Propagation versus Perception: Identification under Distorted Expectations in DSGE Models

Johannes Binswanger* Albert Flak†

March 26, 2026

Abstract

How much of macroeconomic shock persistence is driven by fundamental propagation versus distorted expectations? To separate these channels for estimation, we develop a unified framework in which agents forecast future states using a misspecified transition for exogenous drivers, and derive identification conditions for jointly estimating belief distortions alongside structural parameters in DSGE models. Our main result is that true shock persistence and belief distortions are locally separately identified via distinct signatures in the observable covariance structure. Yet restricted estimation that omits belief channels forces structural parameters to mimic missing dynamics, generating omitted-channel bias. We illustrate both results in the Smets and Wouters (2007) model, jointly estimating over-persistence bias and diagnostic expectations with Survey of Professional Forecasters data. Risk premium shock persistence drops from 0.57 to 0.25, and peak consumption and real wages following a TFP shock are roughly 30 percent lower under joint estimation. Our results have direct implications for how policymakers assess the persistence and consequences of structural shocks.

JEL classification: E32, E71, D84, C51, C32.

Keywords: Belief distortions; DSGE estimation; identification; omitted-channel bias; shock persistence; survey expectations.

*University of St. Gallen, Department of Economics, Switzerland. Email: johannes.binswanger@unisg.ch.

†University of St. Gallen, Department of Economics, Switzerland. Email: albert.flak@unisg.ch.

1 Introduction

How persistent are macroeconomic shocks, and when does measured persistence reflect fundamentals *(propagation)* versus distorted expectations *(perception)*? This question matters for the core objects applied macroeconomists use DSGE models to study: impulse responses, policy transmission, and the persistence parameters of exogenous drivers such as TFP, markups, fiscal shocks, and risk premia. If expectations are systematically distorted, time-series patterns that look like highly persistent fundamentals may also arise because agents propagate shocks differently through their forecasting rules. This concern is not merely hypothetical: in medium-scale DSGE estimation with diagnostic expectations, allowing for distorted forecasting can materially change estimated persistence and transmission of structural shocks (L'Huillier, Singh, and Yoo 2024).

A second question lurks beneath the first. Even if we accept that some belief distortion is present, *which one* is it? Different belief distortions can generate observationally similar macro dynamics once filtered through general equilibrium. When some distortion channels are excluded, the remaining ones — including structural persistence parameters — shift to absorb the missing dynamics. This is the problem of *omitted-channel bias*, and it cuts in two directions simultaneously: persistence estimates may be contaminated by unmodeled belief channels, and belief distortion estimates may be contaminated by each other. Without a framework that nests multiple distortions and provides identification results for their joint estimation, it is difficult to give a clean structural interpretation to any single-distortion finding.

A large empirical and theoretical literature motivates departures from rational expectations and proposes disciplined ways to model them. Prominent examples include diagnostic expectations (Bordalo, Gennaioli, Ma, and Shleifer 2020; Bordalo, Gennaioli, and Shleifer 2018), over-persistence bias (Afrouzi, Kwon, Landier, Ma, and Thesmar 2023; Rozsypal and Schlafmann 2023), narrative-based belief formation (Andre, Haaland, Roth, Wiederholt, and Wohlfart forthcoming; Flynn and Sastry 2024; Shiller 2019), and limited-attention or cognitive-discounting mechanisms (Gabaix 2016, 2020). A newer strand has brought some of these mechanisms into dynamic general equilibrium and estimated them in DSGE environments. In particular, Bianchi, Ilut, and Saijo (2024) and L'Huillier, Singh, and Yoo (2024) develop careful general-equilibrium implementations of diagnostic expectations and show that allowing for distorted forecasting can materially change estimated persistence and transmission. These contributions have substantially advanced the empirical viability of distorted-expectations DSGE

models. At the same time, empirical practice still faces the identification challenge described above: different belief distortions can generate similar macro dynamics, and many papers necessarily focus on one distortion at a time in environments that are not nested. Importantly, this does not mean prior contributions are doing something wrong. It reflects a practical constraint: without a unifying representation that nests multiple distortions in a common estimable environment, disciplined within-model comparison is difficult and the interpretation of any single-distortion finding remains open.

This paper provides a unified framework and identification results designed to address these challenges: (i) when can physical persistence and belief distortions be separately identified; and (ii) when does omitting some belief channels bias the remaining estimates? Our approach imposes a natural discipline for applied DSGE work: agents share the rational-expectations information set, but their forecasts are formed under a misspecified transition law for the exogenous drivers. Formally, the only difference between the actual and subjective economies is the *perceived transition matrix* Ψ^S governing how agents expect the exogenous state to evolve, while the *actual transition matrix* Ψ^A — the one that drives realized dynamics — remains unaffected by beliefs. In a model where exogenous drivers follow vector autoregressions, Ψ^A collects the true autoregressive coefficients; Ψ^S is their subjective counterpart, which may differ from Ψ^A due to overextrapolation, diagnostic updating, narrative confusion, or other mechanisms. We preserve *structural rationality* — agents understand the equilibrium mapping from fundamentals to outcomes — while relaxing *statistical rationality* — agents may hold misspecified beliefs about how fundamentals evolve and, crucially, about how to form expectations of their future values. This restriction nests a wide class of prominent distortions, including over-persistence bias, diagnostic updating of exogenous fundamentals, and narrative confusion, within a single parameterization that can be estimated jointly using standard likelihood-based methods.

Our framework focuses on distortions to beliefs about *exogenous* drivers rather than about endogenous equilibrium objects. This is not a claim that distorted beliefs about endogenous variables are unimportant; rather, it reflects a natural starting point for a nested, multi-distortion framework. Beliefs about endogenous objects introduce self-referential fixed-point problems: an agent's forecast of future consumption depends on their own current decision, which in turn depends on that forecast (Bianchi, Ilut, and Saijo 2024; L'Huillier, Singh, and Yoo 2024). Handling this self-referentiality requires solution methods that are potentially specific to each distortion type, making joint nested estimation across distortions considerably

more involved. By restricting distortions to the perceived transition of exogenous drivers, we preserve the tractability that makes joint estimation feasible while still capturing the belief mechanisms most directly linked to the persistence and propagation of structural shocks.

The framework delivers three results.

A positive identification result. Using the spectral density of observables as the identification object, we show that true shock persistence ρ and belief distortion parameters θ are locally separately identified from each other in the correctly specified model. The spectral density and the full sequence of autocovariances are in one-to-one correspondence via Fourier inversion under standard regularity conditions, so the result is not specific to the frequency domain: it applies equally to time-domain approaches such as likelihood-based estimation — whether Bayesian or classical — and GMM targeting second moments, which all rely on the same second-order information. The proof uses the complex extension of the spectral density — a rational function of a complex variable z — and a pole-order argument. The intuition is precise: ρ_k owns the exogenous pole at $z = \rho_k$ in the rational function representation of the spectrum, so differentiating the spectrum with respect to ρ_k raises the pole order from first to second. No belief distortion can replicate this, because θ cannot touch the exogenous block Ψ^A and therefore cannot generate a second-order pole at that location. This argument extends to the full DSGE setting, where shocks are latent and endogenous propagation creates additional poles that could in principle obscure the exogenous signature — the pole-order separation shows that it cannot.

Two negative results. Separate identification of ρ from θ does not protect restricted estimation from bias. When the true model features nonzero belief distortions but the econometrician imposes rational expectations or omits some belief channels, parameters shift to approximate the true dynamics within the restricted model class. We characterize these shifts analytically and show that the bias in estimated persistence and estimated belief distortions depends on how similar the omitted channel's effect on the observable spectrum is to the included channels. When two distortions affect the spectrum in nearly collinear directions — as is typical for distortions that both operate through perceived persistence of the same fundamentals — the bias is severe and orthogonality between channels is the exception, not the rule. This provides a disciplined lens for interpreting findings in which allowing for distorted expectations changes estimated persistence and transmission.

OPB and DE as a laboratory. Over-persistence bias (OPB) and diagnostic expectations

(DE) applied to exogenous fundamentals together form a sharp stress test for the identification results. Both distortions raise perceived persistence of fundamental drivers, yet they do so through structurally different mechanisms: OPB shifts the level of perceived persistence, while DE adds overreaction to the innovation in the fundamental. Intuitively, they seem almost impossible to separate — and for most of the spectrum, they are indeed confounded. Yet the framework shows that the two channels leave different signatures precisely because DE operates through a lagged innovation term that OPB does not. This separation is informative not only for the OPB-DE pair but as an illustration of a general principle: what distinguishes distortions in the spectrum is not their economic label but the specific restriction they impose on Ψ^S, and the Ψ^S framework makes this restriction explicit and checkable. The same workflow extends to narrative mixing, cognitive discounting, and combinations thereof.

The identification strategy uses the spectral density of observables as the population object, building on the frequency-domain identification framework for linearized DSGE models developed by Qu and Tkachenko (2012) and extended to global identification by Qu and Tkachenko (2017). The complex extension of the spectral density exploits the rational function structure of linear state-space representations, an approach rooted in Hansen and Sargent (1983). The specific identification mechanism — that differentiating the spectral density with respect to true persistence raises the pole order of the corresponding singularity in a way that no belief distortion can replicate — does not appear to have a direct antecedent in the existing literature.

We illustrate the framework empirically in the medium-scale model of Smets and Wouters (2007), estimated under rational expectations, DE only, and OPB+DE jointly, all using Survey of Professional Forecasters data at one-quarter and four-quarter horizons as direct observables for subjective expectations. Three findings stand out. First, estimated distortion parameters at the representative agent level are small, consistent with the view that consensus professional forecasts do not exhibit strong overreaction at the aggregate level; but small parameters are not economically irrelevant — an over-persistence bias of 0.017 applied to a shock with persistence 0.96 raises agents' perceived four-quarter cumulative effect by roughly eight percentage points. Second, allowing belief distortions materially reshapes estimated dynamics for specific shocks: output, consumption, and real wage responses to a TFP shock are substantially lower and faster-decaying under OPB+DE than under rational expectations, with consumption roughly 30 percent lower at the peak; and the estimated persistence of the risk premium shock falls from 0.57 under rational expectations to 0.25 under OPB+DE, as the belief channel absorbs dynamics that the rational-expectations model attributed to structural

shock persistence. Third, survey data are essential for these findings: without them, the joint OPB+DE specification collapses toward the original Smets and Wouters (2007) estimates and the belief channels are not separately disciplined — a direct empirical illustration of the identifying variation conditions underlying the theory. An additional methodological finding is that including moving-average terms in the shock processes triggers near pole-zero cancellation and identification failure, confirmed by the Komunjer and Ng (2011) rank diagnostics, providing direct empirical confirmation of the conditions the theory warns against. These differences across specifications are not resolved by model fit alone: log data densities differ by fewer than three points across all survey-disciplined specifications, meaning the data do not strongly favor one model over another. This model uncertainty itself has stakes: a policymaker assessing the medium-term consequences of a productivity improvement would conclude that real wage and consumption gains dissipate substantially faster under OPB+DE than under rational expectations.

Persistence parameters of exogenous fundamental drivers are a central empirical object in estimated DSGE models. Since endogenous propagation mechanisms are often weak (Cogley and Nason 1995), these parameters carry a large burden in fitting observed output dynamics, and their posterior estimates in benchmark specifications are frequently very high, in some cases approaching unity (Smets and Wouters 2007). Applied researchers typically estimate them under the maintained assumption of rational expectations, leaving no room for belief distortions to contribute to observed persistence. But there is already suggestive evidence that the two are substitutes in estimation: agents using parsimonious forecasting models systematically overestimate the long-run persistence of economic shocks (Fuster, Laibson, and Mendel 2010), and estimated shock persistence falls when departures from rational expectations are explicitly modeled (L'Huillier, Singh, and Yoo 2024; Milani 2007; Slobodyan and Wouters 2012). These findings are suggestive but do not constitute an identification result: they show that omitting belief distortions inflates persistence estimates, but do not characterize when and why this bias arises or what data discipline the two channels separately. That is the gap this paper fills.

The rest of the paper is organized as follows. Section 2 presents the unified state-space framework and shows how prominent distortions map into structured deviations of the perceived transition law, including combinations of distortions within a single parameterization. Section 3 develops spectral identification results and characterizes omitted-channel bias under restricted estimation. Section 4 derives analytical implications for over-persistence bias, diagnostic expectations applied to exogenous fundamentals, and their joint specification. Sec-

tion 5 describes the estimation framework, including the solution workflow, the incorporation of survey expectations as noisy observables, and the specification choices for the empirical illustration. Section 6 presents the empirical results from estimating the Smets and Wouters (2007) model under distorted expectations, discusses the posterior estimates, impulse response analysis, and the identifying role of survey data. Section 7 concludes.

2 A Unified State-Space Framework for Distorted Expectations

2.1 Scope and Maintained Rationality

We study distorted expectations in linearized DSGE models under a maintained discipline: agents may misperceive the *stochastic evolution of exogenous drivers*, but they do not misperceive the *economic structure* that maps states into choices and outcomes. Concretely, agents understand preferences, technologies, constraints, and policy rules, and they condition on the same realized information as a rational-expectation (RE) agent. Distortions enter only through the perceived conditional law of motion of an exogenous Markov state ζ_t (possibly augmented to accommodate higher-order dynamics). This restriction is not meant to imply that biased beliefs about *endogenous* variables are uninteresting; rather, it isolates a broad and empirically grounded class of belief distortions for which disciplined comparison and identification analysis are tractable in workhorse DSGE models.

The restriction can be interpreted as preserving *structural rationality* while allowing for violations of *statistical rationality*. Structural rationality means that, given any perceived law of motion for exogenous fundamentals, agents solve the correct economic optimization problem implied by the DSGE structure. Statistical rationality means that the perceived law of motion coincides with the data-generating process. Our framework maintains the former and relaxes the latter: belief distortions are captured by a misspecified conditional distribution for $\zeta_{t+1} \mid \zeta_t$, while the realized law of motion of ζ_t remains objective and unaffected by beliefs.

To see the scope clearly, consider a single exogenous driver w_t that follows the actual AR(1)

$$w_{t+1} = \rho\, w_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \text{i.i.d.}(0, \sigma^2).$$

Under the actual law, $E_t[w_{t+1}] = \rho\, w_t$. A large class of belief distortions can be represented

as a *misspecified conditional mean* of the same realized state, for example

$$E_t^S[w_{t+1}] = \tilde{\Psi}(\rho, \theta)\, s_t,$$

where s_t is the (possibly augmented) state observed at time t (e.g. $s_t = w_t$ under simple misspecification, or $s_t = (w_t, w_{t-1})'$ when the perceived forecasting rule depends on lags or on perceived "news"). The key discipline is that beliefs affect choices through $E_t^S[\cdot]$, but do *not* affect the realized evolution of w_t: the realized process remains $w_{t+1} = \rho w_t + \varepsilon_{t+1}$ under the data-generating law. In the general model, this is exactly the distinction between the actual transition $\Psi^A(\vartheta)$ and the subjective transition $\Psi^S(\vartheta, \theta)$: subjective beliefs tilt forecasts, but they do not change the physical propagation of exogenous shocks.

We view this scope restriction as a natural starting point for applied macro work. Our aim is to study identification and omitted-channel bias when multiple distortions are considered jointly. When multiple distortions are allowed to compete, empirical discipline requires a representation in which belief distortions can be nested and compared without changing the underlying DSGE structure. Restricting distortions to the exogenous law of motion delivers exactly this: competing distortions correspond to structured deviations in Ψ^S (and, in extensions, in the perceived shock mapping), while preferences, technologies, and policy rules remain fixed.

By contrast, allowing distortions to apply directly to expectations about *endogenous* equilibrium variables is a genuinely different modeling problem. The core difficulty is self-referentiality: endogenous objects are jointly determined by agents' choices, so a belief distortion that depends on a perceived "surprise" in an endogenous variable typically depends on the agent's own current action. This creates fixed-point (and potentially time-consistency) issues that require additional modeling choices beyond those needed for exogenous belief distortions.

A simple illustration makes this vivid. Consider a consumption Euler equation in a setting in which it reduces to a linear condition of the form $c_t = E_t[c_{t+1}]$.[1] Now suppose one attempts to introduce an "overreaction-to-surprises" distortion (known as "diagnostic expectations" in

[1] For example, this linear Euler equation arises exactly under quadratic utility with $r = 0$ in the textbook consumption-smoothing problem, and it also appears as a local approximation in standard linearized DSGE environments after log-linearization.

the literature) directly in consumption forecasts by writing

$$E_t^S[c_{t+1}] = E_t[c_{t+1}] + \theta\big(c_t - E_{t-1}[c_t]\big), \qquad \theta > 0.$$

This specification is inherently *self-referential*: at time t, the agent chooses c_t using $E_t^S[c_{t+1}]$, while the distortion term depends on c_t itself. Substituting the distorted forecast into the Euler equation produces a condition in which c_t appears on both sides through the "surprise" term, so equilibrium behavior must solve a fixed-point problem that pins down *simultaneously* the choice c_t and the distorted belief about c_{t+1}. In richer DSGE models with forward-looking constraints and endogenous state variables, this feedback is precisely what makes distortions about endogenous objects technically and conceptually distinct. This issue is analyzed in detail for diagnostic expectations in general equilibrium by Bianchi, Ilut, and Saijo (2024) and L'Huillier, Singh, and Yoo (2024), who develop solution methods tailored to that self-referential setting.

Our objective in this paper is different. We develop identification and omitted-channel-bias results for a tractable and empirically disciplined class of belief distortions: coherent subjective laws of motion for exogenous fundamentals (and, when needed, augmented exogenous states). This class is broad enough to nest prominent distortions studied in macroeconomics, while remaining compatible with standard DSGE solution and likelihood-based estimation methods. The next subsection formalizes the environment, the subjective and objective transition laws, and the expectations-wedge representation that makes these distortions a plug-in element in linear DSGE models.

2.2 Environment, Beliefs, and the Expectations-Wedge Representation

Time is discrete. At the beginning of period t, the exogenous state realizes and becomes observable. Agents observe the full exogenous state and all predetermined endogenous states, and then choose current endogenous variables. Throughout, we assume a complete-information environment about the current state and focus on distortions in perceived laws of motion, rather than filtering, signal extraction, or learning.

Endogenous variables and exogenous state. Let $x_t \in \mathbb{R}^{n_x}$ collect endogenous variables, including both jump variables and endogenous predetermined states. When it is useful to

separate predetermined components, we write $x_t = (x_t^j, x_t^{en})'$, where x_t^{en} collects endogenous predetermined states (e.g. capital, habits, indexation states), and x_t^j collects jump variables.

Let $\zeta_t \in \mathbb{R}^{n_\zeta}$ denote the (possibly augmented) exogenous Markov state. The state ζ_t may include true economic fundamentals as well as spurious predictors (e.g. narrative or sentiment indices). Higher-order ARMA dynamics or additional predictors are accommodated by augmenting ζ_t so that the system is first-order Markov.

Actual versus subjective laws of motion. There is a single realized history $\{\zeta_\tau\}_{\tau \le t}$ of the exogenous state. Agents are assumed to observe ζ_t at the time decisions are made. Throughout, subjective and objective expectations condition on the *same* realized information; what differs is only the perceived conditional law of motion used to form forecasts. In particular, $E_t^S[\cdot]$ denotes conditional expectations under the subjective transition for $\zeta_{t+1} \mid \zeta_t$, while $E_t[\cdot]$ denotes conditional expectations under the actual (data-generating) transition for $\zeta_{t+1} \mid \zeta_t$. To avoid proliferating notation, we write the realized state simply as ζ_t (there is no separate "subjective" state path).[2]

Under the *actual* (data-generating) law,

$$\zeta_{t+1} = \Psi^A(\vartheta)\,\zeta_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \text{i.i.d.}(0, \Sigma_\varepsilon). \tag{1}$$

Here ϑ collects the model's true *structural* parameters. Agents may instead perceive a distorted law of motion, with the distortions parameterized by a vector θ:

$$\zeta_{t+1} = \Psi^S(\vartheta, \theta)\,\zeta_t + \varepsilon_{t+1}^S, \qquad E_t^S[\varepsilon_{t+1}^S] = 0. \tag{2}$$

The innovation ε_{t+1}^S is a *subjective* innovation introduced solely to complete a coherent subjective probability measure over future states; it is not a separately realized shock.

A key maintained restriction is that subjective beliefs are *coherent*: $E_t^S[\cdot]$ is an ordinary conditional expectation under (2). In particular, linearity and the law of iterated expectations hold under E_t^S. This restriction excludes time-inconsistent expectation operators, but it still nests many prominent distortions once written as a subjective state-space model (e.g. diagnostic expectations via state augmentation).

[2] In empirical work, the econometrician typically observes a vector of time-series observables y_t and treats ζ_t as a latent state inferred from the model-implied state-space likelihood (e.g. via the Kalman filter in linear-Gaussian applications).

