}_{-1,0}\left[K_0\right] + \alpha^\theta_{K_0}K_0 + \alpha^\theta_{\varepsilon_1}\varepsilon_1.$$

The time 1 trade-off is given by

$$C_1^\theta = \mathbb{E}_1^\theta\left[C_2^{RE}\right].$$

The right hand side equals

$$\begin{aligned}
\mathbb{E}_1^\theta\left[C_2^{RE}\right] &= (1+\theta)\mathbb{E}_1\left[Y_2 + K_1^\theta - K_2^{RE}\right] - \theta\mathbb{E}_{-1}\left[Y_2 + K_1^{RE} - K_2^{RE}\right] \\
&= (1+\theta)\mathbb{E}_1\left[\overline{Y} + \varepsilon_2(1-\alpha^{RE}_{\varepsilon_2}) + K_1^\theta(1-\alpha^{RE}_{K_1})\right] - \theta\mathbb{E}_{-1}\left[\overline{Y} + \varepsilon_2(1-\alpha^{RE}_{\varepsilon_2}) + K_1^{RE}(1-\alpha^{RE}_{K_1})\right] \\
&= \overline{Y} + (1-\alpha^{RE}_{K_1})\left[(1+\theta)K_1^\theta - \theta\mathbb{E}_{-1}\left[K_1^{RE}\right]\right] \\
&= \overline{Y} + \frac{1}{2}\left[(1+\theta)K_1^\theta - \frac{2}{3}\theta\mathbb{E}_{-1}\left[K_0\right]\right]
\end{aligned}$$

where we have substituted in the RE policy $K_2^{RE} = \alpha^{RE}_{K_1}K_1 + \alpha^{RE}_{\varepsilon_2}\varepsilon_2$ in the second line and substituted in $\alpha^{RE}_{K_1} = 1/2$ and $\alpha^{RE}_{K_0} = 2/3$ in the fourth line. Connecting this with the left hand side, we have

$$\varepsilon_1 + K_0 - K_1^\theta = \frac{1}{2}\left[(1+\theta)K_1^\theta - \frac{2}{3}\theta\mathbb{E}_{-1}\left[K_0\right]\right].$$

Plugging in the conjectured solution $K_1^\theta = \alpha^\theta_{\mathbb{N}_{-1,0}[K_0]}\mathbb{N}_{-1,0}\left[K_0\right] + \alpha^\theta_{K_0}K_0 + \alpha^\theta_{\varepsilon_1}\varepsilon_1$ and equating coefficients give us

$$\alpha^\theta_{\mathbb{N}_{-1,0}[K_0]} = -\frac{2\theta}{3(3+\theta)}, \quad \alpha^\theta_{K_0} = \frac{2}{3}, \quad \alpha^\theta_{\varepsilon_1} = \frac{2}{3+\theta}.$$

To obtain the time 2 policy function, we consider the conjecture

$$K_2^\theta = \alpha^\theta_{\mathbb{N}_{0,1}\left[K_1^\theta\right]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2.$$

The time 2 trade-off is given by

$$C_2^\theta = \mathbb{E}_2^\theta\left[C_3^{RE}\right].$$

The right hand side equals

$$\begin{aligned}
\mathbb{E}_2^\theta\left[C_3^{RE}\right] &= (1+\theta)\mathbb{E}_2\left[Y_3 + K_2^\theta\right] - \theta\mathbb{E}_0\left[Y_3 + K_2^{RE}\right] \\
&= \overline{Y} + (1+\theta)K_2^\theta - \theta\mathbb{E}_0\left[K_2^{RE}\right] \\
&= \overline{Y} + (1+\theta)K_2^\theta - \frac{1}{2}\theta\mathbb{E}_0\left[K_1^{RE}\right]
\end{aligned}$$

where we substituted in $\alpha_{K_1}^{RE} = 1/2$. Connecting this with the left hand side, we have

$$\varepsilon_2 + K_1 - K_2^\theta = (1+\theta)K_2^\theta - \frac{1}{2}\theta\mathbb{E}_0 K_1^{RE}.$$

Plugging in the conjectured solution $K_2^\theta = \alpha^\theta_{\mathbb{N}_{0,1}[K_1^\theta]}\mathbb{N}_{0,1}\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2$ and equating coefficients give us

$$\alpha^\theta_{\mathbb{N}_{0,1}[K_1^\theta]} = -\frac{\theta}{2(2+\theta)}, \quad \alpha^\theta_{K_1} = \frac{1}{2}, \quad \alpha^\theta_{\varepsilon_2} = \frac{1}{2+\theta}.$$

□

B.6 Proof of Proposition 7

The proposition below considers naïveté and sophistication under $J = 1$. The naïveté problem was described in the main text. Entering period 2, the sophisticated agent's problem is

$$\max_{K_2^\theta}\left[u(C_2^\theta) + \mathbb{E}_2^\theta u(C_3^\theta)\right] \tag{49}$$

where now

$$C_2^\theta = Y_2 + K_1^\theta - K_2^\theta; C_3^\theta = Y_3 + K_2^\theta - K_3^\theta. \tag{50}$$

Sophistication means that at time 1 the agent understands that her future action is dictated by equation (49) (as well as $K_3^\theta = 0$). Thus, the sophisticated agent solves

$$\max_{K_1^\theta}\left\{u(C_1^\theta) + \mathbb{E}_1^\theta\left[u(C_2^\theta) + u(C_3^\theta)\right]\right\}, \tag{51}$$

where $C_1^\theta = Y_1 + K_0 - K_1^\theta$, while C_2^θ and C_3^θ are determined as in (50). We assume that the comparison group for K_2^θ is $\mathbb{E}_{2-J}K_2^\theta$, i.e. the conditional expectation of the DE savings choice at time 2 made J periods ago by the former sophisticated self, under the true density.

Proposition 7. *When $J = 1$, the naïveté and sophistication policy functions are the same and recover the DE optimal choices based on time-consistency.*

Proof. **Policies under naïveté.** Conjecture

$$K_1^\theta = \alpha^\theta_{K_0}K_0 + \alpha^\theta_{\varepsilon_1}\varepsilon_1; \quad K_2^\theta = \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2.$$

The time 2 trade-off is given by

$$C_2^\theta = \mathbb{E}_2^\theta\left[C_3^{RE}\right]$$

The right hand side equals

$$\begin{aligned}\mathbb{E}_2^\theta\left[C_3^{RE}\right] &= (1+\theta)\mathbb{E}_2\left[Y_3+K_2^\theta\right]-\theta\mathbb{E}_1\left[Y_3+K_2^{RE}\right]\\ &= \overline{Y}+(1+\theta)K_2^\theta-\theta\mathbb{E}_1\left[K_2^{RE}\right]\\ &= \overline{Y}+(1+\theta)K_2^\theta-\frac{1}{2}\theta K_1,\end{aligned}$$

where we substituted in $\alpha_{K_1}^{RE}=1/2$ in the third line. Connecting this with the left hand side, we have

$$\varepsilon_2+K_1^\theta-K_2^\theta=(1+\theta)K_2^\theta-\frac{1}{2}\theta K_1^\theta.$$

Plugging in the conjectured solution $K_2^\theta=\alpha_{K_1}^\theta K_1^\theta+\alpha_{\varepsilon_2}^\theta\varepsilon_2$ and equating coefficients give us

$$\alpha_{K_1}^\theta=\frac{1}{2},\quad \alpha_{\varepsilon_2}^\theta=\frac{1}{2+\theta}.$$

By Lemma 2 the time 1 trade-off is given by

$$C_1^\theta=\mathbb{E}_1^\theta\left[C_2^{RE}\right].$$

The right hand side equals

$$\begin{aligned}\mathbb{E}_1^\theta\left[C_2^{RE}\right] &= (1+\theta)\mathbb{E}_1\left[Y_2+K_1^\theta-K_2^{RE}\right]-\theta\mathbb{E}_0\left[Y_2+K_1^{RE}-K_2^{RE}\right]\\ &= (1+\theta)\mathbb{E}_1\left[\overline{Y}+\varepsilon_2(1-\alpha_{\varepsilon_2}^{RE})+K_1^\theta(1-\alpha_{K_1}^{RE})\right]-\theta\mathbb{E}_0\left[\overline{Y}+\varepsilon_2(1-\alpha_{\varepsilon_2}^{RE})+K_1^{RE}(1-\alpha_{K_1}^{RE})\right]\\ &= \overline{Y}+(1-\alpha_{K_1}^{RE})\left[(1+\theta)K_1^\theta-\theta\mathbb{E}_0\left[K_1^{RE}\right]\right]\\ &= \overline{Y}+\frac{1}{2}\left[(1+\theta)K_1^\theta-\frac{2}{3}\theta K_0\right]\end{aligned}$$

where we have substituted in the RE policy $K_2^{RE}=\alpha_{K_1}^{RE}K_1^\theta+\alpha_{\varepsilon_2}^{RE}\varepsilon_2$ in the second line and substituted in $\alpha_{K_1}^{RE}=1/2$ and $\alpha_{K_0}^{RE}=2/3$ in the fourth line. Connecting this with the left hand side, we have

$$\varepsilon_1+K_0-K_1^\theta=\frac{1}{2}\left[(1+\theta)K_1^\theta-\frac{2}{3}\theta K_0\right].$$

Plugging in the conjectured solution $K_1^\theta=\alpha_{K_0}^\theta K_0+\alpha_{\varepsilon_1}^\theta\varepsilon_1$ and equating coefficients give us

$$\alpha_{K_0}^\theta=\frac{2}{3},\quad \alpha_{\varepsilon_1}^\theta=\frac{2}{3+\theta}.$$