A natural question is why beliefs can remain distorted even though agents observe the realized exogenous state ζ_t over time. We view (2) as a disciplined reduced-form representation of *systematic* forecasting errors: agents condition on the same realized history as in the rational-expectations benchmark, but they apply a misspecified transition rule when mapping this history into forecasts. This is standard in the expectations-bias literature and is not specific to our setting. Psychologically, the premise is that forming statistically correct beliefs is effortful and maybe partly displaced by heuristics that produce forecasts that "feel right" even when they are wrong ("fast" versus "slow" thinking), so errors need not be self-correcting simply because outcomes are observed (Kahneman 2011). Diagnostic expectations provide a canonical example in which a coherent but distorted forecasting rule persists as a stable feature of expectation formation (Bordalo, Gennaioli, and Shleifer 2018). Modeling $\Psi^S(\vartheta, \theta)$ as a coherent but misspecified law keeps the equilibrium tractable and allows multiple belief channels to be estimated jointly; modeling explicit learning dynamics is an important extension but is not required for the identification and omitted-channel logic developed below.

Structural equilibrium conditions under subjective expectations. Belief distortions affect equilibrium allocations only through agents' conditional expectations. The log-linearized structural equilibrium conditions take the form

$$A_{-1}x_{t-1} + A_0 x_t + A_{+1}E_t^S[x_{t+1}] + B\zeta_t = 0, \tag{3}$$

where (A_{-1}, A_0, A_{+1}, B) are pinned down by preferences, technology, and policy rules. These structural objects do *not* depend on belief parameters.

Subjective solution and policy rule. Given (ϑ, θ), agents solve (3) taking (2) as the law of motion for ζ_t. A standard linear RE solver (applied under subjective beliefs) yields a policy function

$$x_t = P_x^S(\vartheta, \theta)\, x_{t-1} + P_\zeta^S(\vartheta, \theta)\, \zeta_t. \tag{4}$$

This policy rule summarizes behavior: agents act *as if* the subjective transition $\Psi^S(\vartheta, \theta)$ were the true law of motion. The econometrician, in contrast, evaluates the likelihood of the resulting behavior under the actual transition $\Psi^A(\vartheta)$.

Expectation wedge and objective-expectations representation. Using (4), one-step-ahead outcomes satisfy $x_{t+1} = P_x^S x_t + P_\zeta^S \zeta_{t+1}$. Therefore subjective and objective forecasts

differ only through the conditional mean of ζ_{t+1}:

$$\begin{aligned} E_t^S[x_{t+1}] &= P_x^S x_t + P_\zeta^S E_t^S[\zeta_{t+1}] = P_x^S x_t + P_\zeta^S \Psi^S \zeta_t, \\ E_t[x_{t+1}] &= P_x^S x_t + P_\zeta^S E_t[\zeta_{t+1}] = P_x^S x_t + P_\zeta^S \Psi^A \zeta_t, \end{aligned} \tag{5}$$

where $E_t[\cdot]$ denotes the econometrician's conditional expectation under (1). The corresponding expectation wedge is

$$E_t^S[x_{t+1}] - E_t[x_{t+1}] = P_\zeta^S(\vartheta, \theta)\big(\Psi^S(\vartheta, \theta) - \Psi^A(\vartheta)\big)\zeta_t. \tag{6}$$

Substituting (6) into (3) yields an equivalent representation written entirely with the objective expectation operator $E_t[\cdot]$:

$$A_{-1}x_{t-1} + A_0 x_t + A_{+1}E_t[x_{t+1}] + \big(B + \Omega(\vartheta, \theta)\big)\zeta_t = 0, \tag{7}$$

where the wedge loading matrix is

$$\Omega(\vartheta, \theta) := A_{+1}P_\zeta^S(\vartheta, \theta)\big(\Psi^S(\vartheta, \theta) - \Psi^A(\vartheta)\big). \tag{8}$$

The wedge system (7) is observationally equivalent to the subjective economy by construction, but it is solved under the *actual* exogenous transition (1) using standard RE tools. This representation is the key bridge between economic modeling under subjective beliefs and likelihood-based estimation under the objective law of motion.

Benchmark and local simplification. Our identification results are local around a rational-expectations benchmark $(\vartheta_0, 0)$ such that $\Psi^S(\vartheta_0, 0) = \Psi^A(\vartheta_0)$, implying $\Omega(\vartheta_0, 0) = 0$. Let $P_\zeta^A := P_\zeta^S(\vartheta_0, 0)$. For any scalar distortion parameter θ_i that enters only through Ψ^S, the product rule applied to (8) yields

$$\frac{\partial \Omega(\vartheta, \theta)}{\partial \theta_i} = A_{+1}\frac{\partial P_\zeta^S}{\partial \theta_i}\big(\Psi^S - \Psi^A\big) + A_{+1}P_\zeta^S\frac{\partial(\Psi^S - \Psi^A)}{\partial \theta_i}.$$

At the benchmark where $\Psi^S - \Psi^A = 0$, the first term vanishes and we obtain the simplification used repeatedly below:

$$\left.\frac{\partial \Omega(\vartheta, \theta)}{\partial \theta_i}\right|_{\theta=0} = A_{+1}P_\zeta^A \left.\frac{\partial \Psi^S(\vartheta, \theta)}{\partial \theta_i}\right|_{\theta=0}. \tag{9}$$

2.3 Examples: prominent distortions as structured misspecifications of Ψ^S

This subsection provides concrete examples that map prominent expectation distortions into the perceived transition matrix $\Psi^S(\vartheta, \theta)$. Across all examples, the maintained discipline is the same: the realized exogenous history is generated by the actual law $\Psi^A(\vartheta)$, while distorted expectations enter only through the perceived conditional mean $E_t^S[\zeta_{t+1}] = \Psi^S(\vartheta, \theta)\zeta_t$. Some distortions require augmenting the exogenous state ζ_t so that the perceived forecasting rule is coherent (i.e. admits a finite-dimensional VAR(1) representation).

For transparency, we illustrate the constructions using a standard benchmark in which persistent exogenous fundamentals $u_t \in \mathbb{R}^{n_\rho}$ follow an actual diagonal AR(1),

$$u_{t+1} = R(\rho)\, u_t + \Sigma_u \varepsilon_{t+1}, \qquad R(\rho) := \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho}), \qquad |\rho_k| < 1, \tag{10}$$

with $\varepsilon_{t+1} \sim$ i.i.d.$(0, I)$. All examples extend to non-diagonal Ψ^A and to higher-order ARMA dynamics by augmenting ζ_t.

(1) Over-persistence bias (OPB).

A baseline OPB distortion is a misspecification of perceived persistence: agents behave as if each fundamental is more (or less) persistent than it truly is. In (10), OPB can be written as

$$E_t^S[u_{t+1}] = \big(R(\rho) + \Theta^{\mathrm{OPB}}\big)u_t, \qquad \Theta^{\mathrm{OPB}} := \mathrm{diag}(\theta_1^{\mathrm{OPB}}, \ldots, \theta_{n_\rho}^{\mathrm{OPB}}), \tag{11}$$

so that the perceived transition is

$$\Psi^S(\rho, \theta^{\mathrm{OPB}}) = R(\rho) + \Theta^{\mathrm{OPB}}, \qquad \Psi^S(\rho, 0) = R(\rho), \qquad |\rho_k + \theta_k^{\mathrm{OPB}}| < 1 \ \ \forall k.$$

This simple OPB does not require any state augmentation.

The framework can also accommodate richer "horizon-dependent" over-persistence patterns in which the implied multi-step forecast $E_t^S[u_{t+h}]$ behaves as if persistence were a function of h (as in reduced-form representations of extrapolation and overreaction). Such specifications can be implemented by augmenting ζ_t with auxiliary states that reproduce the desired multi-step

forecast profile while preserving a coherent VAR(1) subjective law.[3]

(2) Diagnostic expectations (DE): coherent augmentation by "news".

Diagnostic expectations capture overreaction to recent surprises; see Bordalo, Gennaioli, and Shleifer (2018). In the AR(1) benchmark (10), define the realized one-step innovation ("surprise")

$$\nu_t := \Sigma_u \varepsilon_t = u_t - R(\rho) u_{t-1}.$$

To represent DE coherently in finite dimension, augment the exogenous Markov state as

$$\zeta_t := \begin{pmatrix} u_t \\ \nu_t \end{pmatrix} \in \mathbb{R}^{2n_\rho}.$$

Under the actual law, the augmented state evolves as

$$\zeta_{t+1} = \Psi^A(\rho)\, \zeta_t + \Lambda^A \varepsilon_{t+1}, \qquad \Psi^A(\rho) = \begin{pmatrix} R(\rho) & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Lambda^A = \begin{pmatrix} \Sigma_u \\ \Sigma_u \end{pmatrix}. \tag{12}$$

A convenient DE parameterization is an overweighting of the most recent realized surprise in the conditional mean of u_{t+1}:

$$E_t^S[u_{t+1}] = R(\rho) u_t + \Theta^{\mathrm{DE}} \nu_t, \qquad \Theta^{\mathrm{DE}} := \mathrm{diag}(\theta_1^{\mathrm{DE}}, \ldots, \theta_{n_\rho}^{\mathrm{DE}}), \qquad \theta_k^{\mathrm{DE}} \geq 0, \tag{13}$$

together with $E_t^S[\nu_{t+1}] = 0$. Hence DE admits the coherent subjective VAR(1) representation

$$E_t^S[\zeta_{t+1}] = \Psi^S(\rho, \theta^{\mathrm{DE}})\, \zeta_t, \qquad \Psi^S(\rho, \theta^{\mathrm{DE}}) = \begin{pmatrix} R(\rho) & \Theta^{\mathrm{DE}} \\ 0 & 0 \end{pmatrix}, \qquad \Psi^S(\rho, 0) = \Psi^A(\rho). \tag{14}$$

Agents correctly perceive the surprise state ν_t as having no own persistence (bottom row is zero); the distortion operates only through the perceived predictive loading of recent surprises into the mean of u_{t+1}. Equivalently, substituting $\nu_t = u_t - R(\rho) u_{t-1}$ yields the perceived AR(2)-type forecasting rule in u_t alone:

$$E_t^S[u_{t+1}] = \big(R(\rho) + \Theta^{\mathrm{DE}}\big) u_t - \Theta^{\mathrm{DE}} R(\rho) u_{t-1}.$$

[3] A leading reference that motivates horizon-dependent extrapolation forces is Afrouzi, Kwon, Landier, Ma, and Thesmar (2023). Our framework is not intended to replicate the full information-theoretic structure in that paper, but it can accommodate reduced-form linear representations of horizon-dependent perceived propagation.

(3) Combining DE and OPB.

A central objective of the paper is to allow multiple distortions to be present and estimated jointly. DE and OPB combine transparently under the same augmented state $\zeta_t = (u_t, \nu_t)'$:

$$E_t^S[u_{t+1}] = \big(R(\rho) + \Theta^{\mathrm{OPB}}\big)u_t + \Theta^{\mathrm{DE}}\nu_t, \qquad E_t^S[\nu_{t+1}] = 0, \tag{15}$$

which corresponds to the perceived transition

$$\Psi^S(\rho, \theta^{\mathrm{OPB}}, \theta^{\mathrm{DE}}) = \begin{pmatrix} R(\rho) + \Theta^{\mathrm{OPB}} & \Theta^{\mathrm{DE}} \\ 0 & 0 \end{pmatrix}, \qquad |\rho_k + \theta_k^{\mathrm{OPB}}| < 1 \ \ \forall k. \tag{16}$$

This nests rational expectations at $(\theta^{\mathrm{OPB}}, \theta^{\mathrm{DE}}) = (0, 0)$.

A small but important detail is that the "surprise" term in (36) is anchored to the *objective* law of motion. By definition,

$$\nu_{k,t} := u_{k,t} - \rho_k u_{k,t-1} = \sigma_k \varepsilon_{k,t},$$

so $\nu_{k,t}$ is the realized one-step innovation under the actual AR(1) and is therefore *invariant* to belief parameters. In particular, over-persistence bias changes the perceived conditional mean through the coefficient on $u_{k,t}$, but it does not "move the reference point" used to define the innovation state. Put differently, in the joint OPB+DE specification the DE channel loads on the realized innovation component $\nu_{k,t}$, whereas OPB loads on the level state $u_{k,t}$; OPB does not enter the construction of $\nu_{k,t}$ itself. This is exactly why DE generates the additional lag/innovation coefficient $b_{1,k} = -\rho_k \theta_k^{\mathrm{DE}}$ in (37), while OPB affects only the perceived persistence component of $b_{0,k}$.

Importantly, inspecting (16) alone does *not* settle whether $R(\rho)$, Θ^{OPB}, and Θ^{DE} are separately identified from a given dataset: OPB and DE enter the same conditional-mean equation for u_{t+1}, and both alter equilibrium behavior through the same general-equilibrium mapping (policy functions and wedges). This motivates the later results on omitted-channel bias and local confounding: even if the unrestricted model is well-posed, a restricted empirical specification that omits one channel may induce the remaining channel(s) to partially proxy for it.

(4) Narrative confusion (mixing of fundamentals).

Narrative distortions can be represented as misattribution or mixing across fundamental drivers; see Andre, Haaland, Roth, Wiederholt, and Wohlfart (forthcoming) for a prominent narrative-based application. For illustration, let $u_t := (d_t, s_t)'$ collect a demand driver d_t and a supply driver s_t. Under the actual law of motion,

$$u_{t+1} = \Psi^A(\rho)\, u_t + \varepsilon_{t+1}, \qquad \Psi^A(\rho) := \mathrm{diag}(\rho_d, \rho_s), \qquad \varepsilon_{t+1} = (\varepsilon^d_{t+1}, \varepsilon^s_{t+1})'.$$

A simple narrative-confusion specification can be written as a *confusion matrix* $C(\lambda)$ acting on the observed current fundamentals:

$$E^S_t[u_{t+1}] = \Psi^A(\rho)\, C(\lambda)\, u_t, \qquad C(\lambda) := \begin{pmatrix} 1-\lambda_{sd} & \lambda_{sd} \\ \lambda_{ds} & 1-\lambda_{ds} \end{pmatrix}, \qquad \lambda_{sd}, \lambda_{ds} \in [0,1].$$

The matrix $C(\lambda)$ is row-stochastic. Economically, it captures *misattribution across drivers*: agents forecast next period's demand using a convex combination of current demand and supply, and analogously for supply, while maintaining the correct persistence parameters ρ_d, ρ_s for each driver.

Expanding the vector form yields the scalar narrative forecasts:

$$E^S_t[d_{t+1}] = \rho_d\big((1-\lambda_{sd})\, d_t + \lambda_{sd}\, s_t\big),$$
$$E^S_t[s_{t+1}] = \rho_s\big(\lambda_{ds}\, d_t + (1-\lambda_{ds})\, s_t\big).$$

Rational expectations are nested at $C(\lambda) = I_2$ (i.e. $\lambda_{sd} = \lambda_{ds} = 0$).

(5) Narrative "spurious predictors" (phantom states).

Many narratives (including several emphasized in Shiller (2019)) can be interpreted as agents treating an observed but non-fundamental signal as predictive for future fundamentals. To capture this, enlarge the exogenous state to include a spurious predictor ξ_t and define $\zeta_t = (u_t, \xi_t)'$. A transparent benchmark is that, under the *actual* law, ξ_t is *not predictive* for u_{t+1} once u_t is given (no Granger predictability), which can be imposed by

$$\begin{pmatrix} u_{t+1} \\ \xi_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} \Psi^A_{uu} & 0 \\ \Psi^A_{\xi u} & \Psi^A_{\xi\xi} \end{pmatrix}}_{\Psi^A} \begin{pmatrix} u_t \\ \xi_t \end{pmatrix} + \begin{pmatrix} \varepsilon^u_{t+1} \\ \varepsilon^\xi_{t+1} \end{pmatrix}, \qquad E[\varepsilon^u_{t+1}(\varepsilon^\xi_{t+1})'] = 0. \tag{17}$$

The zero block $\Psi^A_{u\xi} = 0$ and orthogonality of innovations ensure that ξ_t contains no predictive

content for u_{t+1} under the actual DGP.

Under *subjective* beliefs, a narrative distortion can be represented by allowing agents to treat ξ_t as predictive:

$$E_t^S[u_{t+1}] = \Psi_{uu}^S(\vartheta,\theta)\, u_t + \Psi_{u\xi}^S(\vartheta,\theta)\, \xi_t, \qquad \Psi_{u\xi}^S(\vartheta,\theta) \neq 0, \tag{18}$$

while maintaining coherence through a full perceived VAR(1) law for ζ_t. This "phantom predictor" channel is fully covered by the framework and the subsequent identification/omitted-channel logic. For parsimony, our main theoretical statements below typically omit ξ_t and focus on distortions to perceived propagation of the true fundamentals u_t; however, the theory applies equally to settings with spurious predictors such as (17)–(18).

(6) Cognitive discounting (damping).

A parsimonious reduced-form representation of cognitive discounting is that agents down-weight the predictive content of fundamentals when forming forecasts. This kind of attenuation is central in sparse/inattentive formulations; see Gabaix (2016) and Gabaix (2020). A convenient parameterization that nests rational expectations at zero is

$$E_t^S[u_{t+1}] = K(\theta^{\mathrm{CD}})\, R(\rho)\, u_t, \qquad K(\theta^{\mathrm{CD}}) := \mathrm{diag}(1-\theta_1^{\mathrm{CD}}, \ldots, 1-\theta_{n_\rho}^{\mathrm{CD}}), \qquad \theta_k^{\mathrm{CD}} \in [0,1]. \tag{19}$$

Thus,

$$\Psi^S(\rho, \theta^{\mathrm{CD}}) = K(\theta^{\mathrm{CD}})\, R(\rho), \qquad \Psi^S(\rho, 0) = R(\rho),$$

and the distortion appears as a diagonal "selection/attenuation" operator pre-multiplying benchmark propagation.

(7) A second combined example: DE with narrative confusion.

To further illustrate how the framework allows for combinations of distortions, consider a mix *how* agents attribute predictable components across fundamentals (narratives) with *how* they extrapolate from recent surprises (DE).

Let $u_t = (d_t, s_t)'$ follow the actual diagonal AR(1) law $u_{t+1} = \Psi^A(\rho)u_t + \nu_{t+1}$ with $\Psi^A(\rho) = \mathrm{diag}(\rho_d, \rho_s)$ and ν_t the realized one-step innovation. Let $C(\lambda)$ denote the narrative confusion matrix defined above. Diagnostic expectations can then be combined with narrative confusion

by specifying the subjective conditional mean as

$$E_t^S[u_{t+1}] = \Psi^A(\rho)\, C(\lambda)\, u_t \;+\; \Theta^{\mathrm{DE}}\, \nu_t, \qquad \Theta^{\mathrm{DE}} = \mathrm{diag}(\theta_d^{\mathrm{DE}}, \theta_s^{\mathrm{DE}}), \qquad \theta_i^{\mathrm{DE}} \geq 0, \tag{20}$$

with $E_t^S[\nu_{t+1}] = 0$ under the standard augmented-state representation. Rational expectations are nested at $(C(\lambda), \Theta^{\mathrm{DE}}) = (I_2, 0)$.

These examples illustrate the two central implementation points of the paper: (i) many prominent distortions correspond to transparent, structured deviations in the perceived transition $\Psi^S(\vartheta, \theta)$, sometimes after minimal state augmentation; (ii) multiple distortions can be combined by stacking their parameters into a single vector θ and solving/estimating the DSGE model under the joint perceived law. The remainder of the paper exploits this structure to derive sharp implications for identification and omitted-channel bias, and then applies the logic to the benchmark distortions above.

3 Identification and Omitted-Channel Bias

This section derives the paper's main implications for what can and cannot be learned from macro time-series data when expectations are distorted. We proceed in three stages. First, we translate the distorted-expectations DSGE environment into a population object for identification: the spectral density of observables, viewed as an L^2 mapping that admits a local geometry of derivative directions. Second, within the correctly specified model class, we show that true shock persistence (embedded in the actual transition Ψ^A) is locally separable from belief distortions (embedded in the perceived transition Ψ^S) under standard regularity conditions. Third, we study restricted estimation in which some belief channels are omitted and show how the remaining parameters shift to approximate the missing dynamics, generating omitted-channel bias and confounding among included distortions. We close with practical implications for applied DSGE estimation and for future work.

3.1 State-space and spectrum

We translate the distorted-expectations DSGE environment from Section 2.2 into a single population object for identification: the spectral density of observables. The key point is that, once (ϑ, θ) are fixed, the model delivers a linear state-space system for the stacked state $s_t =$

$(x_t^{en}, \zeta_t)'$ and observables y_t. Here ϑ collects the structural parameters governing preferences, technology, policy rules, shock variances, and the *actual* exogenous law of motion (including the persistence vector ρ), while θ collects parameters that shift the *subjective* conditional law used in expectation formation.

Our identification strategy is then conceptually simple. First, we show that the one-way causal flow from exogenous fundamentals into endogenous variables imposes a block-triangular structure on the state transition matrix. This is purely structural: it follows from exogeneity and does not depend on the particular DSGE model. Second, we study how this structure appears in the transfer function and the spectral density. This yields a sharp separation between how true persistence parameters move the spectrum versus how belief distortions move it. Finally, we use the same spectral geometry to characterize what happens under misspecification: if the econometrician restricts θ (or omits some distortions), the remaining parameters adjust to best approximate the true spectrum, generating omitted-channel bias.