Policies under sophistication. Conjecture

$$K_1^\theta=\alpha_{K_0}^{\theta,s}K_0+\alpha_{\varepsilon_1}^{\theta,s}\varepsilon_1;\quad K_2^\theta=\alpha_{K_1}^{\theta,s}K_1^\theta+\alpha_{\varepsilon_2}^{\theta,s}\varepsilon_2.$$

The time 2 trade-off is given by

$$C_2^\theta=\mathbb{E}_2^\theta\left[C_3^\theta\right]$$

The right hand side equals

$$\begin{aligned}\mathbb{E}_2^\theta C_3^\theta &= (1+\theta)\mathbb{E}_2\left[Y_3 + K_2^\theta\right] - \theta\mathbb{E}_1\left[Y_3 + K_2^\theta\right] \\ &= \overline{Y} + (1+\theta)K_2^\theta - \theta\mathbb{E}_1\left[K_2^\theta\right] \\ &= \overline{Y} + (1+\theta)K_2^\theta - \theta\alpha_{K_1}^{\theta,s}K_1^\theta.\end{aligned}$$

Connecting this with the left hand side, we have

$$\varepsilon_2 + K_1^\theta - K_2^\theta = (1+\theta)K_2^\theta - \theta\alpha_{K_1}^{\theta,s}K_1^\theta.$$

Plugging in the conjectured solution $K_2^\theta = \alpha_{K_1}^{\theta,s}K_1^\theta + \alpha_{\varepsilon_2}^{\theta,s}\varepsilon_2$ and equating coefficients give us

$$\alpha_{K_1}^{\theta,s} = \frac{1}{2}, \quad \alpha_{\varepsilon_2}^{\theta,s} = \frac{1}{2+\theta}.$$

The time 1 trade-off is given by

$$C_1^\theta = \mathbb{E}_1^\theta\left[C_2^\theta + \frac{\partial K_2^\theta}{\partial K_1^\theta}\left(C_3^\theta - C_2^\theta\right)\right].$$

but the indirect effect of current choice captured by the last term disappears under $J = 1$, so

$$C_1^\theta = \mathbb{E}_1^\theta\left[C_2^\theta\right].$$

The right hand side equals

$$\begin{aligned}\mathbb{E}_1^\theta\left[C_2^\theta\right] &= (1+\theta)\mathbb{E}_1\left[Y_2 + K_1^\theta - K_2^\theta\right] - \theta\mathbb{E}_0\left[Y_2 + K_1^\theta - K_2^\theta\right] \\ &= (1+\theta)\mathbb{E}_1\left[\overline{Y} + \varepsilon_2(1-\alpha_{\varepsilon_2}^{\theta,s}) + K_1^\theta(1-\alpha_{K_1}^{\theta,s})\right] - \theta\mathbb{E}_0\left[\overline{Y} + \varepsilon_2(1-\alpha_{\varepsilon_2}^{\theta,s}) + K_1^\theta(1-\alpha_{K_1}^{\theta,s})\right] \\ &= \overline{Y} + (1-\alpha_{K_1}^{\theta,s})\left[(1+\theta)K_1^\theta - \theta\mathbb{E}_0\left[K_1^\theta\right]\right] \\ &= \overline{Y} + \frac{1}{2}\left[(1+\theta)K_1^\theta - \alpha_{K_0}^{\theta,s}\theta K_0\right]\end{aligned}$$

where we have substituted in the DE policy $K_2^\theta = \alpha_{K_1}^{\theta,s}K_1 + \alpha_{\varepsilon_2}^{\theta,s}\varepsilon_2$ in the second line and substituted in $\alpha_{K_1}^{\theta,s} = 1/2$ in the fourth line. Connecting this with the left hand side, we have

$$\varepsilon_1 + K_0 - K_1^{\theta,s} = \frac{1}{2}\left[(1+\theta)K_1^\theta - \alpha_{K_0}^{\theta,s}\theta K_0\right].$$

Plugging in the conjectured solution $K_1^\theta = \alpha_{K_0}^{\theta,s}K_0 + \alpha_{\varepsilon_1}^{\theta,s}\varepsilon_1$ and equating coefficients give us

$$\alpha_{K_0}^{\theta,s} = \frac{2}{3}, \quad \alpha_{\varepsilon_1}^{\theta,s} = \frac{2}{3+\theta}.$$

□

B.7 Proof of Proposition 8

The Proposition below considers the solution to the three-period model under distant memory ($J = 2$) and sophistication.

Proposition 8. *When $J = 2$, under sophistication, the time 2 policy function is given by*

$$K_2^\theta = \alpha^{\theta,s}_{\mathbb{E}_0[K_1]}\mathbb{E}_0[K_1^\theta] + \alpha^{\theta,s}_{K_1}K_1^\theta + \alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2,$$

where the coefficients are identical to the naïveté case (modified to expressed in terms of coefficients $\alpha^\theta_{\mathbb{E}_0[K_1]}$, $\alpha^\theta_{K_1}$ and $\alpha^\theta_{\varepsilon_2}$) and are given by

$$\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} = \frac{\theta}{2(2+\theta)};\ \alpha^{\theta,s}_{K_1} = \frac{1}{2+\theta};\ \alpha^{\theta,s}_{\varepsilon_2} = \frac{1}{2+\theta}.$$

The time 1 policy function is given by

$$K_1^\theta = \alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\mathbb{E}_{-1}\left[K_0\right] + \alpha^{\theta,s}_{K_0}K_0 + \alpha^{\theta,s}_{\varepsilon_1}\varepsilon_1,$$

which compared to the naïveté policy function in Proposition 5 (modified to expressed in terms of coefficients $\alpha^\theta_{\mathbb{E}_{-1}[K_0]}$, $\alpha^\theta_{K_0}$ and $\alpha^\theta_{\varepsilon_1}$) is characterized by the following properties (1) $\alpha^{\theta,s}_{\varepsilon_1} < \alpha^\theta_{\varepsilon_1}$; (2) $\alpha^{\theta,s}_{K_0} < \alpha^\theta_{K_0}$ if $\theta < 1$, and $\alpha^{\theta,s}_{K_0} > \alpha^\theta_{K_0}$ if $\theta > 1$; (3), $\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} > \alpha^\theta_{\mathbb{E}_{-1}[K_0]}$ if $\theta < 1$, and $\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} < \alpha^\theta_{\mathbb{E}_{-1}[K_0]}$ if $\theta > 1$.

Proof. For the time 2 policy, consider the conjecture

$$K_2^\theta = \alpha^{\theta,s}_{\mathbb{E}_0[K_1]}\mathbb{E}_0\left[K_1^\theta\right] + \alpha^\theta_{K_1}K_1^\theta + \alpha^\theta_{\varepsilon_2}\varepsilon_2.$$

The time 2 trade-off is given by

$$C_2^\theta = \mathbb{E}_2^\theta\left[C_3^\theta\right]$$

The right hand side equals

$$\begin{aligned}\mathbb{E}_2^\theta\left[C_3^\theta\right] &= (1+\theta)\mathbb{E}_2\left[Y_3 + K_2^\theta\right] - \theta\mathbb{E}_0\left[Y_3 + K_2^\theta\right]\\ &= \overline{Y} + K_2^\theta + \theta\left[K_2^\theta - \mathbb{E}_0\left[K_2^\theta\right]\right]\\ &= \overline{Y} + \alpha^{\theta,s}_{\mathbb{E}_0[K_1]}\mathbb{E}_0\left[K_1^\theta\right] + \alpha^{\theta,s}_{K_1}K_1^\theta + \alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2 + \theta\left[\alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2 + \alpha^{\theta,s}_{K_1}\left(K_1^\theta - \mathbb{E}_0\left[K_1^\theta\right]\right)\right].\end{aligned}$$

Connecting this with the left hand side, we have

$$\varepsilon_2 + K_1^\theta - K_2^\theta = \alpha^{\theta,s}_{\mathbb{E}_0[K_1]}\mathbb{E}_0\left[K_1^\theta\right] + \alpha^{\theta,s}_{K_1}K_1^\theta + \alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2 + \theta\left[\alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2 + \alpha^{\theta,s}_{K_1}\left(K_1^\theta - \mathbb{E}_0\left[K_1^\theta\right]\right)\right].$$

Plugging in the conjectured solution $K_2^\theta = \alpha^{\theta,s}_{\mathbb{E}_0[K_1]}\mathbb{E}_0\left[K_1^\theta\right] + \alpha^{\theta,s}_{K_1}K_1^\theta + \alpha^{\theta,s}_{\varepsilon_2}\varepsilon_2$ and equating coefficients give us

$$\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} = \frac{1}{2(2+\theta)}, \quad \alpha^{\theta,s}_{K_1} = \frac{1}{2+\theta}, \quad \alpha^{\theta,s}_{\varepsilon_2} = \frac{1}{2+\theta}.$$

For the time 1 policy function, conjecture

$$K_1^\theta = \alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\mathbb{E}_{-1}\left[K_0\right] + \alpha^{\theta,s}_{K_0}K_0 + \alpha^{\theta,s}_{\varepsilon_1}\varepsilon_1.$$

The time 1 tradeoff is given by

$$C_1^\theta = \mathbb{E}_1^\theta\left[C_2^\theta + \alpha^{\theta,s}_{K_1}(\mathbb{E}_2\left[C_3^\theta\right] - C_2^\theta)\right], \tag{52}$$

where the term $\alpha^{\theta,s}_{K_1}(\mathbb{E}_2 C^\theta_3 - C^\theta_2)$ captures the fact that the sophisticated agent internalizes the fact that the current choice affects the future tradeoff. This term is zero when $J = 1$ because in that case the plan is time consistent. The right hand side equals

$$\begin{aligned}
&\mathbb{E}^\theta_1\left[C^\theta_2 + \alpha^{\theta,s}_{K_1}(\mathbb{E}_2 C^\theta_3 - C^\theta_2)\right] = (1-\alpha^{\theta,s}_{K_1})\mathbb{E}^\theta_1 C^\theta_2 + \alpha^{\theta,s}_{K_1}\mathbb{E}^\theta_1 C^\theta_3 \\
&= (1-\alpha^{\theta,s}_{K_1})\left\{(1+\theta)\mathbb{E}_1\left[Y_2 + K^\theta_1 - K^\theta_2\right] - \theta\mathbb{E}_{-1}\left[Y_2 + K^\theta_1 - K^\theta_2\right]\right\} \\
&+ \alpha^{\theta,s}_{K_1}\left\{(1+\theta)\mathbb{E}_1\left[Y_3 + K^\theta_2\right] - \theta\mathbb{E}_{-1}\left[Y_3 + K^\theta_2\right]\right\}
\end{aligned}$$

After some algebra, we find that this equals

$$\begin{aligned}
&= \overline{Y} + (1-\alpha^{\theta,s}_{K_1})(1+\theta)\left[\left(1-\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} - \alpha^{\theta,s}_{K_1}\right)\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\mathbb{E}_{-1}[K_0] + \alpha^{\theta,s}_{K_0}K_0\right) + (1-\alpha^{\theta,s}_{K_1})\alpha^{\theta,s}_{\varepsilon_1}\varepsilon_1\right] \\
&- (1-\alpha^{\theta,s}_{K_1})\theta\left(1-\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} - \alpha^{\theta,s}_{K_1}\right)\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} + \alpha^{\theta,s}_{K_0}\right)\mathbb{E}_{-1}[K_0] \\
&+ \alpha^{\theta,s}_{K_1}(1+\theta)\left[\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\mathbb{E}_{-1}[K_0] + \alpha^{\theta,s}_{K_0}K_0\right) + \alpha^{\theta,s}_{K_1}\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]}\mathbb{E}_{-1}[K_0] + \alpha^{\theta,s}_{K_0}K_0 + \alpha^{\theta,s}_{\varepsilon_1}\varepsilon_1\right)\right] \\
&- \alpha^{\theta,s}_{K_1}\theta\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} + \alpha^{\theta}_{K_1}\right)\left(\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} + \alpha^{\theta,s}_{K_0}\right)\mathbb{E}_{-1}[K_0]
\end{aligned}$$

The left hand side is given by

$$C^\theta_1 = \overline{Y} + \varepsilon_1 + K_0 - K^\theta_1.$$

We then connect the left hand side to the right hand side and equate coefficients after substituting in the conjectured solution for K^θ_1. Equating coefficients, we have

$$\begin{aligned}
\alpha^{\theta,s}_{\varepsilon_1} &= \frac{1}{1+(1+\theta)\left[(1-\alpha^{\theta,s}_{K_1})^2 + (\alpha^{\theta,s}_{K_1})^2\right]} = \frac{(2+\theta)^2}{(2+\theta)^2 + (1+\theta)\left[(1+\theta)^2+1\right]} \\
\alpha^{\theta,s}_{K_0} &= \frac{1}{1+(1+\theta)\left[(1-\alpha^{\theta,s}_{K_1})(1-\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} - \alpha^{\theta,s}_{K_1}) + \alpha^{\theta,s}_{K_1}(\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} + \alpha^{\theta,s}_{K_1})\right]} \\
&= \frac{2(2+\theta)^2}{2(2+\theta)^2 + (1+\theta)\left[(1+\theta)(1+2\theta)+3\right]} \\
\alpha^{\theta,s}_{\mathbb{E}_{-1}[K_0]} &= \frac{\theta\left[(1-\alpha^{\theta,s}_{K_1})(1-\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} - \alpha^{\theta,s}_{K_1}) + \alpha^{\theta,s}_{K_1}(\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} + \alpha^{\theta,s}_{K_1})\right]}{1+(1-\alpha^{\theta,s}_{K_1})(1-\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} - \alpha^{\theta,s}_{K_1}) + \alpha^{\theta,s}_{K_1}(\alpha^{\theta,s}_{\mathbb{E}_0[K_1]} + \alpha^{\theta,s}_{K_1})}\alpha^{\theta,s}_{K_0} \\
&= \frac{\theta\left[(1+2\theta)(1+\theta)+3\right]}{2(2+\theta)^2 + (1+2\theta)(1+\theta)+3}\alpha^{\theta,s}_{K_0}
\end{aligned}$$

which give the specific coefficients in Proposition 8. When we compare this sophistication solution to the naïveté one, we find the patterns stated in Proposition 8.

□

The solution for the sophisticated choice K^θ_2 follows the same logic as for naïveté choice, leading to the result in Proposition 8 that the optimal coefficients are the same. The subtle difference here is the comparison group formation. The naïveté solution can leverage the

law of motion for K_1^{RE}, so that $\mathbb{E}_0\left[K_1^{RE}\right]$ can be immediately plugged in the determination of time 2 savings. In contrast, the corresponding $\mathbb{E}_0\left[K_1^{\theta}\right]$ is more difficult to transparently assess because it requires computing a feedback effect between the K_1^{θ} chosen by the time 1 sophisticated DE agent, which in turn is a function of expectations about K_2^{θ}.

There are three conceptual forces that affect the coefficients of sophisticated time 1 policy compared to their naïveté case. First, the agent now anticipates that she will over-consume (relative to her naive beliefs) at time 2 out of K_1, as the forecasted response of future savings out of capital entering period 2 is smaller than under naïveté, i.e. $\alpha_{K_1}^{\theta,s} < \alpha_{K_1}^{RE}$. This force alone, coming from the $\mathbb{E}_1^{\theta}\left[C_2^{\theta}\right]$ term in (52), leads the agent to consume more today out of ε_1 to achieve consumption smoothing between period 1 and 2. Second, the misalignment of her perceived tradeoffs means that following a positive innovation ε_1, from the viewpoint of current self, the time 2 self will under-consume in period $t = 3$ relative to $t = 2$. This constitutes an indirect effect, i.e. the second term in (52), that leads to more savings. The race between these two forces is dominated here by the former, direct effect, as $\alpha_{K_1}^{\theta,s} < 0.5$, and thus the agent ends up saving less out of ε_1 than under naïveté, i.e. $\alpha_{\varepsilon_1}^{\theta,s} < \alpha_{\varepsilon_1}^{\theta}$. Third, there is the conceptual difference of the comparison groups. With sophistication, the informational state $\mathbb{E}_0\left[K_1^{\theta}\right]$ (a) matters for the K_2^{θ} solution in Proposition 8 but also (b) needs to be itself based on K_1^{θ}, a choice that in turn is affected by $\mathbb{E}_1^{\theta}\left[K_2^{\theta}\right]$ in equation (52). The effect of this fixed point consideration is less transparent, as it turns out to amplify or dampen, through a non-monotonic relationship with θ, the optimal responses of sophisticated time 1 savings to K_0 and $\mathbb{E}_{-1}\left[K_0\right]$ compared to the naïveté case.

B.8 Proof of Proposition 6

We first guess and verify the RE solution. Conjecture consumption policy

$$\begin{aligned} C_t^{RE} &= \frac{r}{1+r}\left(K_{t-1}^{RE} + \varepsilon_t + \frac{1+r}{r}\overline{Y}\right) \\ &= \frac{r}{1+r}\left(K_{t-1}^{RE} + \varepsilon_t\right) + \overline{Y} \end{aligned}$$

and the resulting savings

$$\begin{aligned} K_t^{RE} &= (1+r)\left[K_{t-1}^{RE} + \overline{Y} + \varepsilon_t - \frac{r}{1+r}\left(K_{t-1}^{RE} + \varepsilon_t\right) - \overline{Y}\right] \\ &= (1+r)\frac{1}{1+r}\left[K_{t-1}^{RE} + \varepsilon_t\right] = K_{t-1}^{RE} + \varepsilon_t \end{aligned}$$

Check the FOC by plugging in the above conjectures

$$\begin{aligned}
C_t^{RE} &= \mathbb{E}_t(C_{t+1}^{RE}) \\
\frac{r}{1+r}\left[K_{t-1}^{RE}+\varepsilon_t\right]+\overline{Y} &= \mathbb{E}_t\left[\frac{r}{1+r}\left[K_t^{RE}+\varepsilon_{t+1}\right]+\overline{Y}\right] \\
\frac{r}{1+r}\left[K_{t-1}^{RE}+\varepsilon_t\right] &= \mathbb{E}_t\left[\frac{r}{1+r}\left[K_{t-1}^{RE}+\varepsilon_t+\varepsilon_{t+1}\right]\right] \\
\frac{r}{1+r}\left[K_{t-1}^{RE}+\varepsilon_t\right] &= \frac{r}{1+r}\left[K_{t-1}^{RE}+\varepsilon_t\right]
\end{aligned}$$

so the both sides are indeed equal.