Let $s_t := (x_t^{en}, \zeta_t)'$ denote the stacked state. For any parameter vector (ϑ, θ), agents' subjective beliefs pin down a policy rule for endogenous predetermined states, and combining that behavior with the actual law of motion for ζ_t yields a linear state-space representation for (s_t, y_t):

$$s_t = F(\vartheta, \theta)\, s_{t-1} + G(\vartheta, \theta)\, \varepsilon_t, \qquad y_t = H s_t + \eta_t, \tag{21}$$

where $\varepsilon_t \sim$ i.i.d.$(0, \Sigma_\varepsilon)$ and $\eta_t \sim$ i.i.d.$(0, \Sigma_\eta)$, with $\eta_t \perp \varepsilon_\tau$ for all τ. We assume this representation is stable (all eigenvalues of $F(\vartheta, \theta)$ lie strictly inside the unit circle), so the model implies a unique covariance-stationary solution and the spectral density matrix $S_y(\omega; \vartheta, \theta)$ is well-defined for all $\omega \in [-\pi, \pi]$.

The frequency-domain mapping from structural shocks to observables is captured by the transfer function

$$T(\omega; \vartheta, \theta) := H(I - F(\vartheta, \theta)e^{-i\omega})^{-1} G(\vartheta, \theta).$$

Consequently, for any frequency $\omega \in [-\pi, \pi]$, the implied spectral density matrix of observables is

$$S_y(\omega; \vartheta, \theta) = T(\omega; \vartheta, \theta)\, \Sigma_\varepsilon\, T(\omega; \vartheta, \theta)^* + \Sigma_\eta, \tag{22}$$

where the asterisk denotes the conjugate transpose, i.e. $T(\omega)^* := \overline{T(\omega)}'$, combining matrix transposition with complex conjugation of all entries. This ensures $S_y(\omega; \vartheta, \theta)$ is Hermitian positive semidefinite for all $\omega \in [-\pi, \pi]$.

To study identification, we treat the entire spectral density curve as a single object. Define the L^2-valued spectrum mapping

$$\mathcal{S} : (\vartheta, \theta) \;\mapsto\; \big[\omega \mapsto \mathrm{vec}\big(S_y(\omega; \vartheta, \theta)\big)\big] \in L^2\left([-\pi, \pi]; \mathbb{C}^{n_y^2}\right). \tag{23}$$

We assume the solution and measurement mapping are smooth enough that $\mathcal{S}$ is Fréchet differentiable at the evaluation point of interest. Local identifiability of ρ_k with respect to θ at $(\vartheta_0, 0)$ requires

$$\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0) \;\notin\; \mathrm{Range}\big(D_\theta \mathcal{S}(\vartheta_0, 0)\big),$$

that is, the effect of ρ_k on $\mathcal{S}$ cannot be replicated by any perturbation in θ.

This condition is necessary but not sufficient for clean estimation of ρ_k in the presence of θ. The stronger condition of local non-confounding requires orthogonality in L^2: ρ_k is locally non-confounded with θ at $(\vartheta_0, 0)$ if

$$\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0) \;\perp\; \mathrm{Range}\big(D_\theta \mathcal{S}(\vartheta_0, 0)\big),$$

that is, if

$$\left\langle \frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0),\, \frac{\partial}{\partial \theta_j}\mathcal{S}(\vartheta_0, 0) \right\rangle_{L^2} = 0 \qquad \forall\, j.$$

Orthogonality implies identifiability but is strictly stronger: it ensures that the effects of ρ_k and θ on $\mathcal{S}$ are first-order uncorrelated, so that estimation of each parameter is not distorted by uncertainty in the other. Throughout, all inner products are taken with respect to the Hermitian inner product on $L^2([-\pi, \pi]; \mathbb{C}^{n_y^2})$, defined as

$$\langle f, g \rangle_{L^2} := \int_{-\pi}^{\pi} f(\omega)^*\, g(\omega)\, d\omega.$$

Throughout, we consider local statements around a benchmark $(\vartheta_0, 0)$ satisfying $\Psi^S(\vartheta_0, 0) = \Psi^A(\vartheta_0)$ (the rational-expectations benchmark).

3.2 Identification in the correctly specified model

To make the separation between *physical* persistence and belief-driven propagation as transparent as possible, we begin from the standard benchmark in which exogenous drivers evolve independently as AR(1) processes. This restriction is not conceptual—higher-order ARMA dy-

namics can be handled by state augmentation—but the diagonal AR(1) benchmark isolates the key mechanism behind the main theorem: true persistence parameters pin down the location (and, under differentiation, the order) of the exogenous poles.

ASSUMPTION 1 (Diagonal AR(1) exogenous dynamics) *Under the actual data-generating process, exogenous drivers are mutually independent AR(1), so that*

$$\zeta_t = \Psi^A(\rho)\,\zeta_{t-1} + \varepsilon_t, \qquad \Psi^A(\rho) = \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho}), \qquad |\rho_k| < 1.$$

The persistence vector $\rho = (\rho_1, \ldots, \rho_{n_\rho})$ is a subvector of ϑ.

With the exogenous dynamics established, we now turn to the full state-space representation of the economy. In standard macroeconomic models—whether under rational or distorted beliefs—endogenous variables react to exogenous states, but exogenous states evolve independently of endogenous variables. This fundamental asymmetry mechanically restricts the shape of the system's transition matrix, placing the true exogenous persistence in a highly specific structural position.

LEMMA 1 (Block-triangular propagation) *Suppose the exogenous state follows*

$$\zeta_t = \Psi^A(\rho)\zeta_{t-1} + \varepsilon_t$$

under the actual law of motion. Suppose further that, under subjective beliefs, the solved policy rule for endogenous predetermined states takes the form

$$x_t^{en} = P_x^{en}(\vartheta, \theta)\, x_{t-1}^{en} + P_\zeta^{en}(\vartheta, \theta)\, \zeta_t. \tag{24}$$

Define $s_t = (x_t^{en}, \zeta_t)'$. Then s_t necessarily evolves according to (21) with

$$F(\vartheta, \theta) = \begin{pmatrix} F_{xx}(\vartheta, \theta) & F_{x\zeta}(\vartheta, \theta) \\ 0 & \Psi^A(\rho) \end{pmatrix}, \qquad G(\vartheta, \theta) = \begin{pmatrix} G_x(\vartheta, \theta) \\ I \end{pmatrix}, \tag{25}$$

where $F_{xx}(\vartheta, \theta) = P_x^{en}(\vartheta, \theta)$, $F_{x\zeta}(\vartheta, \theta) = P_\zeta^{en}(\vartheta, \theta)\Psi^A(\rho)$, and $G_x(\vartheta, \theta) = P_\zeta^{en}(\vartheta, \theta)$. In particular, the lower-right block of F equals the true *$\Psi^A(\rho)$ and is invariant to belief distortions θ.*

All proofs are contained in Appendix A. Lemma 1 formally isolates the physical environment from agents' beliefs within the state-space matrix $F(\vartheta, \theta)$. The top block row, comprising F_{xx} and $F_{x\zeta}$, represents the endogenous propagation mechanism. These blocks are functions of the belief distortions θ because agents' policy rules depend on how they forecast the future. In contrast, the bottom block row captures the exogenous fundamentals. The lower-right block is structurally fixed at the true data-generating process $\Psi^A(\rho)$, and by the very definition of exogeneity, the lower-left block is strictly zero. This block-triangular structure serves as the mechanical engine for our identification result. A standard property of block-triangular matrices is that their eigenvalues are simply the union of the eigenvalues of their diagonal blocks. Because θ enters only the upper block row of $F(\vartheta, \theta)$, belief distortions can alter the endogenous eigenvalues (the roots of F_{xx}) and change how shocks propagate through the system via $F_{x\zeta}$. However, it is mathematically impossible for θ to alter the exogenous eigenvalues fixed in $\Psi^A(\rho)$. As we will show next, this structural separation implies that true persistence ρ and belief distortions θ leave fundamentally different mathematical signatures—specifically, different pole orders—in the spectral density.

To characterize persistence and identification, it is useful to work with a complex extension of the spectrum. Replace the unit-circle argument $e^{i\omega}$ by a general complex variable z and define the matrix-valued rational function

$$\mathscr{S}(z; \vartheta, \theta) := H(I - F(\vartheta, \theta)z^{-1})^{-1} G(\vartheta, \theta)\Sigma_\varepsilon G(\vartheta, \theta)'(I - F(\vartheta, \theta)'z)^{-1}H' + \Sigma_\eta. \tag{26}$$

Evaluating on the unit circle ($z = e^{i\omega}$) recovers the observable spectrum: $\mathscr{S}(e^{i\omega}; \vartheta, \theta) = S_y(\omega; \vartheta, \theta)$.

A *pole* of $\mathscr{S}(z; \vartheta, \theta)$ is a complex number z at which the resolvent $(I - Fz^{-1})^{-1}$ is singular. Equivalently, poles occur at values of z such that $\det(I - Fz^{-1}) = 0$, i.e. at eigenvalues of the transition matrix F. For a diagonal AR(1) exogenous block with persistence ρ_k, the corresponding pole is located on the real axis at $z = \rho_k$.

LEMMA 2 (Pole structure under block-triangularity) *Suppose Assumption 1 holds and Lemma 1 applies. Assume $F(\vartheta, \theta)$ is stable. Then, for any complex z at which the resolvents are well-defined (i.e. such that $I - F_{xx}(\vartheta, \theta)z^{-1}$ and $I - \Psi^A(\rho)z^{-1}$ are invertible), the following holds:*

1. *The resolvent $(I - F(\vartheta, \theta)z^{-1})^{-1}$ is block upper triangular and admits the block inverse*

representation

$$(I - Fz^{-1})^{-1} = \begin{pmatrix} (I - F_{xx}z^{-1})^{-1} & (I - F_{xx}z^{-1})^{-1}F_{x\zeta}z^{-1}(I - \Psi^A z^{-1})^{-1} \\ 0 & (I - \Psi^A z^{-1})^{-1} \end{pmatrix}. \tag{27}$$

The poles of $(I - Fz^{-1})^{-1}$ *are precisely the points* z *at which this resolvent fails to exist.*

2. *Because* $\Psi^A(\rho) = \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho})$*, the exogenous resolvent factor satisfies*

$$(I - \Psi^A z^{-1})^{-1} = \mathrm{diag}\big((1 - \rho_1 z^{-1})^{-1}, \ldots, (1 - \rho_{n_\rho} z^{-1})^{-1}\big),$$

and therefore has (at most) simple poles at $z = \rho_k$*,* $k = 1, \ldots, n_\rho$*. Moreover, in* (27) *the factor* $(I - \Psi^A z^{-1})^{-1}$ *appears* at most to the first power *in each entry of* $(I - Fz^{-1})^{-1}$*.*

3. *Consequently, any appearance of the exogenous factor* $(1 - \rho_k z^{-1})^{-1}$ *in entries of* $(I - Fz^{-1})^{-1}$ *arises only through the factor* $(I - \Psi^A z^{-1})^{-1}$ *in* (27)*. In particular,* $(I - Fz^{-1})^{-1}$ *cannot contain a term proportional to* $(1 - \rho_k z^{-1})^{-2}$ unless $z = \rho_k$ *is also a pole of* $(I - F_{xx}z^{-1})^{-1}$*, i.e. unless* ρ_k *is an eigenvalue of* $F_{xx}(\vartheta, \theta)$*.*

Lemma 2 strengthens Lemma 1 in a way that is central for identification. Block-triangularity pins down the *location* of exogenous roots: the true persistence parameters ρ_k enter the fixed block $\Psi^A(\rho)$ and therefore determine poles at $z = \rho_k$. Lemma 2 shows that this separation survives the matrix inversion required to construct the transfer function: the exogenous resolvent factor $(I - \Psi^A z^{-1})^{-1}$ enters $(I - Fz^{-1})^{-1}$ in a highly restricted way, appearing at most to the first power.

The key implication is about *pole order*. At the evaluation point, the observable spectrum $\mathscr{S}(z; \vartheta_0, 0)$ inherits (under visibility, see Assumption 2 below) a simple pole at $z = \rho_{0,k}$. Differentiating with respect to ρ_k necessarily differentiates the factor $(1 - \rho_k z^{-1})^{-1}$ and therefore produces a second-order term $(1 - \rho_k z^{-1})^{-2}$ (plus lower-order terms). In contrast, belief distortions θ cannot differentiate the fixed exogenous block $\Psi^A(\rho)$ and therefore cannot create second-order exogenous poles.

Higher-order singularities at $z = \rho_{0,k}$ can only arise if an endogenous pole coincides with $\rho_{0,k}$. Assumption 3 below rules out this knife-edge coincidence at the evaluation point, delivering the pole-order separation used in Theorem 1.

Having established that true persistence ρ governs the location of the exogenous poles, we

require two final regularity conditions to ensure this mathematical signature is statistically useful. First, the econometrician must actually be able to see the shock in the data. While Lemma 2 guarantees that the pole at $z = \rho_k$ exists in the underlying state dynamics, it might not survive the mapping into the observed variables, $y_t = Hs_t + \eta_t$. If a specific fundamental driver is completely masked by the observation matrix H—for instance, through aggregation or variables perfectly offsetting—the theoretical pole is algebraically canceled by a zero in the transfer function. In complex analysis, this is known as a "removable" pole. Assumption 2 formally rules this out. It requires that the shock leaves a visible trace in the spectral density matrix. Equivalently, it requires that the shock contributes a strictly non-zero weight—measured mathematically as a "residue"—to the variance or covariance of at least one pair of observed variables. This is the frequency-domain equivalent of the standard system observability condition required for estimating any state-space model.

ASSUMPTION 2 (Exogenous root visibility) *For each $k \in \{1, \dots, n_\rho\}$, the matrix-valued rational function $z \mapsto \mathscr{S}(z; \vartheta_0, 0)$ has a non-removable pole at $z = \rho_k$. Equivalently, the residue matrix*

$$R_k := \lim_{z \to \rho_k} (z - \rho_k)\, \mathscr{S}(z; \vartheta_0, 0)$$

is nonzero, i.e. at least one entry (i, j) of $\mathscr{S}(z; \vartheta_0, 0)$ has a nonzero residue at $z = \rho_k$.

This condition rules out *exact* pole–zero cancellation of the exogenous AR(1) component in the mapping from states into observables. Operationally, visibility means: changing ρ_k changes the model-implied second moments of at least one observable series in a detectable way. Exact failure is a knife-edge algebraic event, but in practice the relevant empirical issue is often *weak visibility*: the exogenous AR(1) component may load only very weakly into the chosen observables, so that the likelihood (or posterior) contains little information about ρ_k beyond the prior. For this reason, it is useful to treat visibility as a quantitative diagnostic in the solved state-space system rather than as something that can be "read off" from the data by inspection.

In typical DSGE applications the structural shocks ε_t are not observed; they are inferred from observables using the model (e.g. via the Kalman filter). This does not make visibility circular. Visibility is a property of the *model-implied* mapping from innovations into observables—that is, of the solved system matrices (F, G, H)—not a property of observing shocks directly.

A practical way to assess visibility at a parameter point (prior mode, posterior mode, or a representative draw) is to use diagnostics computed from the solved state-space representation. First, one can check whether the k-th exogenous shock explains a nontrivial share of variation in at least one observable component, for instance via forecast-error variance decompositions at multiple horizons or via the model-implied contribution of that shock to the spectrum. If the contribution is essentially zero at all horizons for all observables, then ρ_k cannot be learned from the chosen y_t regardless of the estimation method.

Second, one can check visibility directly in the object used by our identification theory: the spectrum mapping. Numerically perturb ρ_k by a small amount δ and compute the L^2 sensitivity

$$\left\| \frac{\mathcal{S}(\vartheta_0 + \delta e_k, 0) - \mathcal{S}(\vartheta_0, 0)}{\delta} \right\|_{L^2}.$$

As $\delta \to 0$, this finite difference converges to the L^2 norm of the Fréchet derivative $\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0)$, providing a direct numerical measure of the local sensitivity of the spectrum to ρ_k. If this sensitivity is close to zero, the data contain little information about ρ_k for the given observable set, which corresponds to weak visibility even if exact pole cancellation is ruled out.

Beyond ensuring visibility, we need to rule out the internal propagation mechanisms of the economy perfectly mimicking the fundamental shock. If physical frictions or belief distortions create an endogenous eigenvalue that exactly equals the exogenous persistence ρ_k, their respective poles collide. When this happens, the endogenous and exogenous channels resonate at the exact same frequency, creating a statistical singularity where propagation (i.e. fundamentals) and perception (i.e. belief distortions) can no longer be distinguished. Like the visibility condition, this is a regularity assumption that rules out knife-edge algebraic coincidences; closely related conditions appear routinely in the DSGE literature to prevent repeated roots and non-diagonalizable state transitions.

ASSUMPTION 3 (Pole separation) *At the evaluation point $(\vartheta_0, 0)$, no exogenous persistence root coincides with an endogenous eigenvalue:*

$$\rho_{0,k} \neq \lambda_j(F_{xx}(\vartheta_0, 0)) \qquad \textit{for all } k = 1, \ldots, n_\rho \textit{ and all eigenvalues } \lambda_j(\cdot).$$

The pole-separation condition can be monitored directly by comparing ρ_k to the eigenvalues of F_{xx} in the solved model, which are available as a by-product of the standard solution algorithm. Near-collisions (e.g. $|\lambda_j(F_{xx}) - \rho_k|$ small) are not forbidden by the theory, but they

typically signal weak empirical separation of exogenous and endogenous propagation and may manifest as ridges or near-flat directions in the likelihood surface around ρ_k.

With these regularity conditions in place, we can now state our primary identification result. By combining the one-way causal flow of exogenous states (Lemma 1) with the geometric properties of the spectral density (Lemma 2), Theorem 1 establishes that—under standard macroeconometric baseline conditions—the true physical persistence of fundamental shocks can be unambiguously disentangled from agents' belief distortions. Mathematically, this occurs because exogenous persistence uniquely generates higher-order poles under differentiation, ensuring its mathematical signature is linearly independent from the effects of endogenous beliefs.

Theorem 1 (Persistence identification from the spectrum) *Suppose the economy admits the block-triangular structure of Lemma 1 and satisfies Assumptions 1, 2, and 3. Assume the spectral mapping $\mathcal{S}(\vartheta, \theta)$ is Fréchet differentiable at the evaluation point $(\vartheta_0, 0)$, and partition the structural parameters as $\vartheta = (\rho, \psi)$, where ρ contains the exogenous persistence parameters. Then, evaluated at $(\vartheta_0, 0)$, the derivative of the spectrum with respect to any true persistence parameter ρ_k is linearly independent of the subspace spanned by the derivatives with respect to the belief distortions θ:*

$$\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0) \;\notin \operatorname{Range}\left(D_\theta \mathcal{S}(\vartheta_0, 0)\right) \qquad \text{for every } k = 1, \ldots, n_\rho. \tag{28}$$

In economic terms, this implies that true persistence ρ and belief distortions θ are locally separately identified when the full spectral density is used as the observable object.

Intuition of Theorem 1. To build intuition for Theorem 1, it is useful to translate the pole-order argument into a familiar time-domain comparison. Consider a stylized univariate mechanism in which an observed outcome y_t loads on a persistent exogenous state x_t:

$$x_t = \rho x_{t-1} + \varepsilon_t, \qquad y_t = \alpha x_t, \qquad |\rho| < 1.$$

In lag-operator form this implies

$$y_t = \alpha(1 - \rho L)^{-1}\varepsilon_t.$$

Replacing L by z^{-1} yields the transfer function

$$H(z) = \alpha(1 - \rho z^{-1})^{-1},$$

which has a simple pole at $z = \rho$.

The identification logic in Theorem 1 is based on how the spectrum changes under marginal parameter perturbations. Differentiating with respect to α (in analogy to the belief parameter θ in the DSGE model) gives

$$\frac{\partial y_t}{\partial \alpha} = (1 - \rho L)^{-1}\varepsilon_t, \qquad \frac{\partial H(z)}{\partial \alpha} = (1 - \rho z^{-1})^{-1},$$

so the pole at $z = \rho$ remains first order: changing α scales the response to a given physical propagation mechanism.

In contrast, differentiating with respect to true persistence necessarily differentiates the autoregressive propagator:

$$\frac{\partial y_t}{\partial \rho} = \alpha\, L(1 - \rho L)^{-2}\varepsilon_t, \qquad \frac{\partial H(z)}{\partial \rho} = \alpha\, z^{-1}(1 - \rho z^{-1})^{-2}.$$

The denominator is squared, generating a genuine second-order pole at $z = \rho$. In the time domain, the expansion

$$(1 - \rho L)^{-2} = \sum_{j=0}^{\infty}(j + 1)\rho^j L^j$$

implies

$$L(1 - \rho L)^{-2} = \sum_{t=1}^{\infty} t\, \rho^{t-1} L^t,$$

so changing ρ introduces the additional polynomial factor t in the impulse-response coefficients (up to sign changes when $\rho < 0$). Intuitively, persistence changes the internal propagation mechanism itself, not merely the amplitude of a given propagation.