To solve for the DE poicy function, we take the FOC with respect to K_t^θ in (23):

$$u'(C_t^\theta) = \mathbb{E}_t^\theta\left[\mathcal{V}'(K_t^\theta)\right],$$

and use the envelope theorem in (24):

$$\mathcal{V}'(K_t^\theta) = u'(C_{t+1}^{RE}).$$

Combining the two, we have

$$u'(C_t^\theta) = \mathbb{E}_t^\theta\left(u'(C_{t+1}^{RE})\right).$$

We use the budget constraint to replace C_t^θ in the left hand side and obtain:

$$\begin{aligned}
K_{t-1}^\theta+\overline{Y}+\varepsilon_t-\frac{K_t^\theta}{1+r} &= \mathbb{E}_t^\theta\left[\frac{r}{1+r}\left(K_t^\theta+\varepsilon_{t+1}\right)+\overline{Y}\right] \\
K_{t-1}^\theta+\varepsilon_t-\frac{K_t^\theta}{1+r} &= \frac{r}{1+r}\mathbb{E}_t^\theta\left(K_t^\theta+\varepsilon_{t+1}\right)
\end{aligned}$$

Applying DE and using the fact that $\mathbb{E}_{t-J}K_t^{RE} = \mathbb{E}_{t-J}\left[K_{t-1}^{RE}+\varepsilon_t\right]$, we have that the expectation on the right hand side is equal to:

$$\mathbb{E}_t^\theta\left[K_t^\theta+\varepsilon_{t+1}\right] = K_t^\theta+\theta\left[K_t^\theta-\mathbb{E}_{t-J}K_t^{RE}\right] = K_t^\theta+\theta\left[K_t^\theta-\mathbb{E}_{t-J}\left(K_{t-1}^{RE}\right)\right]$$

Then

$$K_{t-1}^\theta+\varepsilon_t-\frac{K_t^\theta}{1+r} = \frac{r}{1+r}\left[K_t^\theta+\theta\left[K_t^\theta-\mathbb{E}_{t-J}\left(K_{t-1}^{RE}\right)\right]\right]$$

Rearrange:

$$\begin{aligned}
(1+r)K_{t-1}^\theta+r\theta K_{t-1}^\theta+(1+r)\varepsilon_t &= K_t^\theta+rK_t^\theta+r\theta K_t^\theta+r\theta K_{t-1}^\theta-r\theta\mathbb{E}_{t-J}\left(K_{t-1}^{RE}\right) \\
\left[1+r(1+\theta)\right]K_{t-1}^\theta+(1+r)\varepsilon_t &= \left[1+r(1+\theta)\right]K_t^\theta-r\theta\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right)
\end{aligned}$$

Then:

$$K_t^\theta = K_{t-1}^\theta-\frac{r\theta}{1+r(1+\theta)}\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right)+\frac{1+r}{1+r(1+\theta)}\varepsilon_t.$$

where $\mathbb{N}_{t-J,t-1}\left(K_{t-1}^\theta\right) = K_{t-1}^\theta-\mathbb{E}_{t-J}\left(K_{t-1}^{RE}\right)$.

Consistent with the discussion above and the naïveté assumption, we assume that memory

is based on the rational expectation solution. This is how the agent perceives capital should have evolved from the point of view of time $t-J$. Thus, we have $\mathbb{E}_{t-J}\left(K_{t-1}^{RE}\right)=K_{t-J}^{\theta}$ and the solution becomes:

$$K_t^{\theta}=K_{t-1}^{\theta}-\frac{r\theta}{1+r\left(1+\theta\right)}\left[K_{t-1}^{\theta}-K_{t-J}^{\theta}\right]+\frac{1+r}{1+r\left(1+\theta\right)}\varepsilon_t.$$

We can also rewrite this as:

$$K_t^{\theta}=\frac{1}{1+r\left(1+\theta\right)}\left[K_{t-1}^{\theta}+r\theta K_{t-J}^{\theta}+\left(1+r\right)\varepsilon_t\right].$$

C Equilibrium Conditions of the New Keynesian Model

1. Capital Euler equation:

$$\mu_t^{\theta}=\beta\mathbb{E}_t^{\theta}\left[(C_{t+1}^{RE}-bC_t^{\theta})^{-1}(R_{t+1}^{k,RE}u_{t+1}^{RE}-a(u_{t+1}^{RE}))+\mu_{t+1}^{RE}(1-\delta)\right],$$

   where μ_t^{θ} is the Lagrangian multiplier on the capital accumulation equation.

2. Utilization choice:

$$R_t^{k,\theta}=R^k(u_t^{\theta})^{\tau}$$

3. Investment first-order condition:

$$(C_t^{\theta}-bC_{t-1}^{\theta})^{-1}=\mu_t^{\theta}\left\{1-\frac{\kappa}{2}\left(\Delta I_t^{\theta}-\gamma\right)^2-\kappa\left(\Delta I_t^{\theta}-\gamma\right)\Delta I_t^{\theta}\right\}+\beta\mathbb{E}_t^{\theta}\left[\mu_{t+1}^{RE}\kappa\left(\Delta I_{t+1}^{RE}-\gamma\right)\left(\Delta I_{t+1}^{RE}\right)^2\right]$$

4. Investment growth:

$$\Delta I_t^{\theta}=I_t^{\theta}/I_{t-1}^{\theta}$$

5. Consumption Euler equation:

$$Q_t^{\theta}=\frac{\beta R_t^{\theta}}{\Pi}\mathbb{E}_t^{\theta}\left[Q_{t+1}^{RE}\right]$$

6. Definition of Q_t^{θ} :

$$\frac{Q_t^{\theta}}{Q_{t-1}^{\theta}}=\frac{\Pi}{\Pi_t^{\theta}}\left(\frac{C_t^{\theta}-bC_{t-1}^{\theta}}{C_{t-1}^{\theta}-bC_{t-2}^{\theta}}\right)^{-1}$$

7. Capital accumulation:

$$K_t^{\theta}=(1-\delta)K_{t-1}^{\theta}+\left\{1-\frac{\kappa}{2}\left(\frac{I_t^{\theta}}{I_{t-1}^{\theta}}-\gamma\right)^2\right\}I_t^{\theta}$$

8. Real wage:

$$\widetilde{W}_t^{\theta}=MC_t^{\theta}(1-\alpha)\frac{Y_t^{\theta}}{N_t^{\theta}}$$

   where $\widetilde{W}_t^{\theta}\equiv W_t^{\theta}/P_t^{\theta}$ is the real wage.

9. Capital rental rate:

$$R_t^{k,\theta} = MC_t^\theta \alpha \frac{Y_t^\theta}{K_{t-1}^\theta}$$

10. Production function:

$$Y_t^\theta = (u_t^\theta K_{t-1}^\theta)^\alpha (\gamma^t N_t^\theta)^{1-\alpha}$$

11. Optimal price setting:

$$Q_t^\theta \left\{ -\frac{1}{\lambda_f - 1} Y_t^\theta + \frac{\lambda_f}{\lambda_f - 1} MC_t^\theta Y_t^\theta - \varphi_p \left(\Pi_t^\theta - \Pi\right) \Pi_t^\theta Y_t^\theta \right\}$$
$$+ \frac{\beta \varphi_p}{\Pi} \mathbb{E}_t^\theta \left[Q_{t+1}^{RE} \left(\Pi_{t+1}^{RE} - \Pi\right) (\Pi_{t+1}^{RE})^2 Y_{t+1}^{RE} \right] = 0$$

12. Optimal wage setting:

$$Q_t^\theta \left[\left(-\frac{1}{\lambda_n - 1} \right) N_t^\theta + (C_t^\theta - bC_{t-1}^\theta) \left(\frac{\lambda_n}{\lambda_n - 1} \right) \left(N_t^\theta\right)^{1+\eta} \frac{1}{\widetilde{W}_t^\theta} - \varphi_w \left(\Pi_{w,t}^\theta - \gamma\Pi\right) \Pi_{w,t}^\theta \right]$$
$$+ \frac{\beta \varphi_w}{\Pi} \mathbb{E}_t^\theta \left[Q_{t+1}^{RE} \left(\Pi_{w,t+1}^{RE} - \gamma\Pi\right) (\Pi_{w,t+1}^{RE})^2 \right] = 0$$

13. Nominal wage inflation:

$$\Pi_{w,t}^\theta = \Pi_t^\theta \frac{\widetilde{W}_t^\theta}{\widetilde{W}_{t-1}^\theta}$$

14. Resource constraint:

$$C_t^\theta + I_t^\theta + \frac{\varphi_p}{2} \left(\Pi_t^\theta - \Pi\right)^2 Y_t^\theta + \frac{\varphi_w}{2} \left(\Pi_{w,t}^\theta - \gamma\Pi\right)^2 \widetilde{W}_t^\theta + a(u_t^\theta) K_{t-1}^\theta = Y_t^\theta$$

15. GDP:

$$Y_t^{G,\theta} = Y_t^\theta - \frac{\varphi_p}{2} \left(\Pi_t^\theta - \Pi\right)^2 Y_t^\theta - \frac{\varphi_w}{2} \left(\Pi_{w,t}^\theta - \gamma\Pi\right)^2 \widetilde{W}_t^\theta - a(u_t^\theta) K_{t-1}^\theta$$

16. Taylor rule:

$$\frac{R_t^\theta}{R} = \left(\frac{R_{t-1}^\theta}{R} \right)^{\rho_R} \left\{ \left(\frac{\widetilde{\Pi}_t^\theta}{\Pi} \right)^{\phi_\pi} \left(\frac{Y_t^{G,\theta}}{\gamma Y_{t-1}^{G,\theta}} \right)^{\phi_Y} \right\}^{1-\rho_R} exp(\varepsilon_t)$$

D Solution Algorithm

We start from a linear RE system

$$\underset{n\times n}{\mathbf{\Gamma}_0} \underset{n\times 1}{\mathbf{x}_t^{RE}} = \underset{n\times n}{\mathbf{\Gamma}_1} \underset{n\times 1}{\mathbf{x}_{t-1}^{RE}} + \underset{n\times n_s}{\mathbf{\Psi}} \underset{n_s\times 1}{\varepsilon_t} + \underset{n\times n_e}{\mathbf{\Pi}} \underset{n_e\times 1}{\eta_t^{RE}}$$

where $\mathbf{x}_t^{RE}$, ε_t and η_t^{RE} are vectors of endogenous variables, shocks, and expectation errors, respectively. A recursive law of motion can be obtained, using for example Sims (2000), as:

$$\mathbf{x}_t^{RE} = \mathbf{T}^{RE}\mathbf{x}_{t-1}^{RE} + \mathbf{R}^{RE}\varepsilon_t.$$

Note that the solution can be divided based on the non-expectation $\left(\widetilde{\mathbf{x}}_t^{RE}\right)$ and expectation terms $\left(\mathbb{E}_t\mathbf{y}_{t+1}^{RE}\right)$:

$$\begin{bmatrix} \widetilde{\mathbf{x}}_t^{RE} \\ {\scriptstyle (n-n_e)\times 1} \\ \mathbb{E}_t\mathbf{y}_{t+1}^{RE} \\ {\scriptstyle n_e\times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11}^{RE} & \mathbf{T}_{12}^{RE} \\ {\scriptstyle (n-n_e)\times(n-n_e)} & {\scriptstyle (n-n_e)\times n_e} \\ \mathbf{T}_{21}^{RE} & \mathbf{T}_{22}^{RE} \\ {\scriptstyle n_e\times(n-n_e)} & {\scriptstyle n_e\times n_e} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_{t-1}^{RE} \\ {\scriptstyle (n-n_e)\times 1} \\ \mathbb{E}_{t-1}\mathbf{y}_t^{RE} \\ {\scriptstyle n_e\times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{R}_1^{RE} \\ {\scriptstyle (n-n_e)\times n_s} \\ \mathbf{R}_2^{RE} \\ {\scriptstyle n_e\times n_s} \end{bmatrix} \varepsilon_t$$

where $\mathbf{y}_{t+1}^{RE}$ is a subset of $\widetilde{\mathbf{x}}_{t+1}^{RE}$.

Define:

$$\mathbf{x}_t^{\theta} = \begin{bmatrix} \widetilde{\mathbf{x}}_t^{\theta} \\ {\scriptstyle (n-n_e)\times 1} \\ \left(\mathbb{E}_t\mathbf{y}_{t+1}^{RE}\right)^{\theta} \\ {\scriptstyle n_e\times 1} \end{bmatrix}$$

Note that $\left(\mathbb{E}_t\mathbf{y}_{t+1}^{RE}\right)^{\theta}$ denotes the realized value for rational expectations, so it is different from $\mathbb{E}_t^{\theta}\mathbf{y}_{t+1}^{RE}$. We have:

$$\mathbb{E}_t\mathbf{y}_{t+1}^{RE} = \mathbf{M}\mathbf{T}^{RE}\mathbf{x}_t^{\theta} = \left(\mathbb{E}_t\mathbf{y}_{t+1}^{RE}\right)^{\theta}$$

where $\mathbf{M}$ is a matrix that extract the relevant elements from $\mathbf{T}^{RE}\mathbf{x}_t^{\theta}$. Note that the equation needs to be included to the system of equations for the DE model because it provides the law of motion for the realized expectations. To see this,

$$\begin{aligned}(\mathbb{E}_t\mathbf{y}_{t+1}^{RE})^{\theta} &= \underbrace{[\mathbf{M}_1 : \mathbf{0}]}_{\mathbf{M}} \begin{bmatrix} \mathbf{T}_{11}^{RE} & \mathbf{T}_{12}^{RE} \\ \mathbf{T}_{21}^{RE} & \mathbf{T}_{22}^{RE} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_t^{\theta} \\ (\mathbb{E}_t\mathbf{y}_{t+1}^{RE})^{\theta} \end{bmatrix} \\ &= \mathbf{M}_1\mathbf{T}_{11}^{RE}\widetilde{\mathbf{x}}_t^{\theta} + \mathbf{M}_1\mathbf{T}_{12}^{RE}(\mathbb{E}_t\mathbf{y}_{t+1}^{RE})^{\theta}\end{aligned}$$

so

$$-\mathbf{M}_1\mathbf{T}_{11}^{RE}\widetilde{\mathbf{x}}_t^{\theta} + (\mathbf{I} - \mathbf{M}_1\mathbf{T}_{12}^{RE})(\mathbb{E}_t\mathbf{y}_{t+1}^{RE})^{\theta} = 0.$$

It is useful to divide variables $\mathbf{x}_t^{RE}$ in the original gensys system into non-expectation terms and expectation terms:

$$\begin{bmatrix} \mathbf{\Gamma}_{0,11} & \mathbf{\Gamma}_{0,12} \\ {\scriptstyle (n-n_e)\times(n-n_e)} & {\scriptstyle (n-n_e)\times n_e} \\ \mathbf{\Gamma}_{0,21} & \mathbf{\Gamma}_{0,22} \\ {\scriptstyle n_e\times(n-n_e)} & {\scriptstyle n_e\times n_e} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_t^{RE} \\ {\scriptstyle (n-n_e)\times 1} \\ \mathbb{E}_t\mathbf{y}_{t+1}^{RE} \\ {\scriptstyle n_e\times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}_{1,11} & \mathbf{\Gamma}_{1,12} \\ {\scriptstyle (n-n_e)\times(n-n_e)} & {\scriptstyle (n-n_e)\times n_e} \\ \mathbf{\Gamma}_{1,21} & \mathbf{\Gamma}_{1,22} \\ {\scriptstyle n_e\times(n-n_e)} & {\scriptstyle n_e\times n_e} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{x}}_{t-1} \\ {\scriptstyle (n-n_e)\times 1} \\ \mathbb{E}_{t-1}\mathbf{y}_t^{RE} \\ {\scriptstyle n_e\times 1} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{\Psi}_1 \\ {\scriptstyle (n-n_e)\times n_s} \\ \mathbf{\Psi}_2 \\ {\scriptstyle n_e\times n_s} \end{bmatrix} \underset{n_s\times 1}{\varepsilon_t} + \begin{bmatrix} \mathbf{\Pi}_1 \\ {\scriptstyle (n-n_e)\times n_e} \\ \mathbf{\Pi}_2 \\ {\scriptstyle n_e\times n_e} \end{bmatrix} \underset{n_e\times 1}{\eta_t^{RE}}$$

Then, the model under DE can be expressed using matrix notation as:

$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta \mathbb{E}_t^\theta \mathbf{y}_{t+1}^{RE} + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t \tag{53}$$

where $\mathbf{\Gamma}_0^\theta$ includes the RE restrictions:

$$\begin{bmatrix} \underset{(n-n_e)\times(n-n_e)}{\mathbf{\Gamma}_{0,11}} & \underset{(n-n_e)\times n_e}{\mathbf{0}} \\ \underset{n_e\times(n-n_e)}{-\mathbf{M}_1\mathbf{T}_{11}^{RE}} & \underset{n_e\times n_e}{\mathbf{I}-\mathbf{M}_1\mathbf{T}_{12}^{RE}} \end{bmatrix} \begin{bmatrix} \underset{(n-n_e)\times 1}{\widetilde{\mathbf{x}}_t^\theta} \\ \underset{n_e\times 1}{\left(\mathbb{E}_t\mathbf{y}_{t+1}^{RE}\right)^\theta} \end{bmatrix} = \begin{bmatrix} \underset{(n-n_e)\times n_e}{-\mathbf{\Gamma}_{0,12}} \\ \underset{n_e\times n_e}{\mathbf{0}} \end{bmatrix} \mathbb{E}_t^\theta \mathbf{y}_{t+1}^{RE}$$
$$+ \begin{bmatrix} \underset{(n-n_e)\times(n-n_e)}{\mathbf{\Gamma}_{1,11}} & \underset{(n-n_e)\times n_e}{\mathbf{\Gamma}_{1,12}} \\ \underset{n_e\times(n-n_e)}{\mathbf{0}} & \underset{n_e\times n_e}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{(n-n_e)\times 1}{\widetilde{\mathbf{x}}_{t-1}^\theta} \\ \underset{n_e\times 1}{\left(\mathbb{E}_{t-1}\mathbf{y}_t^{RE}\right)^\theta} \end{bmatrix} + \begin{bmatrix} \underset{(n-n_e)\times n_s}{\mathbf{\Psi}_1} \\ \underset{n_e\times n_s}{\mathbf{0}} \end{bmatrix} \varepsilon_t$$

Then:

$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta E_t^\theta \mathbf{y}_{t+1}^{RE} + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t$$
$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta \left[(1+\theta)\, \mathbb{E}_t \mathbf{y}_{t+1}^{RE} - \sum_{j=1}^{J} \theta \alpha_j \mathbb{E}_{t-j} \mathbf{y}_{t+1}^{RE} \right] + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t$$

Suppose that we do not need all elements in $\mathbf{x}_t^\theta$ to form expectations about the future.[28] In particular, we have

$$\mathbf{y}_t^{RE} = \mathbf{M}\mathbf{x}_t^{RE}$$
$$\mathbf{x}_t^{RE} = \mathbf{T}^{RE}\mathbf{x}_{t-1}^{RE} + \mathbf{R}^{RE}\varepsilon_t$$

but can be reduced to

$$\mathbf{y}_t^{RE} = \widetilde{\mathbf{M}}\widetilde{\mathbf{x}}_t^{RE}$$
$$\widetilde{\mathbf{x}}_t^{RE} = \widetilde{\mathbf{T}}^{RE}\widetilde{\mathbf{x}}_{t-1}^{RE} + \widetilde{\mathbf{R}}^{RE}\varepsilon_t$$

Then (53) becomes

$$\mathbf{\Gamma}_0^\theta \mathbf{x}_t^\theta = \mathbf{\Gamma}_2^\theta \left[(1+\theta)\, \mathbf{M}\mathbf{T}^{RE}\mathbf{x}_t^\theta - \sum_{j=1}^{J} \theta \alpha_j \widetilde{\mathbf{M}} \left(\widetilde{\mathbf{T}}^{RE}\right)^{j+1} \widetilde{\mathbf{x}}_{t-j}^\theta \right] + \mathbf{\Gamma}_1^\theta \mathbf{x}_{t-1}^\theta + \mathbf{\Psi}^\theta \varepsilon_t. \tag{54}$$

[28] The method can easily allow for the case where we need full elements in $\mathbf{x}_t^\theta$ to form expectations. The advantage of the current method is that its state space is smaller and hence is useful for a DSGE estimation, among other things.