In the full DSGE environment, belief distortions θ can affect endogenous propagation through the upper blocks of the transition matrix. However, under block-triangularity the true exogenous block $\Psi^A(\rho)$ is fixed. As a result, differentiating with respect to ρ necessarily differentiates the exogenous resolvent and produces second-order exogenous poles, while differentiating with respect to θ cannot. This is the economic content of the pole-order separation underlying Theorem 1.

How to read Theorem 1. A frequent apprehension in applied macroeconomics is that introducing non-rational expectation formation (e.g. diagnostic expectations or narratives) may compromise identification by creating a confounding "black box" in which belief parameters simply absorb the role of physical fundamentals. Theorem 1 dispels this particular fear: under the maintained regularity conditions and at the rational-expectations benchmark $(\vartheta_0, 0)$, the spectral impact of changing true exogenous persistence ρ_k cannot be replicated by any local change in belief distortions θ. In this precise sense, enriching the endogenous propagation mechanism with distorted expectations does not mechanically collapse true persistence and belief distortions into one another.

The core mechanism is structural. Under Assumption 1 and by Lemma 1, the stacked transition matrix contains the true exogenous transition $\Psi^A(\rho)$ as a fixed block. As a result, each persistence parameter ρ_k anchors a pole at $z = \rho_k$ in the spectrum. Belief distortions θ operate through agents' policy rules and therefore change *how* exogenous innovations load into observables (the loadings/residues), and they can alter endogenous propagation through the upper block of $F(\vartheta, \theta)$. However, they cannot change the fixed exogenous block $\Psi^A(\rho)$. This is why changes in ρ and changes in θ generate different pole-order signatures in the spectrum, delivering local separate identification of ρ from θ.

The theorem relies on two interpretable regularity conditions that can be monitored in applied work:

- *Visibility* (Assumption 2): the exogenous AR(1) component must leave a non-removable trace in observables. If a fundamental shock's contribution to observables is exactly cancelled by the observation structure, its persistence cannot be identified by any method, regardless of the estimation approach. In practice, this corresponds to the shock contributing nontrivially to at least one observable variance/covariance component in the model-implied spectrum.
- *Pole separation* (Assumption 3): no endogenous eigenvalue coincides exactly with $\rho_{0,k}$ at the evaluation point. If an endogenous eigenvalue equals the exogenous persistence, the corresponding poles collide and the pole-order argument used for separation breaks down. In practice, one can monitor the eigenvalues of F_{xx} at the prior mode (or posterior mode) and verify that they do not collapse onto the assumed physical persistence rates.

Implications for time-domain estimation and identification. Although Theorem 1 is stated in the frequency domain, its implications carry directly over to time-domain estimation.

The spectral density $S_y(\omega; \vartheta, \theta)$ and the autocovariance sequence of y_t are in one-to-one correspondence via Fourier inversion under standard regularity conditions, so equality of spectral densities implies equality of second moments. In linear Gaussian state-space models, these moments fully characterize the distribution of observables. Consequently, local identification in the frequency domain implies local identification in the time domain: if the spectral effects of ρ_k cannot be replicated by variations in θ, then the same holds for any estimation procedure based on the implied second-order structure.

More generally, likelihood-based, GMM, and Bayesian estimation methods operate on this same second-order information, differing only in how discrepancies across frequencies or moments are weighted. Theorem 1 therefore establishes a structural identification result that is invariant to the choice of estimation method: true persistence parameters ρ and belief distortions θ generate locally distinct patterns in the observable dynamics, and these differences cannot be eliminated by reweighting the data. The frequency-domain representation is thus not a restriction on the estimation method, but a convenient analytical device that makes the identification geometry transparent.

3.3 Restricted estimation and omitted-channel bias

What Theorem 1 does *not* say. First, Theorem 1 is a statement *within the correctly specified model class* using the full spectral density as the observable object. It separates true persistence ρ from belief distortions θ, but it does not assert full identification of all parameters, nor does it guarantee that distinct belief distortions (e.g. θ_1 versus θ_2) are distinguishable from each other or from endogenous structural frictions that also operate through the endogenous propagation block.

Second, and crucially for empirical practice, *identification in the unrestricted model does not rule out omitted-channel bias under restricted estimation.* If the true data are generated by (ϑ_0, θ_0) with $\theta_0 \neq 0$, but the econometrician estimates a restricted model imposing $\theta = 0$ (or omitting some distortions), the estimation problem becomes a best-approximation problem: choose restricted parameters to approximate the true spectrum within the restricted model class. Locally, this yields a pseudo-true parameter expansion in which the restricted estimates shift in the direction of the L^2 projection of the omitted distortion effect onto the derivative span of the restricted model.

Below, we distinguish two empirically relevant cases. Proposition 1 characterizes how im-

posing rational expectations ($\theta = 0$) generically biases estimates of persistence ρ (and other structural parameters). Proposition 2 extends the same logic to partially restricted models: when one class of distortions is included but another is omitted, the included distortion parameters generically absorb part of the omitted channel, biasing inference about which distortions are "present" in the data.

PROPOSITION 1 (Local omitted-channel bias in persistence under restricted belief distortions) *Let $\mathcal{S}(\vartheta, \theta) \in L^2([-\pi, \pi]; \mathbb{C}^{n_y^2})$ denote the spectrum mapping defined in (23), and partition the structural parameters as $\vartheta = (\rho, \psi)$, where ρ collects the exogenous persistence parameters and ψ collects all remaining structural parameters. Fix an evaluation point $(\vartheta_0, 0)$ with $\vartheta_0 = (\rho_0, \psi_0)$, and suppose $\mathcal{S}$ is Fréchet differentiable at this point.*

Assume the true data-generating process corresponds to (ϑ_0, θ_0) with θ_0 small, but the econometrician estimates a restricted *model imposing $\theta = 0$ by choosing (ρ, ψ) to best approximate the true spectrum in L^2:*

$$(\rho^*, \psi^*) \in \arg\min_{(\rho,\psi)} \|\mathcal{S}(\rho_0, \psi_0, \theta_0) - \mathcal{S}(\rho, \psi, 0)\|_{L^2}^2 . \tag{29}$$

Let $J_{(\rho,\psi)} := D_{(\rho,\psi)}\mathcal{S}(\vartheta_0, 0)$ and $J_\theta := D_\theta \mathcal{S}(\vartheta_0, 0)$ denote the Jacobian operators evaluated at $(\vartheta_0, 0)$. If $J_{(\rho,\psi)}$ has full column rank (equivalently, the operator $J^*_{(\rho,\psi)} J_{(\rho,\psi)}$ is invertible, where * denotes the adjoint), then the restricted best-approximation parameters satisfy the local expansion

$$\begin{pmatrix} \rho^* \\ \psi^* \end{pmatrix} = \begin{pmatrix} \rho_0 \\ \psi_0 \end{pmatrix} + \left(J^*_{(\rho,\psi)} J_{(\rho,\psi)} \right)^{-1} J^*_{(\rho,\psi)} J_\theta \, \theta_0 + o(\|\theta_0\|). \tag{30}$$

In particular, $\rho^* \neq \rho_0$ generically whenever the omitted distortion effect $J_\theta \theta_0$ has a nonzero projection onto the derivative span $\mathrm{Range}(J_{(\rho,\psi)})$.

Reconciling identification with omitted-channel bias. Theorem 1 and Proposition 1 address two different environments. The theorem is an *identification* statement within the *correctly specified* model class: if the econometrician allows both ρ and θ to vary, then the spectral effect of changing true persistence ρ cannot be replicated by changing belief distortions θ. Proposition 1 instead describes what happens under *misspecification*. If the true data are generated with belief distortions ($\theta_0 \neq 0$) but the econometrician estimates a restricted rational-expectations model imposing $\theta = 0$, then the estimator must approximate the omitted behavioral dynamics using only the remaining parameters. In this case, ρ shifts not because it

lacks theoretical identification, but because the restricted optimization problem forces (ρ, ψ) to compensate for intentionally omitted channels. In typical applications where belief distortions amplify or extend the perceived persistence of exogenous shocks beyond their true level, this compensation takes the form of an upward bias in ρ^*: the restricted model attributes belief-induced propagation to physical persistence of the exogenous process itself.

Geometry: pseudo-true parameters as an L^2 projection. At the benchmark evaluation point $(\vartheta_0, 0)$, the first-order spectral effect of the omitted distortions is the function $J_\theta \theta_0 \in L^2([-\pi, \pi]; \mathbb{C}^{n_y^2})$. The restricted model can only generate spectral deformations that lie in the derivative span $\text{Range}(J_{(\rho,\psi)})$. Linearizing the restricted best-approximation problem therefore yields the least-squares projection

$$\min_{\delta \in \mathbb{R}^{\dim(\rho,\psi)}} \left\| J_\theta \theta_0 - J_{(\rho,\psi)} \delta \right\|_{L^2}^2,$$

whose normal equations are $J_{(\rho,\psi)}^* J_{(\rho,\psi)} \delta = J_{(\rho,\psi)}^* J_\theta \theta_0$. Under the full-column-rank condition, the unique solution is exactly the shift in (30):

$$\delta = \left(J_{(\rho,\psi)}^* J_{(\rho,\psi)} \right)^{-1} J_{(\rho,\psi)}^* J_\theta \theta_0.$$

Here, $\left(J_{(\rho,\psi)}^* J_{(\rho,\psi)} \right)^{-1} J_{(\rho,\psi)}^*$ is the (left) Moore–Penrose pseudoinverse of $J_{(\rho,\psi)}$. The induced orthogonal projector onto the restricted derivative span is

$$\Pi_{(\rho,\psi)} := J_{(\rho,\psi)} \left(J_{(\rho,\psi)}^* J_{(\rho,\psi)} \right)^{-1} J_{(\rho,\psi)}^*,$$

so the restricted model matches the component $\Pi_{(\rho,\psi)}(J_\theta \theta_0)$ of the omitted distortion effect and necessarily leaves the orthogonal component $(I - \Pi_{(\rho,\psi)})(J_\theta \theta_0)$ unmatched. This is the precise sense in which omitted-channel bias arises: whenever $J_\theta \theta_0$ has a nonzero projection onto $\text{Range}(J_{(\rho,\psi)})$, the pseudo-true parameters must move at first order, and $\rho^* \neq \rho_0$ generically.

The cross-Gram operator and the omitted variable bias analogy. The bias formula (30) is the infinite-dimensional analogue of the omitted variable bias formula in linear regression. In a finite-dimensional regression $y = X\beta + Z\gamma + \varepsilon$, omitting Z yields a bias of $(X'X)^{-1}X'Z\gamma$ in the estimate of β. Here $J_{(\rho,\psi)}$ plays the role of the included regressors, J_θ the omitted regressors, and θ_0 the true coefficient on the omitted channel; the L^2 inner product replaces the finite-dimensional dot product. The key object governing bias severity is the *cross-Gram operator* $J_{(\rho,\psi)}^* J_\theta$, whose (i, j) entry is the L^2 inner product $\langle \partial \mathcal{S} / \partial (\rho, \psi)_i, \, \partial \mathcal{S} / \partial \theta_j \rangle_{L^2}$

between the spectral derivative of the i-th included parameter and the j-th omitted channel. The bias is exactly zero if and only if $J^*_{(\rho,\psi)} J_\theta = 0$ — that is, if the derivative spans of the restricted and omitted parameters are orthogonal in L^2. Theorem 1 establishes that $\partial \mathcal{S}/\partial \rho_k \notin \text{Range}(D_\theta \mathcal{S})$, ruling out exact collinearity; but orthogonality of the derivative spans is a strictly stronger condition that need not hold in general. When it fails, the cross-Gram operator is nonzero and the bias is generically present.

Implications for time-domain estimation and estimation methods. Although Proposition 1 is formulated using an L^2 approximation criterion in the frequency domain, the underlying mechanism is not specific to that representation. The spectral density and the autocovariance sequence are in one-to-one correspondence via Fourier inversion, so working with the full spectrum is equivalent to working with the full set of second moments in the time domain. More generally, under standard regularity conditions, likelihood-based, GMM, and Bayesian estimators under misspecification converge to pseudo-true parameters that solve analogous approximation problems for the second-order structure of the data (see, e.g., Chernozhukov and Hong 2003; Gallant and White 1988; White 1982), differing primarily in how discrepancies are weighted across frequencies or moments. The L^2 formulation therefore serves as a transparent benchmark that isolates the common geometric structure of the problem: the pseudo-true parameter arises from projecting the omitted distortion effect $J_\theta \theta_0$ onto the derivative span $\text{Range}(J_{(\rho,\psi)})$. Consequently, the existence of bias is determined by this geometric relationship and is invariant to the estimation method. In particular, whenever $J_\theta \theta_0$ has a nonzero projection onto this span, the estimated structural parameters must adjust to absorb part of its effect under any smooth estimation procedure. While the magnitude of this adjustment depends on the estimator-specific weighting, the presence of bias is a generic implication of the model structure rather than a feature of the L^2 criterion.

From omitting all distortions to omitting some distortions. Proposition 1 captures the extreme restriction $\theta = 0$. A common empirical practice is less extreme: researchers may include one class of belief distortions while omitting another. The next proposition shows that the same projection logic then applies *within* the belief block: when $\theta = (\theta_1, \theta_2)$ and θ_2 is omitted, the included distortions θ_1 generically absorb part of the omitted channel, biasing inference about which belief distortions are quantitatively important.

PROPOSITION 2 (Local confounding under partially restricted belief distortions) *Partition belief distortions as $\theta = (\theta_1, \theta_2)$ and structural parameters as $\vartheta = (\rho, \psi)$. Fix an eval-*

uation point $(\vartheta_0, \theta_{1,0}, 0)$ *and assume the spectrum map* $\mathcal{S}(\vartheta, \theta) \in L^2([-\pi, \pi]; \mathbb{C}^{n_y^2})$ *is Fréchet differentiable at this point.*

Assume the true data are generated by $(\vartheta_0, \theta_{1,0}, \theta_{2,0})$ *with* $\|\theta_{2,0}\|$ *small, but the econometrician estimates a* partially restricted *model that includes* θ_1 *and imposes* $\theta_2 = 0$ *by choosing* (ρ, ψ, θ_1) *to best approximate the true spectrum in* L^2*:*

$$(\rho^*, \psi^*, \theta_1^*) \in \arg\min_{(\rho,\psi,\theta_1)} \|\mathcal{S}(\vartheta_0, \theta_{1,0}, \theta_{2,0}) - \mathcal{S}(\rho, \psi, \theta_1, 0)\|_{L^2}^2 . \tag{31}$$

Let

$$J_1 := D_{(\rho,\psi,\theta_1)}\mathcal{S}(\vartheta_0, \theta_{1,0}, 0), \qquad J_2 := D_{\theta_2}\mathcal{S}(\vartheta_0, \theta_{1,0}, 0),$$

denote the Jacobian operators evaluated at the restricted benchmark. If J_1 *has full column rank (equivalently,* $J_1^* J_1$ *is invertible), then the restricted best-approximation parameters satisfy the local expansion*

$$\begin{pmatrix} \rho^* \\ \psi^* \\ \theta_1^* \end{pmatrix} = \begin{pmatrix} \rho_0 \\ \psi_0 \\ \theta_{1,0} \end{pmatrix} + (J_1^* J_1)^{-1} J_1^* J_2 \, \theta_{2,0} + o(\|\theta_{2,0}\|). \tag{32}$$

In particular, the included distortion parameter θ_1 *is locally biased (i.e.* $\theta_1^* \neq \theta_{1,0}$*) whenever the* θ_1*-component of the projection term in* (32) *is nonzero.*

Behavioral omitted-channel bias and the limits of pole-order separation. Proposition 2 formalizes an empirically relevant mechanism that is distinct from the non-confounding result in Theorem 1. The theorem exploits a pole-order separation to distinguish *physical* persistence ρ from belief distortions θ in the correctly specified model class. By contrast, Proposition 2 shows that this pole-order logic does *not* provide a generic firewall *within* the belief block. When the true economy features an omitted distortion $\theta_{2,0} \neq 0$ but the econometrician estimates a partially restricted model imposing $\theta_2 = 0$, the included distortion block θ_1 and the remaining structural parameters (ρ, ψ) generically reallocate to approximate the omitted spectral deformation.

The mechanism is the same projection geometry as in Proposition 1, but now it operates in a larger restricted span. Linearizing around the restricted benchmark $(\vartheta_0, \theta_{1,0}, 0)$, the omitted channel generates a first-order spectral deformation $J_2\theta_{2,0} \in L^2([-\pi, \pi]; \mathbb{C}^{n_y^2})$. The partially restricted model can only fit deformations in the derivative span Range(J_1), where $J_1 = D_{(\rho,\psi,\theta_1)}\mathcal{S}$. Thus the pseudo-true shift in (32) is the L^2 projection of $J_2\theta_{2,0}$ onto

Range(J_1). Equivalently, defining the orthogonal projector

$$\Pi_1 := J_1 \left(J_1^* J_1\right)^{-1} J_1^*,$$

the best-fitting restricted model matches $\Pi_1(J_2\theta_{2,0})$ and necessarily leaves $(I - \Pi_1)(J_2\theta_{2,0})$ unmatched. Whenever the θ_1-component of this projection is nonzero, the estimated θ_1^* is biased and partially acts as the omitted distortion "in disguise."

This clarifies a substantive limit of pole-order arguments. Under block-triangularity, belief distortions affect the spectrum through endogenous propagation and shock loadings, but they do not differentiate the fixed exogenous resolvent factor. As a result, derivatives with respect to θ_1 and θ_2 generically live in the same pole-order class (they preserve exogenous pole order), so the pole-order separation that protects ρ does not, by itself, separate competing belief distortions from one another.

The contrast with Theorem 1 is therefore sharp and structural. The theorem protects ρ from belief confounding because ρ-derivatives generate second-order poles at $z = \rho_k$ while θ-derivatives generate at most first-order poles there — a structural pole-order separation that follows directly from block-triangularity and the invariance of $\Psi^A(\rho)$ to belief parameters. No analogous separation is available between θ_1 and θ_2: both classes of belief distortions operate through the endogenous propagation block, neither differentiates the fixed exogenous resolvent, and both therefore share the same pole-order class. This is why Theorem 1 delivers a structural identification result while Proposition 2 can only characterize the direction and magnitude of bias under partial misspecification.

Distinguishing θ_1 from θ_2 therefore requires additional identifying content beyond the aggregate time-series spectrum alone: additional theoretical restrictions (e.g. parametric structure on Ψ^S) and/or additional data that directly discipline beliefs (e.g. survey expectations or other moments that load differentially on subjective forecasts).

Omitted-channel bias across belief distortions. For empirical practice, Proposition 2 gives a sharp warning: estimating one belief distortion at a time can generate misleading conclusions about which distortion is "present" or "quantitatively important," because included distortions generically absorb omitted ones. This motivates our nested framework and the subsequent applications, in which we map prominent distortions into explicit restrictions on Ψ^S and compare their implied wedge directions and spectral signatures within a common model environment.

Implications for estimation methods. As in Proposition 1, the bias mechanism characterized here is not specific to the L^2 formulation in the frequency domain. More generally, likelihood-based, GMM, and Bayesian estimators under misspecification solve analogous projection problems in the same tangent space, differing only in how discrepancies across frequencies or moments are weighted. Consequently, the confounding of θ_1 and θ_2 described above is a structural feature of the model rather than an artifact of the estimation method: whenever the omitted distortion induces spectral deformations that lie in the derivative span of the included parameters, those parameters must adjust to absorb part of its effect. While the magnitude of this adjustment depends on the estimator-specific weighting, the presence of confounding is generic.

The role of data. All identification statements above are population results: they characterize how parameters map into the *joint second-order dynamics* of the observables. Nevertheless, the results speak directly to data choice, because the dataset enters the theory through the measurement equation $y_t = Hs_t + \eta_t$. Changing the observable set (and measurement equations) changes the mapping $(\vartheta, \theta) \mapsto \mathcal{S}(\vartheta, \theta)$ and therefore changes which channels can be distinguished.

A first implication concerns separating true persistence ρ from belief distortions θ. Theorem 1 shows that, in a correctly specified model, ρ is locally protected from confounding with θ when the full joint dynamics of y_t are used. However, this conclusion requires that the exogenous AR(1) component is *visible* in the chosen observables (Assumption 2). In applied terms, if the shock or state whose persistence is of interest does not load into any measured series (or is effectively canceled by aggregation or dominated by measurement error), then its persistence cannot be learned from the data regardless of whether belief distortions are included. A second, equally practical point is that belief distortions only become empirically relevant if the chosen observables are sensitive to expectation wedges, i.e. if forward-looking equilibrium restrictions transmit changes in forecasts into measurable outcomes.

A further implication concerns distinguishing *different* belief distortions from one another. Proposition 2 shows that macro time series alone can generically support "proxy" behavior: an included distortion can absorb an omitted one if both move the implied dynamics of y_t in similar directions. Whether this happens is not a universal property of the distortions; it is a joint property of the distortions *and* the observable set. This suggests a concrete empirical strategy: when competing belief channels are plausible, researchers should (i) estimate nested specifications (single-channel and multi-channel), and (ii) expand the observable set in ways

that load differentially on the belief objects that separate the channels.