This becomes:

$$\begin{aligned}\left[\mathbf{\Gamma}_0^\theta - \mathbf{\Gamma}_2^\theta (1+\theta)\mathbf{M}\mathbf{T}^{RE}\right]\mathbf{x}_t^\theta = &\left[\mathbf{\Gamma}_1^\theta - \mathbf{\Gamma}_2^\theta\theta\alpha_1\mathbf{M}(\mathbf{T}^{RE})^2\right]\mathbf{x}_{t-1}^\theta \\ &- \mathbf{\Gamma}_2^\theta\theta\alpha_2\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^3\widetilde{\mathbf{x}}_{t-2}^\theta \\ &\cdots \\ &- \mathbf{\Gamma}_2^\theta\theta\alpha_J\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^{J+1}\widetilde{\mathbf{x}}_{t-J}^\theta \\ &+ \mathbf{\Psi}^\theta\varepsilon_t.\end{aligned}$$

The solution can be obtained inverting the left hand side matrix:

$$\begin{aligned}\mathbf{x}_t^\theta = &(\mathbf{A}_0^\theta)^{-1}\left[\mathbf{\Gamma}_1^\theta - \mathbf{\Gamma}_2^\theta\theta\alpha_1\mathbf{M}(\mathbf{T}^{RE})^2\right]\mathbf{x}_{t-1}^\theta \\ &- (\mathbf{A}_0^\theta)^{-1}\mathbf{\Gamma}_2^\theta\theta\alpha_2\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^3\widetilde{\mathbf{x}}_{t-2}^\theta \\ &\cdots \\ &- (\mathbf{A}_0^\theta)^{-1}\mathbf{\Gamma}_2^\theta\theta\alpha_J\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^{J+1}\widetilde{\mathbf{x}}_{t-J}^\theta \\ &+ (\mathbf{A}_0^\theta)^{-1}\mathbf{\Psi}^\theta\varepsilon_t,\end{aligned}$$

where $\mathbf{A}_0^\theta \equiv \left[\mathbf{\Gamma}_0^\theta - \mathbf{\Gamma}_2^\theta(1+\theta)\mathbf{M}\mathbf{T}^{RE}\right]$.

Writing in a more compact form, we obtain

$$\underbrace{\begin{bmatrix}\mathbf{x}_t^\theta \\ \widetilde{\mathbf{x}}_{t-1}^\theta \\ \vdots \\ \widetilde{\mathbf{x}}_{t-J+1}^\theta\end{bmatrix}}_{\mathbf{z}_t^\theta}$$

$$= \underbrace{\begin{bmatrix}\left(\mathbf{A}_0^\theta\right)^{-1}\left[\mathbf{\Gamma}_1^\theta - \mathbf{\Gamma}_2^\theta\theta\alpha_1\mathbf{M}(\mathbf{T}^{RE})^2\right] & -\left(\mathbf{A}_0^\theta\right)^{-1}\mathbf{\Gamma}_2^\theta\theta\alpha_2\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^3 & \cdots & -\left(\mathbf{A}_0^\theta\right)^{-1}\mathbf{\Gamma}_2^\theta\theta\alpha_J\widetilde{\mathbf{M}}\left(\widetilde{\mathbf{T}}^{RE}\right)^{J+1} \\ \mathbf{S} & \mathbf{0} & \cdots & \mathbf{0} \\ & & \mathbf{I} & \\ \mathbf{0} & \cdots & & \mathbf{0}\end{bmatrix}}_{\mathbf{T}^\theta}$$

$$\underbrace{\begin{bmatrix}\mathbf{x}_{t-1}^\theta \\ \widetilde{\mathbf{x}}_{t-2}^\theta \\ \vdots \\ \widetilde{\mathbf{x}}_{t-J}^\theta\end{bmatrix}}_{\mathbf{z}_{t-1}^\theta} + \underbrace{\begin{bmatrix}\left(\mathbf{A}_0^\theta\right)^{-1}\mathbf{\Psi}^\theta \\ \mathbf{0} \\ \vdots \\ \mathbf{0}\end{bmatrix}}_{\mathbf{R}^\theta}\varepsilon_t,$$

where $\mathbf{S}$ is a selection matrix that relates $\mathbf{x}_t^\theta$ to $\widetilde{\mathbf{x}}_t^\theta$:

$$\widetilde{\mathbf{x}}_t^\theta = \mathbf{S}\mathbf{x}_t^\theta.$$

Finally, we check that all variables over which we take DE present residual uncertainty. To do this, we define a vector $\mathbf{w}_t^{RE} = \mathbf{Q}\mathbf{x}_t^{RE}$ that extracts all relevant linear combinations from the vector $\mathbf{x}_t^{RE}$. This vector contains all and only the variables over which we compute DE. Then, for each element $w_{j,t}^{RE}$ of this vector we verify that the one-step-ahead conditional variance is positive:

$$Var_t\left(w_{j,t+1}^{RE}\right) = (\mathbf{Q}\mathbf{R}^{RE}\mathbf{\Sigma}(\mathbf{Q}\mathbf{R}^{RE})')_{j,j} > 0,$$

where $\mathbf{\Sigma} \equiv \mathbb{E}_t[\varepsilon_{t+1}\varepsilon_{t+1}']$ and $(\cdot)_{j,j}$ indicates the j-th diagonal element of the matrix.

E Estimation method

Our description of the methodology closely follows Christiano et al. (2010). The Bayesian estimation of impulse-response-matching method first computes the 'likelihood" of the data using approximation based on standard asymptotic distribution theory. Let $\hat{\psi}$ denote the impulse response function calculated from a local projection and let $\psi(\theta)$ denote the impulse response function from the DSGE model, which depend on the structural parameters θ. Suppose the DSGE model is correct and let θ_0 denote the true parameter vector; hence $\psi(\theta_0)$ is the true impulse response function. Then we have

$$\sqrt{T}(\hat{\psi} - \psi(\theta_0)) \xrightarrow{d} N(0, W(\theta_0)),$$

where T is the number of observations and $W(\theta_0)$ is the asymptotic sampling variance, which depends on θ_0. The asymptotic distribution of $\hat{\psi}$ can be rewritten as

$$\hat{\psi} \xrightarrow{d} N(\psi(\theta_0), V), \qquad V \equiv \frac{W(\theta_0)}{T}.$$

We use a consistent estimator of V, where the non-diagonal terms are set to zero and the main diagonal elements consist of the sample variance of $\hat{\psi}$.[29] As Christiano et al. (2011) explains, there are two advantages of this approach. First, it improves small sample efficiency and can be justified using a logic similar to the estimation of frequency-zero spectral densities in Newey and West (1987). Second, the interpretation of the estimator is graphically intuitive and transparent: it chooses parameters so that the model-implied impulse responses lie inside a confidence interval around the empirical responses.[30]

The method then calculates the likelihood

$$\mathcal{L}(\psi|\theta) = (2\pi)^{-\frac{N}{2}}|V|^{-\frac{1}{2}}\exp\{-0.5[\hat{\psi} - \psi(\theta)]'V^{-1}[\hat{\psi} - \psi(\theta)]\},$$

[29] Altig et al. (2011) and Christiano et al. (2005) use this approach in a frequentist context.

[30] In contrast, when the non-diagonal terms of V are non-zero, the estimator also takes into account the deviations of the model from data across different impulse responses in a non-intuitive manner.

where N is the total number of elements in the impulse responses to be matched. Intuitively, the likelihood is higher when the model-based impulse response $\psi(\theta)$ is closer to the empirical counterpart $\hat{\psi}$, taking into account the precision of the estimated empirical responses. We use the Bayes law to obtain the posterior distribution $p(\theta|\psi)$:

$$p(\theta|\psi) = \frac{p(\theta)\mathcal{L}(\psi|\theta)}{p(\psi)},$$

where $p(\theta)$ is the prior and $p(\psi)$ is the marginal likelihood. We simulate the posterior distribution $p(\theta|\psi)$ using the random-walk Metropolis-Hasting algorithm.

To conduct model comparisons, we use marginal likelihoods, computed from the MCMC output using the Geweke (1999)'s modified harmonic mean estimator. Inoue and Shintani (2018) provide asymptotic justification for a such exercise. In particular, they show that a model with a higher marginal likelihood is either correct or a better approximation to true impulse responses as the sample size approaches infinity.