In particular, expectation data are natural candidates for strengthening identification across belief distortions because they are direct noisy observables of subjective forecasts. For example, diagnostic expectations are a restriction on how agents map realized surprises into forecasts. Accordingly, survey expectations (ideally at multiple horizons and/or using forecast revisions) can be especially informative about DE-specific components that are difficult to infer from macro aggregates alone. Likewise, for narrative-type distortions that operate through spurious predictors, including narrative proxies (or observables that move with them) helps isolate the distortion because it creates observables that load directly on the relevant non-fundamental state.

3.4 Implications for applied DSGE work estimation

Takeaways for applied researchers. The theory delivers two practical messages. First, allowing for distorted expectations does *not* mechanically destroy inference about the persistence of exogenous fundamentals. Under standard baseline conditions (the shock leaves a non-negligible imprint on at least one observable second moment, and there are no knife-edge root collisions), true persistence parameters ρ are locally non-confounded with belief distortions in the *unrestricted* model that admits both objects.

Second, this theoretical separation does *not* protect estimation from omitted-channel bias. If belief distortions are present in the data but are excluded (or only partially included) in the estimated model, the remaining parameters adjust to approximate the missing dynamics. This is why persistence estimates can move when behavioral channels are omitted: not because ρ is "unidentified" in principle, but because a restricted model is being forced to fit the wrong dynamics.

A disciplined estimation workflow. The projection logic in Propositions 1–2 suggests a simple workflow for applied DSGE users.

1. *Do not estimate belief distortions one at a time by default.* A single-distortion specification can easily attribute to the included distortion what is actually generated by an omitted belief channel (or by omitted structural propagation). If the goal is to learn which distortions matter, the disciplined baseline is a nested specification in which multiple distortions compete within a common model environment.

2. *Treat shifts in estimated ρ as a diagnostic, not as a nuisance.* If ρ changes materially when adding or removing expectation distortions, the natural interpretation is not "the model became unidentified." Rather, the restricted model was previously using ρ (and other parameters) to approximate belief-driven dynamics that were omitted by construction. This is precisely the omitted-channel mechanism formalized in Proposition 1.

3. *Diagnose visibility and near-collisions in the solved state-space system.* The identification arguments are population statements about the *solved* state-space representation and the chosen observables. In particular, "visibility" is not about observing shocks directly (shocks are latent), but about whether the exogenous AR(1) component survives the equilibrium and measurement mapping into y_t without (exact or near) pole–zero cancellation. In practice, it is best treated as a quantitative diagnostic rather than a trivial checkbox. At the prior or posterior mode (and, if desired, on a small set of representative posterior draws), one can monitor: (i) *visibility*: the shock explains a non-negligible share of forecast-error variance for at least one observable at some horizon, or the model-implied spectrum exhibits non-negligible numerical sensitivity to ρ_k; (ii) *no near-collisions*: the endogenous eigenvalues of F_{xx} are not close to the physical persistence rates ρ_k (near-collisions imply weak separation even if exact equality is ruled out).

4. *Use belief data to separate belief channels, not to "make ρ identifiable."* Theorem 1 separates ρ from θ using the structure of exogenous propagation, but it does not generically separate θ_1 from θ_2. When the empirical question is "which belief distortion is present," additional identifying content is often needed. Survey expectations and other belief measures are natural candidates because they load directly on subjective forecasts and can discipline the mapping from primitives to perceived dynamics.

How to interpret "small effects" of distortions in medium-scale models. A natural question is how these arguments should be read in policy-grade DSGE models such as Smets–Wouters. Empirically, it is not unusual to find that adding belief distortions changes likelihoods and conventional impulse responses only modestly. Our theory suggests a disciplined interpretation. Medium-scale models often contain rich endogenous propagation mechanisms (habits, indexation, adjustment costs, moving-average components, etc.), which generate endogenous eigenvalues of F_{xx} that may lie close to the physical persistence rates ρ_k. Such *near-collisions* do not violate our assumptions (which rule out exact coincidences), but they can make pole

separation quantitatively weak, reducing the marginal empirical leverage of adding additional expectation wedges in fit and in standard IRFs. Moreover, even when ρ is protected from confounding with θ in the unrestricted model class, belief distortions can be difficult to disentangle from the existing endogenous propagation channels collected in ψ. In this sense, "small incremental effects" of distorted expectations in a policy-grade model are consistent with the model already providing substantial scaffolding for persistence, rather than evidence that distorted expectations are irrelevant.

Scope, relationship to the literature, and next steps. Our maintained restriction—that belief distortions operate through subjective beliefs about the law of motion of exogenous fundamentals—is not meant to claim that agents never form biased expectations about endogenous objects. Rather, it reflects a modeling tradeoff that is particularly relevant for applied DSGE work. Distortions applied directly to endogenous variables typically create self-referential fixed-point issues (and, depending on the operator, potential time-inconsistencies) that make it difficult to nest multiple distortions in a single estimable environment. By contrast, restricting distortions to exogenous fundamentals preserves a coherent subjective state-space representation and keeps the model within the class that can be solved and estimated with standard linear tools.

This perspective is complementary to work that embeds specific belief operators directly into general equilibrium and then estimates a rich medium-scale model. For example, L'Huillier, Singh, and Yoo (2024) develop an implementation of diagnostic expectations within a New Keynesian framework and evaluate its empirical performance in a medium-scale DSGE environment with many benchmark frictions and shocks. Our contribution does not compete with such implementations; it clarifies general identification logic that applies across belief distortions and highlights why one-distortion-at-a-time estimation (or imposing rational expectations) generically induces pseudo-true parameter shifts.

Finally, the framework points to several natural extensions that are complementary rather than competing with the present contribution:

- *Separating belief distortions from each other:* develop additional identifying restrictions or bring in additional belief data (surveys, expectation panels, narrative measures) to discipline Ψ^S beyond what aggregate macro series can deliver.
- *Partial observability and filtering:* allow agents to observe noisy signals about ζ_t and distort either the signal structure or the perceived transition, raising new identification

questions.

- *Endogenous-object distortions:* extend the analysis to distortions applied directly to expectations of endogenous variables, where self-referential fixed-point issues arise and different tools may be required.
- *Policy counterfactuals and welfare:* quantify how observationally similar belief distortions change the mapping from estimated models to policy counterfactuals and welfare analysis, especially under regime changes.

4 Analytical Applications

In this section, we translate the abstract identification logic into concrete implications for prominent belief distortions used in applied DSGE work. Each application proceeds in the same steps: (i) specify a structured restriction $\Psi^S(\vartheta, \theta)$ that nests rational expectations at $\theta = 0$; (ii) compute the local *tangent direction* $\Delta_i = \partial_{\theta_i} \Psi^S|_{\theta=0}$ that isolates the primitive belief error; (iii) map this tangent direction into the *equilibrium wedge direction* via the wedge loading matrix $\Omega(\vartheta, \theta)$; (iv) use Theorem 1 to read off the resulting pole-order implication for local (non-)confounding with true persistence.

For comparability across distortions, we start from the benchmark in which a vector of persistent exogenous fundamentals $u_t \in \mathbb{R}^{n_\rho}$ follows a diagonal AR(1) under the actual law:

$$u_{t+1} = \operatorname{diag}(\rho_1, \ldots, \rho_{n_\rho})\, u_t + \Sigma_u \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \text{i.i.d.}(0, I), \qquad |\rho_k| < 1, \tag{33}$$

where $\Sigma_u = \operatorname{diag}(\sigma_1, \ldots, \sigma_{n_\rho})$ collects shock volatilities. When no state augmentation is needed (as under OPB), we set the exogenous Markov state equal to fundamentals, $\zeta_t := u_t$. When a distortion depends on realized surprises (as under diagnostic expectations), we augment the Markov state accordingly; in that case we will use $\zeta_t = (u_t, \nu_t)'$ with $\nu_t := \Sigma_u \varepsilon_t$.

Finally, recall the wedge loading matrix from the objective-expectations representation,

$$\Omega(\vartheta, \theta) = A_{+1} P_\zeta^S(\vartheta, \theta)\big(\Psi^S(\vartheta, \theta) - \Psi^A(\vartheta)\big),$$

and the benchmark $\Psi^S(\vartheta_0, 0) = \Psi^A(\vartheta_0)$, implying $\Omega(\vartheta_0, 0) = 0$. At this benchmark, the

product rule yields the simplification used repeatedly below:

$$\left.\frac{\partial \Omega(\vartheta,\theta)}{\partial \theta_i}\right|_{\theta=0} = A_{+1} P_\zeta^A \left.\frac{\partial \Psi^S(\vartheta,\theta)}{\partial \theta_i}\right|_{\theta=0}, \qquad P_\zeta^A := P_\zeta^S(\vartheta_0, 0).$$

4.1 Identifying Over-Persistence Bias

We begin with over-persistence bias (OPB), the simplest benchmark distortion. In this specification, agents correctly perceive the state dimension and shock structure, but misperceive the autoregressive propagation of fundamentals. Because OPB only perturbs the perceived persistence of existing AR(1) drivers, it requires *no* state augmentation: for this subsection we set $\zeta_t := u_t$, so that $\Psi^A(\rho) = \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho})$ as in (33). The stationarity restriction $|\rho_k + \theta_k^{\mathrm{OPB}}| < 1$ ensures beliefs remain coherent and stable.

PROPOSITION 3 (OPB: tangent direction, wedge derivative, and pole order) *Consider the diagonal AR(1) exogenous law in Assumption 1, $\Psi^A(\rho) = \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho})$. Let over-persistence bias (OPB) be parameterized as a misspecification of the perceived law of motion*

$$\Psi^S(\rho, \theta^{\mathrm{OPB}}) = \mathrm{diag}(\rho_1 + \theta_1^{\mathrm{OPB}}, \ldots, \rho_{n_\rho} + \theta_{n_\rho}^{\mathrm{OPB}}), \qquad |\rho_k + \theta_k^{\mathrm{OPB}}| < 1,$$

so that rational expectations are nested at $\theta^{\mathrm{OPB}} = 0$.

Fix an evaluation point $(\vartheta_0, 0)$ with $\vartheta_0 = (\rho_0, \psi_0)$ and $\Psi^S(\rho_0, 0) = \Psi^A(\rho_0)$.

1. **Tangent direction.** *For each component θ_k^{OPB},*

$$\Delta_k^{\mathrm{OPB}} := \left.\frac{\partial \Psi^S(\rho_0, \theta^{\mathrm{OPB}})}{\partial \theta_k^{\mathrm{OPB}}}\right|_{\theta^{\mathrm{OPB}}=0} = E_{kk},$$

   where E_{kk} is the matrix with a 1 in entry (k,k) and zeros elsewhere.

2. **First-order wedge direction at the evaluation point (objective-expectations representation).** *Let $\Omega(\vartheta, \theta^{\mathrm{OPB}})$ denote the wedge loading matrix,*

$$\Omega(\vartheta, \theta^{\mathrm{OPB}}) = A_{+1} P_\zeta^S(\vartheta, \theta^{\mathrm{OPB}}) \big(\Psi^S(\rho, \theta^{\mathrm{OPB}}) - \Psi^A(\rho)\big), \qquad P_\zeta^A := P_\zeta^S(\vartheta_0, 0).$$

   Then, at the evaluation point,

$$\left.\frac{\partial \Omega(\vartheta_0, \theta^{\mathrm{OPB}})}{\partial \theta_k^{\mathrm{OPB}}}\right|_{\theta^{\mathrm{OPB}}=0} = A_{+1} P_\zeta^A E_{kk}.$$

3. **Spectral signature (pole order).** *Under the conditions of Theorem 1, the complex extension* $\mathscr{S}(z;\vartheta,\theta^{\mathrm{OPB}})$ *satisfies*

$$\frac{\partial}{\partial\theta_k^{\mathrm{OPB}}}\mathscr{S}(z;\vartheta_0,0) \text{ has at most a simple pole at } z=\rho_{0,k},$$

whereas

$$\frac{\partial}{\partial\rho_k}\mathscr{S}(z;\vartheta_0,0) \text{ has a genuine second-order pole at } z=\rho_{0,k}.$$

In particular, OPB can locally be separately identified from true persistence ρ_k *when the full spectrum is used as the observable object.*

Proposition 3 isolates three conceptually distinct objects: (i) the primitive cognitive perturbation to the perceived law of motion, (ii) the induced first-order expectations wedge in general equilibrium, and (iii) the resulting econometric signature that separates OPB from true persistence when the full spectrum is used.

Part 1 (tangent direction). The tangent direction $\Delta_k^{\mathrm{OPB}} = E_{kk}$ makes precise what OPB changes *and what it does not.* It perturbs only the perceived own-persistence of the k-th exogenous driver and does not introduce mixing across drivers. This matters because it guarantees that OPB is a localized belief distortion at the level of Ψ^S: any cross-equation or cross-variable effects that appear in the data must come from the general equilibrium mapping, not from the OPB parameterization itself.

Part 2 (first-order wedge direction). A belief distortion matters macroeconomically only insofar as it changes equilibrium behavior. The wedge representation delivers a transparent mapping from the cognitive perturbation to an observable GE object. At the rational-expectations benchmark $(\vartheta_0, 0)$, the wedge loading matrix satisfies

$$\left.\frac{\partial\Omega(\vartheta_0,\theta^{\mathrm{OPB}})}{\partial\theta_k^{\mathrm{OPB}}}\right|_0 = A_{+1}P_\zeta^A E_{kk}.$$

This expression decomposes the first-order macro footprint of OPB into three structural components: E_{kk} selects the misperceived exogenous driver, P_ζ^A maps that driver into equilibrium endogenous outcomes under the benchmark policy function, and A_{+1} selects the forward-looking equilibrium conditions in which the distorted forecast enters. In particular, whenever $A_{+1}P_\zeta^A E_{kk} \neq 0$, OPB generates a nontrivial first-order expectations wedge in the objective-expectations representation. If $A_{+1}P_\zeta^A E_{kk} = 0$, the OPB wedge is first-order irrelevant at

the benchmark (though higher-order effects may still arise away from the benchmark through changes in the policy function).

Part 3 (pole order and non-confounding with ρ). The econometric payoff is that OPB cannot masquerade as true persistence when the observable object is the full spectrum. True persistence ρ_k enters the *actual* transition $\Psi^A(\rho)$ and therefore affects the exogenous resolvent factor $(I - \Psi^A z^{-1})^{-1}$ itself; differentiating with respect to ρ_k necessarily differentiates this resolvent and produces a genuine second-order singularity at $z = \rho_{0,k}$. By contrast, OPB perturbs only agents' perceived transition and enters the objective representation through a wedge term; under block-triangularity, $\Psi^A(\rho)$ is invariant to θ^{OPB}, so $\partial/\partial\theta_k^{\mathrm{OPB}}$ cannot differentiate the exogenous resolvent and cannot generate $(1 - \rho_{0,k} z^{-1})^{-2}$ terms. This pole-order contrast formalizes the intuitive economic dichotomy: changes in ρ_k alter the economy's physical propagation of shocks, while OPB changes how agents *forecast* that propagation. Because these two perturbations generate different spectral footprints, they are locally separately identified when the full spectrum is used.

4.2 Identifying Diagnostic Expectations

We now turn to *diagnostic expectations* (DE) in the sense of Bordalo, Gennaioli, and Shleifer (2018), which capture overreaction to recent realized surprises. Unlike OPB, DE makes current forecasts depend on a *realized innovation* state. To represent this dependence within a coherent Markov law of motion, we augment the exogenous state.

Let $u_t \in \mathbb{R}^{n_\rho}$ follow the diagonal AR(1) benchmark under the actual law, as in (33), and define the realized one-step innovation ("surprise") as

$$\nu_t := \Sigma_u \varepsilon_t = u_t - \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho})\, u_{t-1}.$$

We augment the exogenous Markov state as $\zeta_t := (u_t, \nu_t)' \in \mathbb{R}^{2n_\rho}$. Under the actual law, the augmented state follows the VAR(1)

$$\zeta_{t+1} = \Psi^A(\rho)\, \zeta_t + \Lambda^A \varepsilon_{t+1}, \qquad \Psi^A(\rho) = \begin{pmatrix} \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho}) & 0 \\ 0 & 0 \end{pmatrix}, \qquad \Lambda^A = \begin{pmatrix} \Sigma_u \\ \Sigma_u \end{pmatrix}. \quad (34)$$

Note that the innovation component ν_t has no own-persistence under the actual law (lower block row is zero).

We parameterize DE as an overweighting of the most recent realized surprise in the conditional mean of u_{t+1}:

$$E_t^S[u_{t+1}] = \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho})\, u_t + \Theta^{\mathrm{DE}}\, \nu_t, \qquad \Theta^{\mathrm{DE}} = \mathrm{diag}(\theta_1^{\mathrm{DE}}, \ldots, \theta_{n_\rho}^{\mathrm{DE}}), \qquad \theta_k^{\mathrm{DE}} \geq 0, \quad (35)$$

which nests rational expectations at $\theta^{\mathrm{DE}} = 0$. By design, agents do *not* attribute own-persistence to the surprise state, i.e. $E_t^S[\nu_{t+1}] = 0$. Hence DE admits a coherent subjective VAR(1) representation under the augmented state:

$$E_t^S[\zeta_{t+1}] = \Psi^S(\rho, \theta^{\mathrm{DE}})\, \zeta_t, \qquad \Psi^S(\rho, \theta^{\mathrm{DE}}) = \begin{pmatrix} \mathrm{diag}(\rho_1, \ldots, \rho_{n_\rho}) & \Theta^{\mathrm{DE}} \\ 0 & 0 \end{pmatrix}, \qquad \Psi^S(\rho, 0) = \Psi^A(\rho).$$

Substituting $\nu_t = u_t - \mathrm{diag}(\rho) u_{t-1}$ into (35) yields the equivalent AR(2)-type forecast rule in levels:

$$E_t^S[u_{t+1}] = \big(\mathrm{diag}(\rho) + \Theta^{\mathrm{DE}}\big) u_t - \Theta^{\mathrm{DE}}\, \mathrm{diag}(\rho)\, u_{t-1}.$$

With this coherent subjective law of motion in place, we can evaluate the tangent direction, wedge derivative, and pole-order signature implied by DE.

PROPOSITION 4 (Diagnostic Expectations: tangent direction, wedge derivative, and pole order) *Fix an evaluation point $(\vartheta_0, 0)$ with $\vartheta_0 = (\rho_0, \psi_0)$ and $\Psi^S(\rho_0, 0) = \Psi^A(\rho_0)$.*

1. **Tangent direction.** *For each component θ_k^{DE},*

$$\Delta_k^{\mathrm{DE}} := \left. \frac{\partial \Psi^S(\rho_0, \theta^{\mathrm{DE}})}{\partial \theta_k^{\mathrm{DE}}} \right|_{\theta^{\mathrm{DE}}=0} = E_{k,\, n_\rho + k},$$

where $E_{k, n_\rho + k}$ is the $(2n_\rho) \times (2n_\rho)$ matrix with a 1 in row k and column $n_\rho + k$ and zeros elsewhere.

2. **First-order wedge direction at the evaluation point.** *Let $\Omega(\vartheta, \theta^{\mathrm{DE}})$ denote the wedge loading matrix. At the evaluation point,*

$$\left. \frac{\partial \Omega(\vartheta_0, \theta^{\mathrm{DE}})}{\partial \theta_k^{\mathrm{DE}}} \right|_{\theta^{\mathrm{DE}}=0} = A_{+1} P_\zeta^A\, E_{k,\, n_\rho + k},$$

where $P_\zeta^A := P_\zeta^S(\vartheta_0, 0)$.

3. **Spectral signature and Non-Confounding.** *Under the conditions of Theorem 1*

(applied to the augmented $\Psi^A(\rho)$ above), the complex extension $\mathscr{S}(z;\vartheta,\theta^{\mathrm{DE}})$ satisfies:

$$\frac{\partial}{\partial\theta_k^{\mathrm{DE}}}\mathscr{S}(z;\vartheta_0,0) \text{ has at most a simple pole at } z=\rho_{0,k},$$

whereas the derivative with respect to true persistence satisfies:

$$\frac{\partial}{\partial\rho_k}\mathscr{S}(z;\vartheta_0,0) \text{ has a genuine second-order pole at } z=\rho_{0,k}.$$

In particular, diagnostic expectations cannot be locally be separately identified from true persistence ρ_k when the full spectrum is used as the observable object.

Proposition 4 parallels the OPB application but highlights a different cognitive mechanism. Where OPB perturbs perceived own-persistence, diagnostic expectations (DE) introduce a *news channel*: agents overweight the most recent realized innovation when forecasting future fundamentals. The proposition again separates (i) the primitive cognitive perturbation to the perceived transition law, (ii) the induced first-order GE wedge, and (iii) the resulting spectral signature that rules out local confounding with true persistence when the full spectrum is used.

Part 1 (tangent direction and what DE changes). The tangent direction $\Delta_k^{\mathrm{DE}} = E_{k,n_\rho+k}$ is an off-diagonal perturbation in the augmented subjective transition matrix. It maps the realized surprise component $\nu_{k,t}$ into the perceived conditional mean of the next-period fundamental $u_{k,t+1}$. Thus DE does not (at the parameterization level) "tilt" the perceived AR(1) coefficient on $u_{k,t}$ directly; instead it introduces history dependence by allowing the most recent innovation to act as a predictive state. Equivalently, DE can be written as an AR(2)-type forecast rule in u_t alone, which makes clear that DE operates through *state augmentation / lag structure* rather than through a pure shift in perceived persistence.

Part 2 (first-order wedge direction). The wedge representation shows how this localized misforecast translates into equilibrium behavior. At the rational-expectations benchmark $(\vartheta_0,0)$,

$$\left.\frac{\partial\Omega(\vartheta_0,\theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}}\right|_0 = A_{+1}P_\zeta^A\,E_{k,n_\rho+k}.$$

Using $E_{k,n_\rho+k} = e_k e'_{n_\rho+k}$, this can be written as

$$A_{+1}P_\zeta^A\,E_{k,n_\rho+k} = \left(A_{+1}P_\zeta^A e_k\right)e'_{n_\rho+k}.$$

This highlights the economic content: the diagnostic perturbation makes $\nu_{k,t}$ enter the equilib-

rium conditions only through the sensitivity of forward-looking constraints to the fundamental component $u_{k,t}$, summarized by $A_{+1}P_{\zeta}^{A}e_k$. Hence DE leaves a first-order macro footprint whenever $A_{+1}P_{\zeta}^{A}e_k \neq 0$, i.e. whenever the k-th fundamental matters for at least one forward-looking margin at the benchmark. If $A_{+1}P_{\zeta}^{A}e_k = 0$, DE is first-order irrelevant at the benchmark (though, as usual, higher-order effects may arise away from the benchmark through changes in the policy function).

Part 3 (pole order and non-confounding with ρ). The econometric separation again follows from block-triangularity and the invariance of the true exogenous transition $\Psi^{A}(\rho)$ to belief parameters. True persistence ρ_k enters the *actual* law of motion and therefore enters the exogenous resolvent $(I - \Psi^{A}z^{-1})^{-1}$; differentiating with respect to ρ_k differentiates this resolvent and generates a genuine second-order singularity at $z = \rho_{0,k}$. By contrast, θ_k^{DE} alters only how agents map the realized innovation state into the perceived conditional mean of $u_{k,t+1}$; it affects equilibrium through the wedge and through endogenous propagation, but it cannot differentiate the exogenous resolvent factor itself. Consequently, $\partial\mathscr{S}/\partial\theta_k^{\mathrm{DE}}$ can have at most a simple pole at $z = \rho_{0,k}$. This pole-order contrast formalizes the key distinction: ρ_k governs the physical compounding of the fundamental in the data-generating process, whereas DE governs a belief-based "news" loading that changes forecasts and hence actions without altering the true transition of the fundamental.

4.3 Over-Persistence Bias and Diagnostic Expectations combined

Over-persistence bias (OPB) and diagnostic expectations (DE) are a natural "stress test" for our omitted-channel logic because, at a high level, both distortions push in the same qualitative direction: they raise the perceived persistence of fundamentals. This creates a practical identification concern that is distinct from the pole-order separation in Theorem 1. Theorem 1 separates true persistence ρ from belief distortions θ in the correctly specified model, but it does *not* guarantee that two different belief distortions can be separated from each other using aggregate macro time series.

For applied work, the relevant question is often not only "can ρ be identified in principle," but also: *if the researcher estimates one persistence-type distortion at a time, can the included channel proxy for the omitted one?* To make this question transparent, we study OPB and DE at the level of the perceived law of motion for a single fundamental. This isolates a reduced-form source of confounding that arises *before* imposing any additional DSGE structure. Only

afterward should one ask whether a given DSGE model and dataset provide enough information to discipline the DE-specific component.

Corollary 2 (OPB and DE jointly: overlap in perceived persistence and separation via the innovation/lag coefficient) *Fix a component $k \in \{1, \dots, n_\rho\}$ and consider the AR(1) benchmark*

$$u_{k,t+1} = \rho_k u_{k,t} + \sigma_k \varepsilon_{k,t+1}, \qquad |\rho_k| < 1,$$

with realized surprise (innovation state) $\nu_{k,t} := u_{k,t} - \rho_k u_{k,t-1} = \sigma_k \varepsilon_{k,t}$. Let OPB and DE enter the subjective forecast jointly as

$$E_t^S[u_{k,t+1}] = (\rho_k + \theta_k^{\mathrm{OPB}})\, u_{k,t} + \theta_k^{\mathrm{DE}}\, \nu_{k,t}. \tag{36}$$

Then:

1. **AR(2) representation and the two reduced-form coefficients.** *Equation (36) is equivalent to*

$$E_t^S[u_{k,t+1}] = b_{0,k}\, u_{k,t} + b_{1,k}\, u_{k,t-1}, \qquad b_{0,k} = \rho_k + \theta_k^{\mathrm{OPB}} + \theta_k^{\mathrm{DE}}, \qquad b_{1,k} = -\rho_k\, \theta_k^{\mathrm{DE}}. \tag{37}$$

2. **Non-identification if the perceived forecast is restricted to depend only on $u_{k,t}$.** *Within the restricted perceived-forecast class*

$$E_t^S[u_{k,t+1}] = \tilde{\rho}_k\, u_{k,t} \quad (\textit{equivalently } b_{1,k} \equiv 0),$$

   OPB and DE are not separately identified from the coefficient on $u_{k,t}$: only the sum $\theta_k^{\mathrm{OPB}} + \theta_k^{\mathrm{DE}}$ is pinned down through $b_{0,k} - \rho_k$.

3. **Algebraic separation once the innovation/lag coefficient is retained.** *If $\rho_k \neq 0$, then $(b_{0,k}, b_{1,k})$ pins down $(\theta_k^{\mathrm{OPB}}, \theta_k^{\mathrm{DE}})$ uniquely via*

$$\theta_k^{\mathrm{DE}} = -\frac{b_{1,k}}{\rho_k}, \qquad \theta_k^{\mathrm{OPB}} = b_{0,k} - \rho_k - \theta_k^{\mathrm{DE}}.$$

Implications for applied estimation. Corollary 2 isolates a concrete confounding mechanism that is easy to miss in DSGE estimation: *OPB and DE share the same reduced-form direction in perceived persistence.* In the AR(2) representation (37), both distortions raise the

coefficient on the current level $u_{k,t}$ through $b_{0,k} - \rho_k = \theta_k^{\mathrm{OPB}} + \theta_k^{\mathrm{DE}}$. By contrast, only DE generates the additional innovation/lag coefficient $b_{1,k} = -\rho_k \, \theta_k^{\mathrm{DE}}$. This is the sharp sense in which OPB can act as a proxy for omitted DE (and vice versa): if the maintained perceived-law class effectively collapses to a perceived AR(1) in levels (so that $b_{1,k}$ is excluded or empirically irrelevant), then the data can at best identify a *composite* perceived-persistence distortion, not its decomposition into OPB versus DE.

The connection to Proposition 2 is direct. Estimating OPB-only (imposing $\theta_k^{\mathrm{DE}} = 0$) when the true perceived law has $\theta_k^{\mathrm{DE}} \neq 0$ forces the restricted model to approximate the true perceived dynamics using the remaining degrees of freedom. At the level of (37), the restriction $b_{1,k} \equiv 0$ removes the only reduced-form object that is specific to DE. When that happens, the best-fitting persistence-type distortion necessarily loads into $b_{0,k}$, and the estimated "persistence bias" must be interpreted as a mixture of OPB and DE rather than evidence for one mechanism in isolation.

What the corollary *does not* claim is that OPB and DE are always empirically separable once a DSGE model is imposed. Part (3) shows that separation is algebraically possible *within the perceived-law class* once $b_{1,k}$ is retained. Whether $b_{1,k}$ is actually disciplined by a given dataset depends on the full equilibrium mapping from perceived fundamentals into observables. In practice, the relevant diagnostic is therefore not only the posterior on $(\theta_k^{\mathrm{OPB}}, \theta_k^{\mathrm{DE}})$, but the implied posterior on the reduced-form pair $(b_{0,k}, b_{1,k})$.

Use case: estimating DE versus OPB in applied work. Suppose an applied researcher wants to quantify the relative importance of DE and OPB for a given fundamental $u_{k,t}$. Corollary 2 gives a simple organizing principle: in the joint OPB+DE specification, the perceived one-step-ahead forecast is summarized by

$$b_{0,k} = \rho_k + \theta_k^{\mathrm{OPB}} + \theta_k^{\mathrm{DE}}, \qquad b_{1,k} = -\rho_k \, \theta_k^{\mathrm{DE}}.$$

If the empirical design primarily disciplines perceived persistence in levels—i.e. it is informative about $b_{0,k}$ but weakly informative about $b_{1,k}$—then the data will tend to identify the composite distortion $b_{0,k} - \rho_k$ and leave the split between OPB and DE fragile. In that case, reporting separate posteriors for θ_k^{OPB} and θ_k^{DE} can be misleading unless it is accompanied by evidence that $b_{1,k}$ is actually pinned down.

A disciplined workflow is therefore to estimate OPB-only, DE-only, and OPB+DE jointly and to report the implied $(b_{0,k}, b_{1,k})$ objects. If the posterior for $b_{1,k}$ is wide, highly prior-

sensitive, or unstable across reasonable changes in observables, then the data are not sharply distinguishing DE from OPB, and the safest interpretation is in terms of the composite perceived-persistence deviation $b_{0,k} - \rho_k$. Conversely, if $b_{1,k}$ is sharply estimated and robust, then Part (3) implies that DE can be cleanly separated from OPB, and the decomposition has a transparent meaning: DE is precisely the component that loads on realized surprises (equivalently, that generates the lag/innovation coefficient), while OPB is the residual perceived-persistence shift net of that surprise channel.

Because $b_{1,k}$ is intrinsically tied to how agents map realized surprises into forecasts, data that load more directly on forecast revisions or forecast errors are natural candidates for disciplining this channel. In our framework, survey expectations (possibly at multiple horizons) provide precisely such information, because they link observables to subjective forecasts rather than only to realized macro outcomes.

4.4 A workflow for deriving distortion-specific implications

The analytical applications above are modular because, in our maintained class, belief distortions enter exclusively through the perceived transition $\Psi^S(\vartheta, \theta)$ for the (possibly augmented) exogenous state. As a result, a distortion can be analyzed through three objects: (i) a coherent perceived transition $\Psi^S(\vartheta, \theta)$ (after minimal augmentation if needed), (ii) its *tangent direction* at the rational-expectations benchmark,

$$\Delta_i \;:=\; \left.\frac{\partial \Psi^S(\vartheta, \theta)}{\partial \theta_i}\right|_{\theta=0},$$

and (iii) the induced *first-order wedge direction* in the objective-expectations representation,

$$\left.\frac{\partial \Omega(\vartheta, \theta)}{\partial \theta_i}\right|_{\theta=0} = A_{+1} P_\zeta^A \, \Delta_i, \qquad P_\zeta^A := P_\zeta^S(\vartheta_0, 0).$$

These objects deliver (a) the economic content of the distortion (what changes in forecasts), (b) its general-equilibrium footprint (how it enters equilibrium restrictions), and (c) its identification implications.

A practical derivation recipe is: (i) specify the primitive bias and augment ζ_t until the perceived law is VAR(1); (ii) write $\Psi^A(\rho)$ and $\Psi^S(\vartheta, \theta)$ on the *same realized* state and impose nesting at $\theta = 0$; (iii) compute tangents Δ_i; (iv) map Δ_i into wedge directions via $A_{+1} P_\zeta^A \Delta_i$ to check whether the distortion is locally macro-relevant; (v) interpret identification using our

two generic tools: pole-order separation for ρ versus θ (Theorem 1) and derivative-span overlap for θ_1 versus θ_2 (Proposition 2).

For additional worked templates—including narrative confusion, narrative shocks via spurious predictors, and cognitive discounting—see Appendix B.

5 Estimation Framework

The identification results of Section 3 are population statements about the spectral density of observables. By the Fourier inversion theorem, the spectral density and the full autocovariance sequence stand in a one-to-one correspondence under standard regularity conditions: two processes have the same spectral density if and only if they have the same autocovariances at all lags. Since the Gaussian likelihood in the linear state-space model is fully determined by the autocovariance sequence, identification established in the spectral domain carries over directly to likelihood-based and Bayesian estimation, and to GMM targeting second moments.

5.1 Solving and estimating the model

Fix (ϑ, θ), where ϑ collects structural parameters including the actual transition $\Psi^A(\vartheta)$, and θ collects belief-distortion parameters entering only through the perceived transition $\Psi^S(\vartheta, \theta)$. Consider the distorted-expectations structural system

$$A_{-1}x_{t-1} + A_0 x_t + A_{+1}E_t^S[x_{t+1}] + B\zeta_t = 0, \qquad \zeta_{t+1} = \Psi^A(\vartheta)\zeta_t + \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \text{i.i.d.}(0, \Sigma_\varepsilon).$$

The realized exogenous state ζ_t follows the *actual* transition $\Psi^A(\vartheta)$, while subjective forecasts are formed under the *perceived* transition $\Psi^S(\vartheta, \theta)$.

Step 1: solve the subjective economy. Given (ϑ, θ), agents solve the DSGE under subjective beliefs using $\Psi^S(\vartheta, \theta)$ to form conditional forecasts. A standard linear solver delivers a subjective decision rule

$$x_t = P_x^S(\vartheta, \theta)\, x_{t-1} + P_\zeta^S(\vartheta, \theta)\, \zeta_t. \tag{38}$$

Step 2: construct the expectations wedge. Using (38) and the contrast between perceived and actual transitions, the one-step-ahead forecast wedge is

$$W_t(\vartheta, \theta) := E_t^S[x_{t+1}] - E_t[x_{t+1}] = P_\zeta^S(\vartheta, \theta)\big(\Psi^S(\vartheta, \theta) - \Psi^A(\vartheta)\big)\zeta_t. \tag{39}$$

The wedge-loading matrix $\Omega(\vartheta,\theta) = A_{+1}P^S_\zeta(\vartheta,\theta)(\Psi^S(\vartheta,\theta) - \Psi^A(\vartheta))$ rewrites the system using only the objective expectation operator:

$$A_{-1}x_{t-1} + A_0 x_t + A_{+1}E_t[x_{t+1}] + \big(B + \Omega(\vartheta,\theta)\big)\zeta_t = 0.$$

Step 3: construct the state-space system. Combining the subjective decision rule with the actual transition yields

$$s_t = F(\vartheta,\theta)\, s_{t-1} + G(\vartheta,\theta)\, \varepsilon_t, \qquad y_t = H s_t + \eta_t, \tag{40}$$

with $\eta_t \sim$ i.i.d.$(0,\Sigma_\eta)$ and $\eta_t \perp \varepsilon_\tau$ for all τ. By Lemma 1, the exogenous block of $F(\vartheta,\theta)$ equals $\Psi^A(\vartheta)$ and is invariant to θ — the property that underlies the identification result of Theorem 1.

Step 4: evaluate the likelihood and run Bayesian estimation. Given (40), the likelihood is evaluated with a standard Kalman filter. Bayesian estimation combines the likelihood with priors on (ϑ,θ) and draws from the posterior via Metropolis–Hastings or SMC.

5.2 Survey expectations as noisy observables

Survey expectations can be incorporated as additional observables that load on model-implied subjective forecasts, providing direct empirical discipline on the belief parameters θ. Let q_t^H denote a survey measure at time t over a horizon of H periods. The measurement equation is

$$q_t^H = \delta_0 + E_t^S\left(\sum_{h=1}^{H} v_{t+h}\right) + e_t, \qquad e_t \sim \text{i.i.d.}(0,\sigma_e^2), \tag{41}$$

where v_t is the model counterpart, δ_0 captures a level wedge between survey and model units, and e_t is measurement noise. Because the subjective equilibrium is linear, subjective multi-step forecasts are linear in the state: $E_t^S[s_{t+h}] = (\tilde{F}^S(\vartheta,\theta))^h s_t$, where $\tilde{F}^S(\vartheta,\theta)$ is the one-step transition for the subjective conditional mean. Equation (41) therefore becomes a standard linear measurement row in the Kalman filter:

$$q_t^H = \delta_0 + R(\vartheta,\theta)\, s_t + e_t, \qquad R(\vartheta,\theta) := \sum_{h=1}^{H} \ell_v' H\big(\tilde{F}^S(\vartheta,\theta)\big)^h,$$

where ℓ_v selects the relevant variable from y_t. Survey expectations are treated as noisy observables that discipline belief parameters through the likelihood, not as hard targets imposed on the model.

5.3 Specification choices

Model. We estimate the medium-scale New Keynesian model of Smets and Wouters (2007), which features Calvo price and wage stickiness with partial indexation, habit formation in consumption, investment adjustment costs, variable capital utilization, and a Taylor rule for monetary policy. To account for the non-stationarity of the macroeconomic observables, the model incorporates a deterministic growth path driven by labor-augmenting technological progress, which implies a common constant trend growth rate for the real variables. The model is driven by seven structural shocks: a productivity shock (ε^a), a risk premium shock (ε^b), a government spending shock (ε^g), an investment-specific technology shock (ε^i), a monetary policy shock (ε^r), a price markup shock (ε^p), and a wage markup shock (ε^w). Four parameters are fixed at standard values: the depreciation rate $\delta = 0.025$, the gross wage markup $\phi_w = 1.5$, the government spending share $\bar{g}/\bar{y} = 0.18$, and the curvature parameters for the Kimball aggregators $\varepsilon_p = \varepsilon_w = 10$. All remaining parameters are estimated.

Data and survey measurement. The model is estimated on US data over the sample 1966Q1–2004Q4, using the seven standard Smets and Wouters (2007) macroeconomic observables: output growth ($\Delta \log \text{GDP}$), consumption growth, investment growth, wage growth, log hours worked, inflation, and the federal funds rate. To separately identify physical macroeconomic persistence from subjective belief distortions, we augment this standard dataset with survey expectations from the Survey of Professional Forecasters (SPF). We use SPF consensus forecasts (median) as noisy observables for model-implied subjective expectations, following the framework of Section 5.2. Six survey series are included: one-quarter-ahead and four-quarter-ahead forecasts of output growth ($y^{exp,obs}$, $y^{exp,obs}_{t+4}$), inflation ($\pi^{exp,obs}$, $\pi^{exp,obs}_{t+4}$), and the federal funds rate ($r^{exp,obs}$, $r^{exp,obs}_{t+4}$).

Each survey series enters through a measurement equation of the form (41) with an estimated intercept δ_0 and an estimated measurement error standard deviation σ_e. Crucially, the intercept parameters δ_0 are included to absorb the unconditional means of the raw survey data. By capturing the underlying deterministic trend growth rates and steady-state levels of the macro variables, these intercepts ensure that the model's stationary, mean-zero theoretical ex-

pectations map correctly to the non-stationary empirical forecasts without spuriously loading into the dynamic behavioral parameters. The intercepts δ_0 are given Normal priors centered at plausible scaling values designed to match these trend growth rates: $\mathcal{N}(0.5, 0.2^2)$ for output forecasts at both horizons (except four-quarter-ahead output which uses $\mathcal{N}(1.5, 0.35^2)$), $\mathcal{N}(0.6, 0.2^2)$ for inflation forecasts, and $\mathcal{N}(1.5, 0.5^2)$ for interest rate forecasts. Measurement error standard deviations are given InvGamma$(0.1, 2)$ priors.

Moving-average exclusion. The original Smets and Wouters (2007) specification includes moving-average (MA) terms in the price and wage markup shock processes, governed by coefficients μ_p and μ_w. We exclude these MA terms in all estimated specifications that include survey data. The reason is a severe identification failure when the strict rational expectations (RE) model is forced to fit the inherently smooth and sluggish survey data. Because RE restricts subjective persistence to perfectly equal physical persistence, the optimizer artificially inflates the physical AR roots to minimize survey forecast errors, but then weaponizes the MA terms as statistical degrees of freedom to artificially dampen the resulting excess macroeconomic volatility.

For the price markup shock, the Komunjer and Ng (2011) rank diagnostics confirm this failure directly: the estimated AR root ($\hat{\rho}_p = 0.980$) and MA root ($\hat{\mu}_p = 0.975$) converge to nearly identical values, yielding a collinearity measure of 0.999. This triggers perfect pole-zero cancellation, mathematically violating the minimality conditions for the state-space system and effectively replacing the persistent markup shock with near-white-noise dynamics. Similarly, for the wage markup shock, the optimizer forces the MA root to strictly exceed the AR root ($\hat{\mu}_w > \hat{\rho}_w$), which pathologically inverts the macroeconomic transmission of the shock on impact. Excluding MA terms averts this state-space degeneration, restores minimality, passes the Komunjer and Ng (2011) rank conditions, and forces the model to resolve the tension between subjective and physical persistence structurally rather than through statistical contortions.