F Additional Results

In this Appendix we report some additional results for the estimated DSGE model.

Table 1 reports the priors and the posterior mode for the model parameters of the DE model and RE re-estimated model. Standard deviations are reported in parentheses. The priors are symmetric across the two models and diffuse.

Figure 7 reports the consumption impulse responses of DE and RE models without consumption habit. In the estimated DE model without habit, the diagnosticity parameter is estimated to be significantly smaller than the benchmark DE model at $\theta = 0.64$. The mean and standard deviation of the Beta distribution that control the weights for the comparison group are 0.35 and 0.17, respectively, resulting in a more distant and dispersed memory relative to the benchmark estimated DE model with habit. The left panel of Figure 7 shows that the DE model without habit is able to generate boom-bust cycle in consumption. In response to an unexpected Fed rate cut, consumption initially spikes. This is because without habit, according to the Euler equation (43), the lower-than-usual interest rate implies falling consumption (negative expected consumption growth).[31] In the medium run, consumption gradually increases because, as shown in the right panel, unexpectedly high consumption implies unexpectedly high RE consumption $\mathbb{E}_t\left[\widehat{c}_{t+1}^{RE}\right]$ (relative to the comparison group $\mathbb{E}_t^r\left[\widehat{c}_{t+1}^{RE}\right]$) and hence high DE consumption $\mathbb{E}_t^\theta\left[\widehat{c}_{t+1}^{RE}\right]$. As consumption drops, agents become overly pessimistic (low $\mathbb{E}_t^\theta\left[\widehat{c}_{t+1}^{RE}\right]$), feeding into significantly low consumption below trend. An interesting difference of the no habit model compared to the benchmark DE model with habit is that, because agents expect a quicker reversion of consumption to trend, movements

[31] In the benchmark model with habit, this initial spike is absent because habit suppresses the initial spike by breaking the tight link between low rate and negative consumption growth.

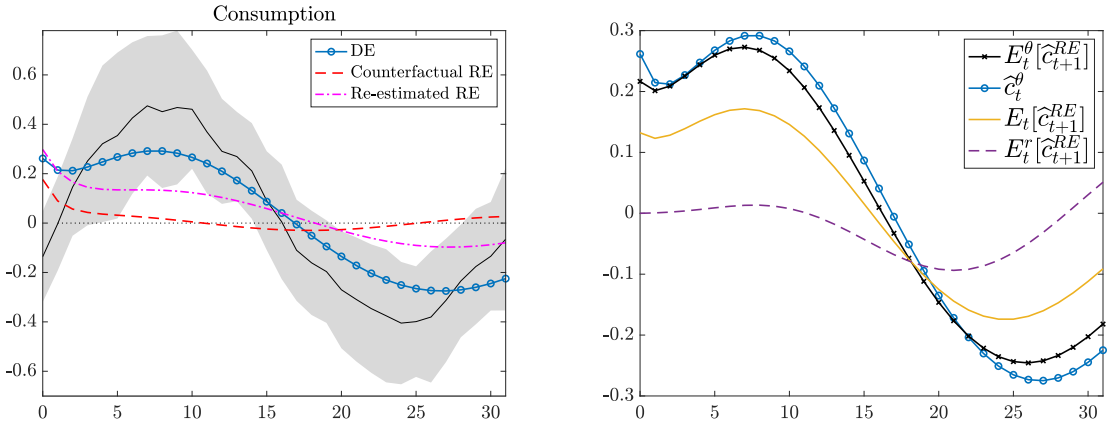
Table 1: Estimated parameters

		Prior			Posterior mode	
		Type	Mean	Std	DE	RE
η	Inverse Frisch elasticity	G	2	0.3	2.11	1.56
					(0.28)	(0.26)
b	Consumption habit	B	0.5	0.2	0.80	0.90
					(0.01)	(0.01)
τ	Utilization cost	IG	1	1	0.22	0.27
					(0.01)	(0.01)
κ	Investment adjustment cost	G	2	0.2	2.97	5.48
					(0.20)	(0.36)
φ_p	Price adjustment cost	G	100	20	195.4	291.28
					(21.2))	(31.6)
φ_w	Wage adjustment cost	G	100	20	88.6	78.8
					(20.3)	(18.0)
ρ_R	Taylor rule smoothing	B	0.5	0.2	0.009	0.82
					(0.008)	(0.018)
ϕ_π	Taylor rule inflation	N	1.5	0.4	1.000	1.000
					(0.01)	(0.025)
ϕ_Y	Taylor rule output	N	0.1	0.05	0.67	0.23
					(0.02)	(0.05)
$100\sigma_R$	Monetary policy shock	IG	1	1	0.15	0.21
					(0.01)	(0.02)
θ	Diagnosticity parameter*	N	0	0.2	1.97	–
					(0.10)	
μ	Memory distribution mean	B	0.5	0.2	0.17	–
					(0.01)	
σ	Memory distribution stdev	G	0.2	0.05	0.03	–
					(0.004)	
Log marginal likelihood					**-345**	**-369**

Notes: 'DE' corresponds to the model with diagnostic expectations and 'RE' corresponds to the rational expectations version. B refers to the Beta distribution, N to the Normal distribution, G to the Gamma distribution, IG to the Inverse-gamma distribution. (*For the prior for the diagnoscity parameter, we truncate the Normal distribution above $\theta \geq 0$.) Posterior standard deviations are in parentheses and are obtained from draws using the random-walk Metropolis-Hasting algorithm. The marginal likelihood is calculated using Geweke's modified harmonic mean estimator.

in DE consumption $\mathbb{E}_t^{\theta}\widehat{c}_{t+1}^{RE}$ are smoother and, as a result, the surprise in cumulative inflation $\pi_{J,t}^{*}$ plays a smaller role in determining the dynamics. Finally, note that in contrast to the DE model, both counterfactual and re-estimated RE models generate transitory IRFs that understate the amplitude of empirical consumption response. As a result, the log marginal likelihood of the DE model is $(-388-(-437)=)$ 49 log points higher than the RE model. We conclude that the DE model delivers boom-bust dynamics irrespective of whether it features consumption habit or not.

Figure 7: Consumption paths in a model without habit

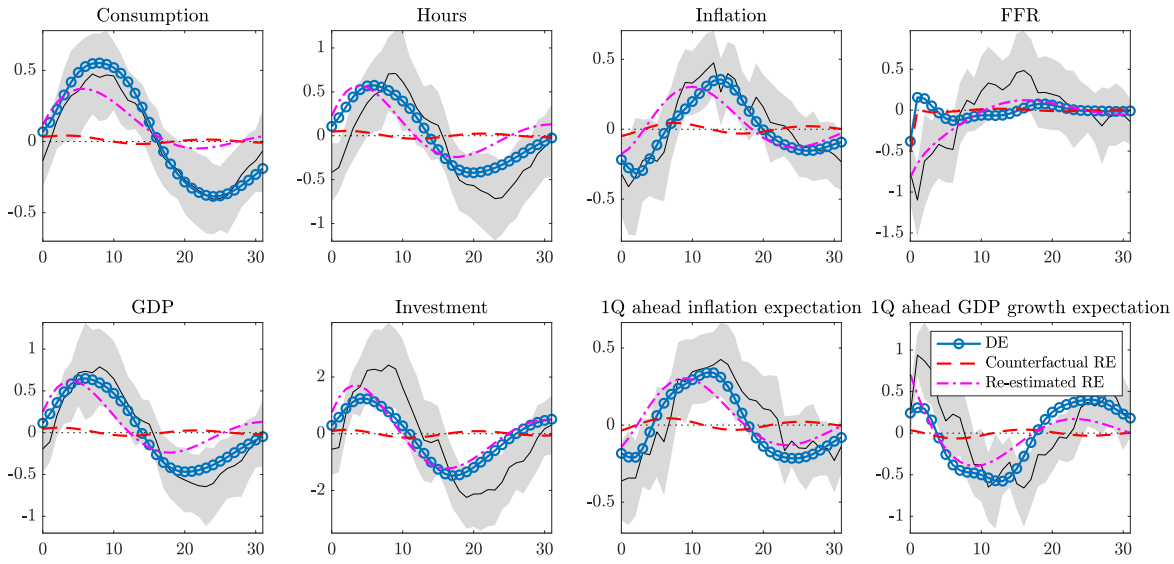


Notes: The left panel shows the consumption IRFs in response to a monetary policy shock from the DE model without habit (blue circled line), counterfactual RE model without habit (red dashed line) and the re-estimated RE model without habit (magenta dashed line). The right panel plots DE expected consumption ($\mathbb{E}_t^{\theta}\left[\widehat{c}_{t+1}^{RE}\right]$), realized equilibrium consumption ($\widehat{c}_t^{\theta}$), RE expected consumption ($\mathbb{E}_t\left[\widehat{c}_{t+1}^{RE}\right]$) and reference expectation ($\mathbb{E}_t^{r}\left[\widehat{c}_{t+1}^{RE}\right]$).

Figure 8 reports the model impulse responses when we use an alterative prior for the diagnosticity parameter θ. Specifically, we consider a Normal prior with mean 1 and standard deviation 0.2. We find that the estimated $\theta = 2.16$, slightly higher than the benchmark estimate of $\theta = 1.97$. Nevertheless, the estimated DE IRF under the alternative prior is very similar to the benchmark IRF reported in the paper.