Nested specifications and observational equivalence. We estimate five nested specifications to systematically disentangle physical macroeconomic frictions from cognitive expectation biases. Two specifications exclude SPF data entirely: the canonical Smets and Wouters (2007) baseline (including MA terms) and our joint behavioral model (OPB+DE, no SPF). Comparing these establishes a baseline of *observational equivalence*: without survey data to anchor beliefs, behavioral distortions and physical persistence are almost perfectly confounded. The likelihood function cannot distinguish between them, causing the behavioral model to mimic the high-amplitude dynamics of the canonical baseline.

Our main three specifications include SPF data (and exclude MA terms): strict RE, diagnostic expectations only (DE), and over-persistence bias plus diagnostic expectations jointly (OPB+DE). This nested structure exposes sheds light on potential omitted variable bias in standard RE models. By explicitly modeling OPB, the joint framework absorbs the subjective sluggishness of the consensus of professional forecasters.

Priors and estimation. Priors for the structural parameters follow Smets and Wouters (2007) closely. Shock persistence parameters ρ_j receive Beta$(0.5, 0.20)$ priors, with upper bounds tightened to 0.98 for the TFP, government spending, and wage markup shocks to avoid the near-unit-root region that creates estimation difficulties in the presence of belief distortions. Shock standard deviations receive InvGamma$(0.1, 2)$ priors. Key structural frictions receive standard priors: investment adjustment costs $\mathcal{N}(4, 1.5^2)$, habit formation Beta$(0.7, 0.1^2)$, Calvo price stickiness Beta$(0.66, 0.05^2)$, Calvo wage stickiness Beta$(0.5, 0.1^2)$, Taylor rule inflation feedback $\mathcal{N}(1.5, 0.25^2)$, and interest rate smoothing Beta$(0.75, 0.10^2)$.

For the belief distortion parameters, the over-persistence bias parameter θ^{OPB} and the diagnostic expectations parameter θ^{DE} receive tight $\mathcal{N}(0, 0.05^2)$ and $\mathcal{N}(0, 0.1^2)$ priors, respectively. These priors are deliberately conservative: by centering them strictly at zero, we impose no ex-ante assumption that cognitive distortions are present, requiring the survey data to provide positive evidence to pull the posterior away from the strict RE null hypothesis. Both parameters are bounded within $[-0.25, 0.25]$ for α_{OPB} and $[-0.5, 0.5]$ for θ_{DE}, ensuring the subjective persistence $\rho_k + \theta^{OPB}$ remains dynamically stable inside the unit circle.

Posterior distributions for all specifications are computed via the Metropolis-Hastings algorithm with two parallel chains of 1,000,000 draws each, discarding the first 20 percent as burn-in. The posterior mode is computed first and used as the starting point for the Markov chain. Komunjer and Ng (2011) identification diagnostics are computed at the posterior mode for each specification to verify that the model satisfies local minimal state-space identification conditions before interpreting posterior estimates.

6 Empirical Illustration: Smets and Wouters (2007) with Distorted Expectations

We illustrate the framework by estimating the Smets and Wouters (2007) model under three nested survey-disciplined specifications: rational expectations (RE), diagnostic expectations

only (DE), and over-persistence bias plus diagnostic expectations jointly (OPB+DE). The goal is intentionally narrow: to show that multi-distortion estimation is feasible in a policy-grade DSGE environment, and to illustrate the practical implications of the identification and omitted-channel bias results developed in Sections 3–4. Specification choices — the MA exclusion, the survey observables, the prior design, and the nested structure — are described in Section 5.

Table 1: Posterior estimates across survey-disciplined specifications

	RE		DE only		OPB+DE	
Parameter	Mean	90% HPD	Mean	90% HPD	Mean	90% HPD
Belief distortion parameters						
θ^{OPB}	—	—	—	—	0.017	[0.001, 0.033]
θ^{DE}	—	—	0.044	[−0.008, 0.096]	0.037	[−0.016, 0.087]
Shock persistence parameters						
ρ_a (TFP)	0.964	[0.950, 0.980]	0.957	[0.944, 0.970]	0.942	[0.922, 0.962]
ρ_b (risk premium)	0.572	[0.226, 0.852]	0.280	[0.167, 0.389]	0.252	[0.154, 0.349]
ρ_g (gov. spending)	0.969	[0.959, 0.980]	0.970	[0.961, 0.980]	0.959	[0.942, 0.977]
ρ_i (investment)	0.632	[0.532, 0.730]	0.644	[0.547, 0.742]	0.626	[0.532, 0.721]
ρ_r (monetary)	0.083	[0.023, 0.140]	0.079	[0.021, 0.133]	0.075	[0.019, 0.130]
ρ_p (price markup)	0.742	[0.690, 0.795]	0.747	[0.693, 0.800]	0.719	[0.660, 0.779]
ρ_w (wage markup)	0.977	[0.972, 0.980]	0.975	[0.970, 0.980]	0.964	[0.950, 0.980]
Key structural parameters						
λ (habit)	0.521	[0.301, 0.769]	0.729	[0.647, 0.822]	0.744	[0.667, 0.822]
φ (adj. costs)	6.813	[3.915, 9.398]	7.513	[5.558, 9.372]	7.636	[5.844, 9.376]
σ_c (risk aversion)	1.458	[1.153, 1.790]	1.452	[1.215, 1.687]	1.435	[1.193, 1.655]
ξ_p (Calvo prices)	0.879	[0.858, 0.903]	0.874	[0.851, 0.899]	0.884	[0.864, 0.903]
ξ_w (Calvo wages)	0.421	[0.307, 0.528]	0.506	[0.429, 0.583]	0.522	[0.452, 0.593]
ι_p (price index.)	0.043	[0.013, 0.070]	0.045	[0.014, 0.074]	0.046	[0.015, 0.076]
ι_w (wage index.)	0.484	[0.275, 0.699]	0.456	[0.251, 0.660]	0.456	[0.225, 0.682]
r_π (Taylor infl.)	1.961	[1.679, 2.253]	1.957	[1.699, 2.215]	1.963	[1.736, 2.213]
ϕ_r (Taylor smooth.)	0.857	[0.829, 0.885]	0.846	[0.820, 0.870]	0.844	[0.820, 0.868]
α (capital share)	0.178	[0.146, 0.210]	0.181	[0.154, 0.209]	0.179	[0.146, 0.210]
Model fit						
Log data density	−1076.7		−1079.0		−1079.7	

Notes: All three specifications use SPF survey data at one-quarter and four-quarter horizons and exclude moving-average terms. RE imposes $\theta^{\text{OPB}} = \theta^{\text{DE}} = 0$. DE only imposes $\theta^{\text{OPB}} = 0$. 90% HPD denotes the 90% highest posterior density interval. Log data density is computed via the modified harmonic mean estimator. The upper bound on shock persistence parameters is set to 0.98 for all specifications; the risk premium posterior is therefore unconstrained at its estimated values.

6.1 Posterior estimates

Table 1 reports posterior means and 90% highest posterior density intervals for the belief distortion parameters, the shock persistence parameters, key structural parameters, and log

data densities across the three survey-disciplined specifications.

Belief distortion parameters. Estimated distortions at the representative agent level using SPF consensus forecasts are small. In the OPB+DE specification, the posterior mean for θ^{OPB} is 0.017 with a 90% HPD interval of $[0.001, 0.033]$, and the posterior mean for θ^{DE} is 0.037 with an interval of $[-0.016, 0.087]$. In the DE-only specification, θ^{DE} rises modestly to 0.044 with an interval of $[-0.008, 0.096]$. The OPB interval just excludes zero from below, providing marginal evidence for a positive over-persistence component; the DE interval straddles zero in both specifications, consistent with the view that consensus professional forecasts do not exhibit strong aggregate overreaction. These estimates are consistent with the representative agent interpretation: the distortions that survive aggregation to the consensus level are small, even if individual-level overreaction may be larger (Bordalo, Gennaioli, Ma, and Shleifer 2020).

The shift in θ^{DE} from 0.044 in the single-distortion specification to 0.037 in the joint specification illustrates the mild within-belief confounding predicted by Proposition 2: when OPB is added alongside DE, the DE estimate falls because OPB absorbs part of the composite perceived-persistence channel. The shift is modest here because both distortion parameters are small, but the direction is exactly as the theory predicts.

Although the point estimates are small, they are not economically irrelevant. An over-persistence bias of $\theta^{\text{OPB}} = 0.017$ applied to a shock with AR(1) persistence $\rho = 0.96$ raises agents' perceived four-quarter cumulative response by approximately eight percentage points relative to the rational benchmark, compounding nonlinearly over the policy horizon.

Shock persistence parameters. The most striking finding in the posterior table is the behavior of the risk premium shock persistence ρ_b. Under RE, the posterior mean is 0.572; under OPB+DE it falls to 0.252; under DE-only it falls to 0.280. This is a drop of more than 0.3 in posterior mean persistence — a large and economically significant shift. It is a direct empirical illustration of Proposition 1: the RE model, unable to attribute any demand persistence to the belief channel, inflates the estimated persistence of the risk premium shock to fit the low-frequency demand dynamics in the data. When belief distortions are allowed, the model redistributes that persistence — the structural risk premium shock is estimated to be substantially less persistent, and the belief channel absorbs the dynamics that the RE model attributed to structural shock persistence.

For TFP (ρ_a) and government spending (ρ_g), the substitution is more modest: ρ_a falls from 0.964 under RE to 0.942 under OPB+DE, and ρ_g from 0.969 to 0.959. The wage markup

persistence ρ_w falls from 0.977 to 0.964. For monetary policy and price markup shocks — which have low or moderate persistence and for which the belief channels have little to absorb — the posteriors are essentially unchanged. This shock-specificity is itself informative: belief distortions are not acting as a universal free parameter absorbing all model misfit, but are operative precisely for the persistent shocks where perceived persistence has the most room to deviate from the rational benchmark.

Structural parameters. Structural parameters are remarkably stable across specifications. Investment adjustment costs, the intertemporal elasticity of substitution, price stickiness, wage stickiness, price and wage indexation, capital utilization, and Taylor rule parameters are all close across RE, DE-only, and OPB+DE. The one notable exception is habit formation, which rises from 0.521 under RE to 0.744 under OPB+DE and 0.729 under DE-only. This movement is consistent with a substitution in the opposite direction: as belief distortions absorb some of the low-frequency demand dynamics that the risk premium shock was generating under RE, habit formation compensates by generating additional internal persistence in consumption dynamics. This is not a failure of identification but an illustration of the interaction between belief channels and structural propagation mechanisms — the two can partially substitute for each other in matching second moments.

Model fit. Log data densities are -1076.7 for RE, -1079.0 for DE-only, and -1079.7 for OPB+DE. The RE model fits marginally better by this metric, with the distortion models within approximately three log points. Differences of this magnitude do not constitute strong evidence in favor of any specification: the data are consistent with all three models, and the choice between them cannot be resolved by fit alone. This model uncertainty itself has stakes, as we document in the following subsections.

6.2 Impulse response analysis

Figures 1–2 display posterior mean impulse responses under RE, DE-only, and OPB+DE for a productivity shock (ε^a), and a government spending shock. These are taken as prototypes for a supply-side productivity, and a demand-side shock — and are the two for which belief distortions produce the most economically meaningful differences across specifications. Impulse responses for the remaining four shocks are reported in Appendix C.

Productivity shock (Figure 1). The productivity shock illustrates the persistence substitution mechanism most clearly. Under RE, the responses of output, consumption, and real

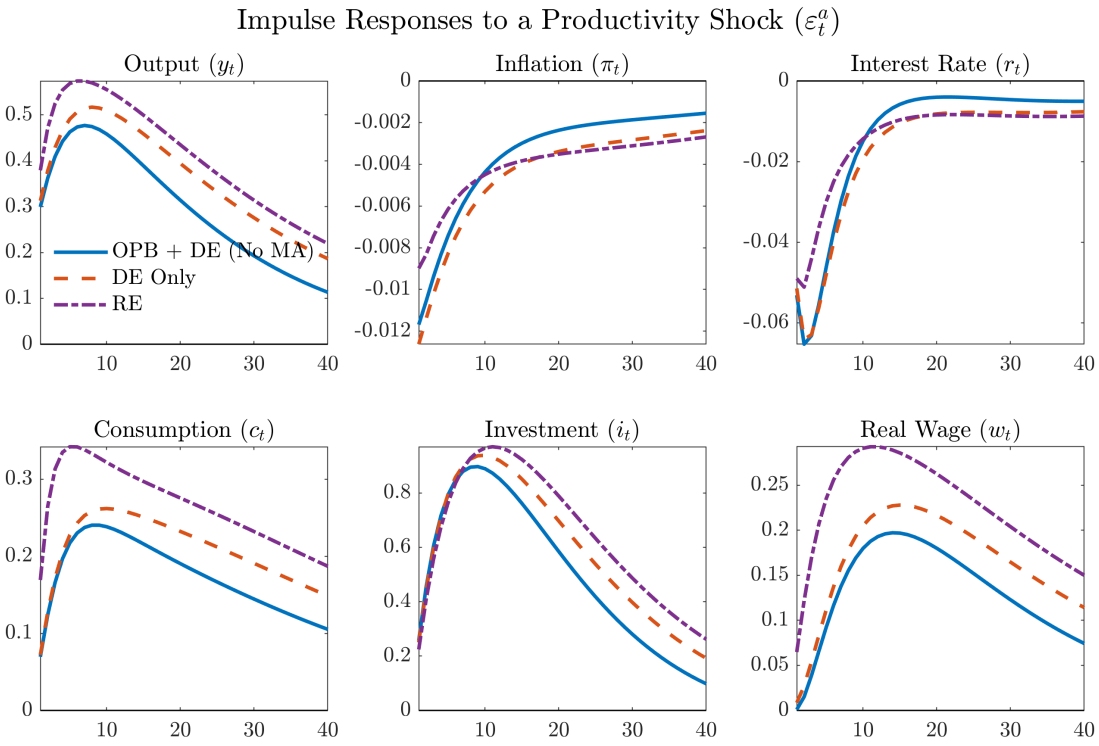


Figure 1: Impulse responses to a productivity shock (ε_t^a).

Notes: Posterior mean responses under three survey-disciplined specifications. All specifications use SPF data at one-quarter and four-quarter horizons and exclude moving-average terms. RE imposes $\theta^{\text{OPB}} = \theta^{\text{DE}} = 0$ (purple dashed); DE only imposes $\theta^{\text{OPB}} = 0$ (orange dashed); OPB+DE estimates both distortions jointly (blue solid). Horizon in quarters.

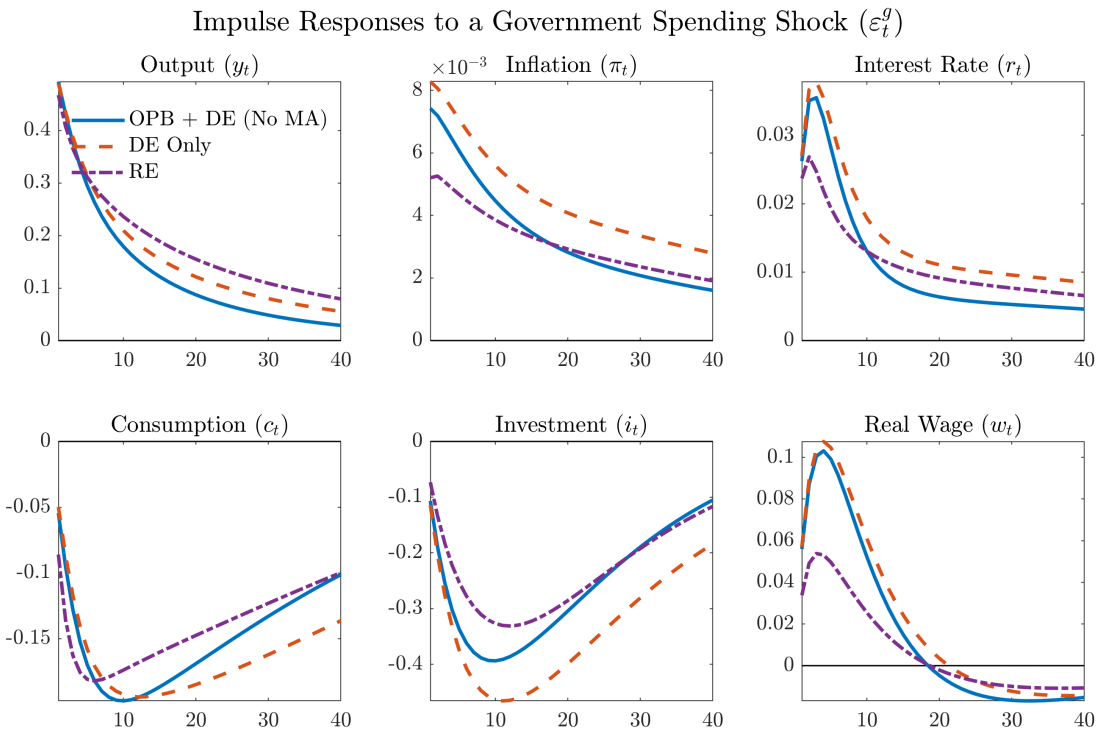


Figure 2: Impulse responses to a government spending shock (ε_t^g).

Notes: See notes to Figure 1.

wages are large and slowly decaying: output peaks around 0.55 and is still above 0.2 at horizon 40; consumption peaks around 0.33 and remains above 0.15 throughout; real wages peak around 0.30 and decay slowly. Under OPB+DE, all three responses are substantially lower and faster-decaying: consumption at its peak is roughly 30 percent lower than under RE, and real wages are approximately 35 percent lower. By horizon 20, the OPB+DE responses have returned close to baseline while the RE responses remain elevated. DE-only is intermediate throughout. Inflation and interest rate responses are more similar across specifications in sign and shape. These differences reflect the persistence substitution documented in the posterior table: under OPB+DE, TFP persistence is lower and some of the low-frequency output variation is attributed to the belief channel, resulting in a less amplified and faster-reverting impulse response. For a policymaker assessing the medium-term consequences of a productivity improvement, the two specifications lead to materially different conclusions about the durability of consumption and real wage gains.

Government spending shock (Figure 2). The government spending shock shows a different pattern. The impact responses are similar across specifications for most variables, but the decay profiles diverge. Under RE, output and the interest rate response persist significantly longer than under OPB+DE. The most striking difference is in inflation: DE-only generates the largest and most persistent inflationary response, above both RE and OPB+DE, reflecting the DE mechanism amplifying agents' overreaction to the government spending innovation. OPB+DE produces the fastest-decaying output response, consistent with lower actual persistence of the fiscal shock under joint estimation. Consumption is crowded out more deeply and for longer under OPB+DE than under RE, while investment crowding out is broadly similar. These differences matter for assessments of fiscal multipliers: the OPB+DE model implies a smaller and shorter-lived output effect of fiscal expansions, with faster crowding out, relative to the RE benchmark.

6.3 The role of survey data

Figure 3 illustrates the identifying role of survey data by comparing three specifications for the TFP shock: OPB+DE estimated with SPF data (our baseline), OPB+DE estimated without SPF data, and the original Smets and Wouters (2007) specification with MA terms and no surveys.

The finding is stark. Without survey data, OPB+DE is nearly indistinguishable from SW

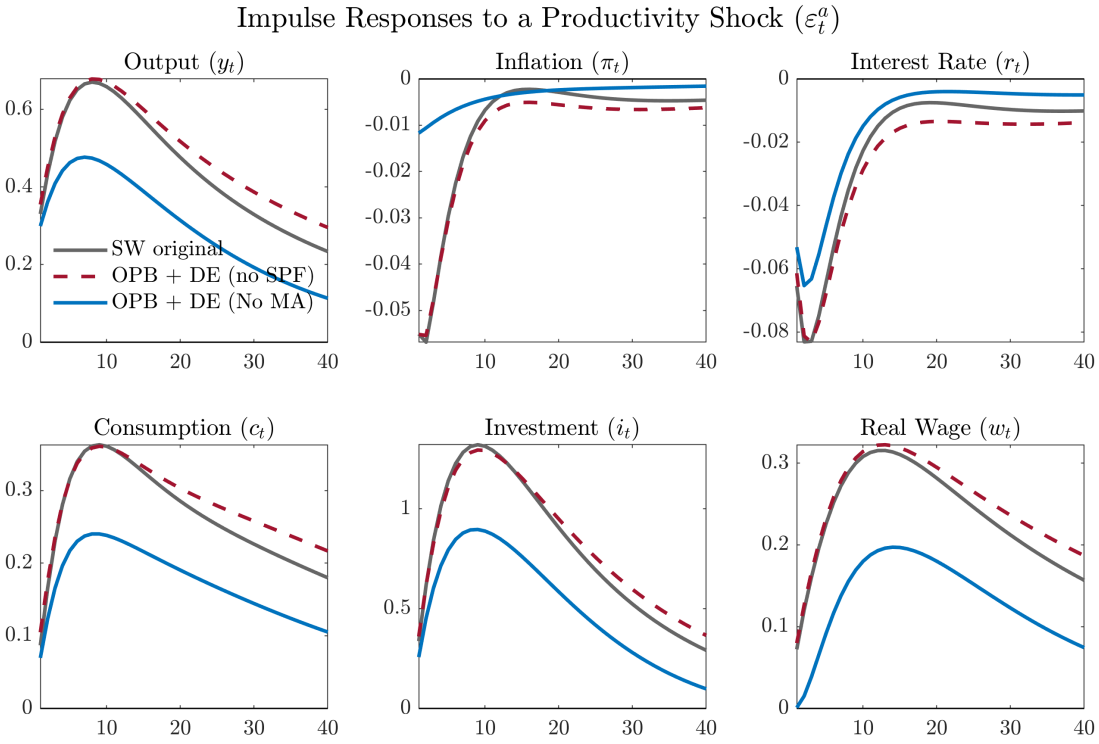


Figure 3: The role of survey data: impulse responses to a productivity shock (ε_t^a).