Figure 9 reports the model impulse responses when we target the inflation and output expectations. When expectations are targeted, the diagnosticity parameter is estimated to be slightly lower than the benchmark DE model at $\theta = 1.77$. The mean and standard deviation of the Beta distribution that control the weights for the comparison group are 0.19 and 0.04, respectively and thus are similar to the estimated values in the benchmark DE model. Figure 9 shows that the DE model is able to replicate the boom-bust cycles in

Figure 8: Impulse responses to a monetary policy shock: Alternative prior for θ



Notes: The interpretation of the plotted lines follow their description for Figure 2. The responses of consumption, hours, GDP and investment are in percentage deviations from the steady states while the inflation, FFR and inflation and output growth expectations are in annual percentage points.

macro variables as well as the responses of survey expectations, although it slightly overstates consumption during the decline after the peak. The counterfactual RE model where we set the diagnosticity parameter θ to 0 while holding fixed other estimated parameters generates transitory and negligible response. The re-estimated RE model misses the magnitude of the bust in consumption, hours and GDP. It also has difficulty matching realized and expected inflation. As a result, the log marginal likelihood of the DE model is $(-636-(-645)=)$ 9 log points higher than the RE model.

Figure 10 reports the impulse response of the marginal utility to an expensionary monetary policy shock, given that the estimated Euler Equation features habits:

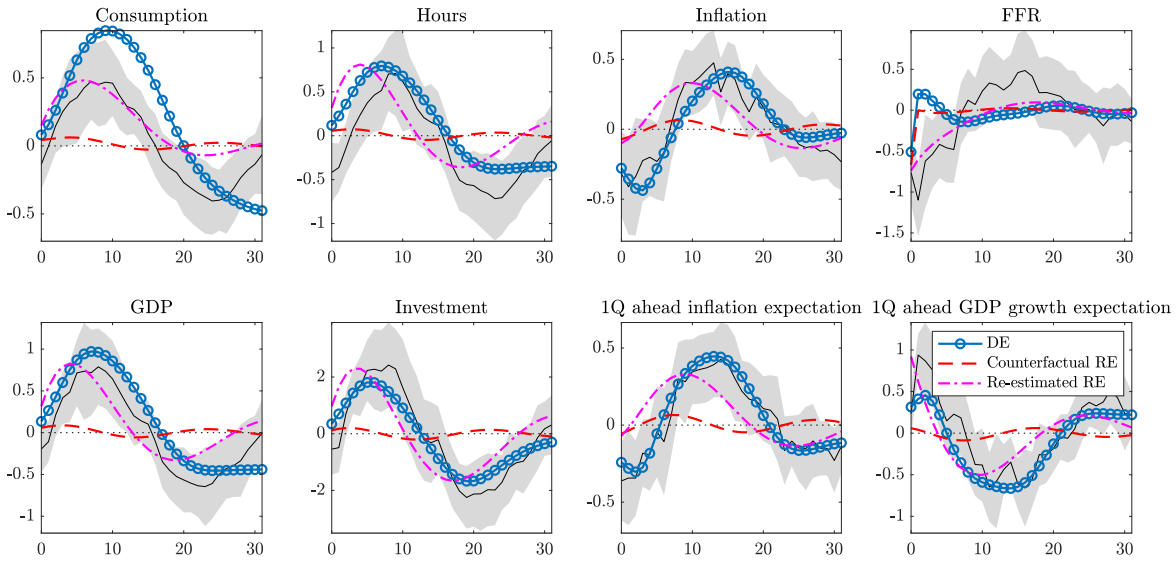
$$-\mathbb{E}_t^{\theta}[\widehat{\lambda}_{t+1}^{RE}]+\widehat{\lambda}_t^{\theta}=\widehat{r}_t^{\theta}-\mathbb{E}_t^{\theta}[\widehat{\pi}_{t+1}^{RE}]-\theta\pi_{J,t}^{*} \tag{55}$$

where

$$\widehat{\lambda}_t^{\theta}=-\frac{\widehat{c}_t^{\theta}-b\gamma^{-1}\widehat{c}_{t-1}^{\theta}}{1-b\gamma^{-1}}. \tag{56}$$

Marginal utility follows a symmetric pattern with respect to consumption, once controlling for habits. The initial increase in consumption is associated with low expected marginal utility that induces expectations of even lower marginal utility. Thus, agents expect consumption

Figure 9: Impulse responses to a monetary policy shock: Survey expectations as targets

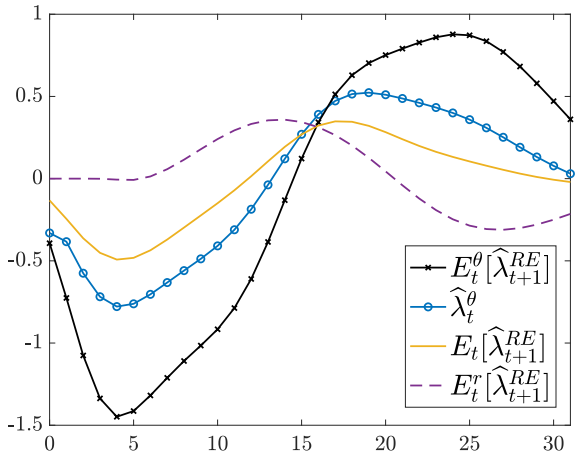


Notes: The interpretation of the plotted lines follow their description for Figure 2. The responses of consumption, hours, GDP and investment are in percentage deviations from the steady states while the inflation, FFR and inflation and output growth expectations are in annual percentage points.

to increase even when controlling for the stock of habits. As the economy progresses in its response to the shock, consumption starts declining and marginal utility to increase. However, reference expectations for marginal utility also start increasing. This is because reference expectations were formed at a time of high consumption. Under RE, agents expect a fairly quick return to the steady state from above, implying consumption lower than the stock of habits, leading to a positive RE marginal utility. However, under DE, the return to the steady state is slower than expected as agents remain overly optimistic for a while. Agents are still surprised by the high consumption, leading to a negative surprise in marginal utility, amplified by DE. Thus, past decisions feed into current beliefs, affecting the duration and amplitude of the cycle. It is only around 15 quarters that reference expectations catch up with the current marginal utility. As consumption moves below trend, agents start expecting a return to the steady from below, generating a negative reference expectation for marginal utility as consumption is expected to be higher than the stock of habits. In the bust phase, agents are surprised by the fact that consumption is still well below trend, leading to a positive surprise in marginal utility, that induces magnified DE of high marginal utility in the future.

How can we rationalize this behavior from the perspective of the Euler equation under

Figure 10: Impulse response of marginal utility



Notes: The Figure shows the DE marginal utility ($\mathbb{E}_t^\theta[\widehat{\lambda}_{t+1}^{RE}]$), realized equilibrium marginal utility ($\widehat{\lambda}_t^\theta$), RE marginal utility ($\mathbb{E}_t[\widehat{\lambda}_{t+1}^{RE}]$) and reference expectation of marginal utility ($\mathbb{E}_t^r[\widehat{\lambda}_{t+1}^{RE}]$).

DE in (43)? As mentioned in the paper, a key role is played by the surprise in cumulative inflation $\pi_{J,t}^*$ with respect to the reference expectations formed in the past. On impact, because of an increase in utilization, inflation declines. This determines a negative surprise in the price level that induces a misperception in the model relevant real interest rate that starts increasing. This perceived high real interest is, in the eyes of the agent, justified in light of a perceived acceleration in consumption that more than compensates for the habit stock. In other words, not only agents expect consumption to be higher in the future, but they also expect the marginal utility to be lower: $-\mathbb{E}_t^\theta[\widehat{\lambda}_{t+1}^{RE}] + \widehat{\lambda}_t^\theta > 0$ implies $\mathbb{E}_t^\theta\left[\widehat{c}_{t+1}^{RE} - b\gamma^{-1}\widehat{c}_t^\theta\right] - \left[\widehat{c}_t^\theta - b\gamma^{-1}\widehat{c}_{t-1}^\theta\right] > 0$. Eventually, inflation starts picking up, leading first to a reduction in the negative surprises for the price level and then eventually to positive surprises. This determines a reversal in the model relevant real interest rate that moves into the negative territory during the bust part of the cycle, when agents find the perceived low real interest rate justified in light of their excessive pessimism. Now not only they expect consumption to decline, but also to do so in a way to increase the marginal utility.

G Alternative Expression of New Keynesian Phillips Curve

In this Appendix we derive an alternative expression of the New Keynesian Phillips Curve (NKPC) of our DE model that we use to discuss the connection of inflation and real activity in Section 4. Consider the NKPC:

$$\widehat{\pi}_t^\theta = \beta\mathbb{E}_t^\theta[\widehat{\pi}_{t+1}^{RE}] + \kappa_p\widehat{mc}_t^\theta,$$

where $\kappa_p \equiv (\varphi_p \Pi^2(\lambda_f - 1))^{-1}$. The shadow RE NKPC reads:

$$\widehat{\pi}_t^{RE} = \beta \mathbb{E}_t[\widehat{\pi}_{t+1}^{RE}] + \kappa_p \widehat{mc}_t^{RE}.$$

Iterating forward the RE version of the NKPC, RE inflation can be expressed as:

$$\widehat{\pi}_t^{RE} = \kappa_p \sum_{i=0}^{\infty} \beta^i \mathbb{E}_t \widehat{mc}_{t+i}^{RE}.$$

Thus, we have:

$$\widehat{\pi}_t^{\theta} = \kappa_p \sum_{i=1}^{\infty} \beta^i \mathbb{E}_t^{\theta}[\widehat{mc}_{t+i}^{RE}] + \kappa_p \widehat{mc}_t^{\theta},$$

where we have used the fact that DE are additive as long as uncertainty is present. This expression makes clear that inflation depends on the DE of future marginal costs for a given starting value of current marginal costs.