Notes: Posterior mean responses under three specifications. OPB+DE with SPF (blue solid): joint estimation with SPF data at one-quarter and four-quarter horizons, no moving-average terms. OPB+DE without SPF (dark red solid): same distortions estimated on macro observables only, moving-average terms included as in SW baseline. SW original (gray solid): Smets and Wouters (2007) benchmark with moving-average terms and no survey data. Horizon in quarters.

original across all panels: output, consumption, investment, and real wages track each other closely throughout the horizon. The belief channels are not separately disciplined from the structural parameters when only macro observables are available — without the direct information that survey forecasts provide about subjective expectations, the estimation cannot identify what share of the observed dynamics reflects distorted beliefs versus structural persistence. We obtain very similar results for the remaining shocks (not shown). The OPB and DE parameters are estimated close to zero in the no-survey specification not because the data reject distortions, but because the macro observables are compatible with both the RE model and the distortion model at similar parameter values.

When SPF data are added, the picture changes substantially. The survey forecasts load directly on model-implied subjective expectations and provide direct empirical content for the belief parameters. The result is a substantially lower and faster-decaying response across all panels under OPB+DE with surveys, clearly separated from both SW original and OPB+DE without surveys.

This finding relates to the identifying variation conditions underlying Theorem 1. The theorem guarantees that true persistence and belief distortions are separately identified in population given the full spectrum of observables. In practice, the macro observables alone do not provide sufficient within-sample variation to discipline the belief channel separately from structural persistence — they are consistent with a wide range of (ρ, θ) combinations. Survey data expand the observable set in a way that directly targets the belief-specific channel, making the separation empirically operational. This is not a contradiction of the identification theory — it is a reminder that population identification results require sufficient empirical counterparts in the observable set, and that survey expectations are precisely the type of data that provide those counterparts for belief distortion parameters.

6.4 Connection to the identification results

The empirical findings connect to the theoretical results of Section 3 at three levels.

Theorem 1 — positive identification. The theorem establishes that, in the correctly specified model, the spectral derivatives with respect to ρ_k and θ point in different directions — ρ_k generates a second-order pole signature that no belief distortion can replicate. The empirical counterpart of this result is that the joint posterior over (ρ_k, θ) should not collapse onto a ridge along which the two parameters trade off freely. Consistent with this, the posterior distributions

for the shock persistence parameters and the belief distortion parameters are not strongly negatively correlated in the OPB+DE specification: the data assign positive probability to combinations of nonzero θ^{OPB} and lower ρ_k simultaneously, rather than treating them as interchangeable. The most transparent illustration is the risk premium shock: the posterior for ρ_b tightens substantially relative to RE — moving from a dispersed posterior centered at 0.572 to a tighter posterior centered at 0.252 — while θ^{OPB} and θ^{DE} simultaneously take on positive values. If the two were not separately identified, we would expect the posterior for ρ_b to remain diffuse as the belief parameters absorb variation — instead, both sets of parameters are disciplined jointly, which is what the theorem predicts.

Proposition 1 — omitted-channel bias in persistence. The risk premium shock provides, again, the most direct empirical illustration. As documented in Section 6.1, ρ_b falls from 0.572 under RE to 0.252 under OPB+DE — a drop of more than 0.3 in posterior mean persistence that is directly interpretable as the omitted-channel bias the proposition characterizes: the restricted estimator inflates the included persistence parameter to absorb the dynamics of the omitted belief channel.

Proposition 2 — within-belief confounding. The mild shift in θ^{DE} from 0.044 in DE-only to 0.037 in OPB+DE illustrates the within-belief confounding of Proposition 2: when OPB is added to the estimation, DE absorbs slightly less of the composite perceived-persistence dynamics. The shift is modest because both distortion parameters are small and the spectral derivatives of OPB and DE are partially collinear for highly persistent shocks — as the analytical results in Section 4 establish — meaning the data cannot cleanly separate their contributions even with survey data. A stronger test of the confounding result would require more individually significant distortion parameters; we regard the current exercise as an illustration of the direction of the bias rather than a precise quantification of its magnitude.

The MA identification failure. An additional finding bears directly on the identification theory. When MA terms are reinstated in the price markup shock process, the Komunjer and Ng (2011) rank diagnostics flag a minimality failure: the AR root $\hat{\rho}_p = 0.980$ and MA root $\hat{\mu}_p = 0.975$ converge to nearly identical values, with a collinearity measure of 0.999. The resulting impulse responses (not shown) are economically implausible — inflation following a wage markup shock fails to decay, consumption turns positive, and investment barely responds. This is pole-zero cancellation operating exactly as the identification theory predicts: when an AR and MA root nearly cancel, the persistent structural shock is effectively removed from the

model. This finding provides direct empirical confirmation of the conditions the pole-order argument warns against, and it validates the MA exclusion as both practically necessary and theoretically motivated.

7 Conclusion

This paper studies a practical and recurring identification problem in applied DSGE work: when time-series evidence points to highly persistent macroeconomic dynamics, how much of that persistence reflects *fundamentals* (the objective propagation of exogenous drivers) and how much reflects *perception* (systematic distortions in expectations that alter equilibrium responses)? Motivated by recent general-equilibrium implementations of distorted expectations, we develop a unifying and tractable framework in which agents are structurally rational but form coherent, misspecified beliefs about the law of motion of exogenous fundamentals. A key payoff is an expectations-wedge representation: any distorted-expectations equilibrium in our class can be rewritten as a rational-expectations system with an additional state-dependent wedge. As a result, the framework is a genuine plug-in: it nests multiple belief distortions in a single state-space system and preserves the standard solution and likelihood-based estimation toolkit used in policy-grade DSGE models. An important limitation is that our framework restricts distortions to beliefs about *exogenous* drivers rather than endogenous equilibrium objects. This is a natural starting point for a nested, multi-distortion framework, but it means our implementation of diagnostic expectations differs from the richer self-referential formulations of Bianchi, Ilut, and Saijo (2024) and L'Huillier, Singh, and Yoo (2024), who allow beliefs about endogenous variables and develop solution methods tailored to that setting. Extending our identification framework to that more general class remains an important open question.

Our theoretical analysis uses the spectral density — equivalently, the full set of second moments across horizons — as the population object for identification. Under mild, economically interpretable conditions — that the exogenous drivers are visible in observables and that endogenous roots do not coincide exactly with exogenous persistence — we show that true shock persistence is locally separable from belief distortions when the econometrician estimates the unrestricted model that allows both. At the same time, we formalize why restricted empirical practice can still produce misleading persistence estimates and misleading inference about belief mechanisms: if some belief channels are omitted, remaining parameters generically shift to approximate the missing dynamics, generating omitted-channel bias in persistence and in

included distortion parameters. These results provide an explicit lens for interpreting why persistence estimates move when belief distortion channels are added or removed, without concluding that persistence is unidentified in principle.

We illustrate the approach in the Smets and Wouters (2007) model, jointly estimating over-persistence bias and diagnostic expectations using Survey of Professional Forecasters data. The empirical findings confirm that the theoretical concerns are quantitatively relevant. Estimated distortions at the representative agent level are small, but their effects compound over the policy horizon: an over-persistence bias of 0.017 applied to a shock with persistence 0.96 raises agents' perceived four-quarter cumulative response by approximately eight percentage points. The most striking finding is the risk premium shock, whose estimated persistence falls from 0.572 under rational expectations to 0.252 under joint OPB+DE estimation — a direct empirical illustration of the omitted-channel bias result, as the rational-expectations model inflates structural shock persistence to absorb dynamics that the belief channel generates under the unrestricted specification. For the TFP shock, output, consumption, and real wages are substantially lower and faster-decaying under OPB+DE, with consumption roughly 30 percent lower at the peak. Survey data are essential for these findings: without them, the joint OPB+DE specification collapses toward the Smets and Wouters (2007) benchmark and the belief channels are not separately disciplined. In a next iteration we plan to extend the empirical specification to additional distortions that fit naturally into our framework, in particular cognitive discounting, narrative confusion, and narrative sentiment shocks, and to operationalize a richer set of identification diagnostics suggested by the theory.

A central reason to care about separating propagation from perception is policy analysis. Belief wedges and structural propagation mechanisms can fit similar second moments in-sample yet imply different counterfactuals under regime changes, alternative policy rules, or alternative shock environments. In our empirical illustration, a policymaker using rational-expectations estimates would assess the risk premium shock as highly persistent and the gains from a TFP improvement as durable; a policymaker using OPB+DE estimates would expect faster mean reversion in both cases and calibrate policy accordingly. An important next step is to incorporate counterfactual policy experiments and welfare-relevant objects, using the same plug-in structure to quantify how belief distortions reshape policy transmission and policy tradeoffs.

Finally, the framework points to several promising research directions. One is to develop a disciplined nested class in which expectations about *endogenous* objects are also dis-

torted, while retaining coherence and tractability — a key challenge being the self-referential fixed-point issues that arise when expectation operators act directly on endogenous variables (Bianchi, Ilut, and Saijo 2024; L'Huillier, Singh, and Yoo 2024). A second direction is a more systematic investigation of how medium-scale baselines such as the Smets and Wouters (2007) model may absorb belief-driven propagation through built-in persistence mechanisms, and how this depends on model architecture and the set of observables and belief data used in estimation. A third, more ambitious extension is to combine the present approach with heterogeneous-agent environments, where belief distortions may interact with cross-sectional state distributions and where additional data — micro expectations and cross-sectional heterogeneity — may help discipline competing belief channels.

References

Afrouzi, Hassan, Spencer Y Kwon, Augustin Landier, Yueran Ma, and David Thesmar (Mar. 2023). "Overreaction in Expectations: Evidence and Theory". *The Quarterly Journal of Economics* 138.3, pp. 1713–1764. ISSN: 0033-5533.

Andre, Peter, Ingar Haaland, Christopher Roth, Mirko Wiederholt, and Johannes Wohlfart (forthcoming). "Narratives about the Macroeconomy". *Review of Economic Studies*.

Bianchi, Francesco, Cosmin Ilut, and Hikaru Saijo (2024). "Diagnostic Business Cycles". *The Review of Economic Studies* 91.1, pp. 129–162.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer (2020). "Overreaction in Macroeconomic Expectations". *American Economic Review* 110.9, pp. 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2018). "Diagnostic Expectations and Credit Cycles". *The Journal of Finance* 73.1, pp. 199–227.

Chernozhukov, Victor and Han Hong (2003). "An MCMC Approach to Classical Estimation". *Journal of Econometrics* 115.2, pp. 293–346.

Cogley, Timothy and James M. Nason (June 1995). "Output Dynamics in Real-Business-Cycle Models". *American Economic Review* 85.3, pp. 492–511.

Flynn, Joel P. and Karthik Sastry (2024). "The Macroeconomics of Narratives". NBER Working Paper 32602.

Fuster, Andreas, David Laibson, and Brock Mendel (2010). "Natural Expectations and Macroeconomic Fluctuations". *Journal of Economic Perspectives* 24.4, pp. 67–84.

Gabaix, Xavier (2016). "Behavioral Macroeconomics via Sparse Dynamic Programming".

Gabaix, Xavier (2020). "A Behavioral New Keynesian Model". *American Economic Review* 110.8, pp. 2271–2327.

Gallant, A. Ronald and Halbert White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

Hansen, Lars Peter and Thomas J. Sargent (1983). "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities". *Econometrica* 51.2, pp. 377–387.

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. Random House.

Komunjer, Ivana and Serena Ng (2011). "Dynamic Identification of Dynamic Stochastic General Equilibrium Models". *Econometrica* 79.6, pp. 1995–2032.

L'Huillier, Jean-Paul, Sanjay R Singh, and Donghoon Yoo (2024). "Incorporating Diagnostic Expectations into the New Keynesian Framework". *The Review of Economic Studies* 91.5, pp. 3013–3046.

Milani, Fabio (Oct. 2007). "Expectations, Learning and Macroeconomic Persistence". *Journal of Monetary Economics* 54.7, pp. 2065–2082.

Qu, Zhongjun and Denis Tkachenko (2012). "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models". *Quantitative Economics* 3.1, pp. 95–132.

— (2017). "Global Identification in DSGE Models Allowing for Indeterminacy". *Review of Economic Studies* 84.3, pp. 1306–1345.

Rozsypal, Filip and Kathrin Schlafmann (2023). "Overpersistence bias in individual income expectations and its aggregate implications". *American Economic Journal: Macroeconomics* 15.4, pp. 331–371.

Shiller, Robert J. (2019). *Narrative Economics. How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Slobodyan, Sergey and Raf Wouters (2012). "Learning in an Estimated Medium-Scale DSGE Model". *Journal of Economic Dynamics and Control* 36.1, pp. 26–46.

Smets, Frank and Rafael Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach". *American Economic Review* 97.3, 586–606.

White, Halbert (1982). "Maximum likelihood estimation of misspecified models". *Econometrica*, pp. 1–25.

A Proofs

Proof of Lemma 1

Substitute the true law $\zeta_t = \Psi^A(\rho)\zeta_{t-1} + \varepsilon_t$ into the policy rule (24):

$$x_t^{en} = P_x^{en} x_{t-1}^{en} + P_\zeta^{en}(\Psi^A \zeta_{t-1} + \varepsilon_t) = P_x^{en} x_{t-1}^{en} + (P_\zeta^{en}\Psi^A)\zeta_{t-1} + P_\zeta^{en}\varepsilon_t.$$

Stack this with $\zeta_t = \Psi^A \zeta_{t-1} + \varepsilon_t$ to obtain

$$\begin{pmatrix} x_t^{en} \\ \zeta_t \end{pmatrix} = \begin{pmatrix} P_x^{en} & P_\zeta^{en}\Psi^A \\ 0 & \Psi^A \end{pmatrix} \begin{pmatrix} x_{t-1}^{en} \\ \zeta_{t-1} \end{pmatrix} + \begin{pmatrix} P_\zeta^{en} \\ I \end{pmatrix} \varepsilon_t,$$

which is (21) with the block structure in (25).

Proof of Lemma 2

By Lemma 1, F is block upper triangular:

$$F = \begin{pmatrix} F_{xx} & F_{x\zeta} \\ 0 & \Psi^A \end{pmatrix}.$$

Hence

$$I - Fz^{-1} = \begin{pmatrix} I - F_{xx}z^{-1} & -F_{x\zeta}z^{-1} \\ 0 & I - \Psi^A z^{-1} \end{pmatrix}.$$

Whenever $I - F_{xx}z^{-1}$ and $I - \Psi^A z^{-1}$ are invertible, the inverse of a block upper triangular matrix satisfies

$$\begin{pmatrix} E & F \\ 0 & G \end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}FG^{-1} \\ 0 & G^{-1} \end{pmatrix}.$$

Applying this identity with $E = I - F_{xx}z^{-1}$, $F = -F_{x\zeta}z^{-1}$, and $G = I - \Psi^A z^{-1}$ yields (27), proving 1.

For 2., since $\Psi^A(\rho)$ is diagonal, each diagonal entry of $(I - \Psi^A z^{-1})^{-1}$ equals $(1 - \rho_k z^{-1})^{-1}$, which has a simple pole at $z = \rho_k$. From (27), $(I - \Psi^A z^{-1})^{-1}$ appears only in the lower-right block and as the right-most factor in the upper-right block, never multiplied by itself; therefore it appears at most to the first power in each entry.

For 3., any factor $(1-\rho_k z^{-1})^{-1}$ entering the resolvent must come from $(I-\Psi^A z^{-1})^{-1}$ in (27). Since this factor appears at most once, a squared term $(1-\rho_k z^{-1})^{-2}$ cannot arise unless an additional simple pole at the same location is contributed by $(I-F_{xx}z^{-1})^{-1}$, i.e. unless ρ_k is also an eigenvalue of $F_{xx}(\vartheta,\theta)$.

Proof of Theorem 1

Fix a component k. We argue by contradiction using pole order.

Step 1: From L^2 equality on the unit circle to a rational identity. Each entry of $\mathscr{S}(z;\vartheta,\theta)$ is a rational function of z, because it is built from matrix inverses of $(I-Fz^{-1})$ and $(I-F'z)$. Stability implies $\mathscr{S}(e^{i\omega};\vartheta_0,0)$ is continuous in ω on $[-\pi,\pi]$ (there are no poles on $|z|=1$). If two continuous functions coincide in $L^2([-\pi,\pi])$, they coincide almost everywhere, hence (by continuity) they coincide everywhere on $[-\pi,\pi]$. Therefore, an L^2 identity for $\omega\mapsto\mathscr{S}(e^{i\omega})$ implies equality of the corresponding rational functions on an uncountable set of z values (the entire unit circle), and thus the rational functions must be identical. In particular, they must have the same poles with the same pole orders.

Step 2: A visible simple pole at $z=\rho_k$ at the evaluation point. By Lemma 2 and pole separation (Assumption 3), $\mathscr{S}(z;\vartheta_0,0)$ has at most a simple pole at $z=\rho_k$ (no other factor generates a singularity at the same location). By visibility (Assumption 2), this pole is non-removable: there exist indices (i,j) and a neighborhood of $z=\rho_k$ such that

$$[\mathscr{S}(z;\vartheta_0,0)]_{ij}=\frac{R_{ij}(\vartheta_0)}{z-\rho_k}+(\text{analytic at } z=\rho_k),\qquad R_{ij}(\vartheta_0)\neq 0, \tag{42}$$

where $R_{ij}(\vartheta_0)$ is the (nonzero) residue of the scalar rational function $z\mapsto[\mathscr{S}(z;\vartheta_0,0)]_{ij}$ at $z=\rho_k$.

Step 3: Differentiating w.r.t. ρ_k creates a second-order pole. Note that the residue $R_{ij}(\vartheta)$ may depend on ρ_k through the DSGE solution. Differentiating (42) w.r.t. ρ_k yields

$$\left[\frac{\partial}{\partial\rho_k}\mathscr{S}(z;\vartheta_0,0)\right]_{ij}=\frac{\partial R_{ij}}{\partial\rho_k}(\vartheta_0)\,\frac{1}{z-\rho_k}+\frac{R_{ij}(\vartheta_0)}{(z-\rho_k)^2}+(\text{analytic at } z=\rho_k). \tag{43}$$

Because $R_{ij}(\vartheta_0)\neq 0$, the derivative $\partial\mathscr{S}/\partial\rho_k$ has a *genuine second-order* pole at $z=\rho_k$ in at least one entry. Differentiating the analytic remainder may produce additional simple poles or analytic terms at $z=\rho_k$, but these do not affect the conclusion: the coefficient on $(z-\rho_k)^{-2}$

is $R_{ij}(\vartheta_0) \neq 0$ regardless, so the second-order pole is genuine and survives.

Step 4: Differentiating w.r.t. θ cannot create a second-order pole at $z = \rho_k$. Under Lemma 1, $\Psi^A(\rho)$ does not depend on θ and appears in $(I - Fz^{-1})^{-1}$ only through the factor $(I - \Psi^A z^{-1})^{-1}$ as in (27). A generic θ_j-derivative of the matrix inverse satisfies

$$\frac{\partial}{\partial \theta_j}(I - Fz^{-1})^{-1} = (I - Fz^{-1})^{-1} \left(\frac{\partial F}{\partial \theta_j}\right) z^{-1} (I - Fz^{-1})^{-1}.$$

One might worry that the two inverted factors could generate a squared exogenous term. However, by Lemma 1,

$$\frac{\partial F}{\partial \theta_j} = \begin{pmatrix} \frac{\partial F_{xx}}{\partial \theta_j} & \frac{\partial F_{x\zeta}}{\partial \theta_j} \\ 0 & 0 \end{pmatrix},$$

so the lower-right block is zero. A direct block multiplication using (27) shows that $\frac{\partial}{\partial \theta_j}(I - Fz^{-1})^{-1}$ can contain the exogenous factor $(I - \Psi^A z^{-1})^{-1}$ in the upper-right block *at most to the first power*, never squared. Moreover, by pole separation, all terms involving $(I - F_{xx} z^{-1})^{-1}$ are analytic at $z = \rho_k$. The same reasoning applies to the right-side inverse $(I - F'z)^{-1}$.

Therefore, for every j, each entry of $\frac{\partial}{\partial \theta_j}\mathscr{S}(z; \vartheta_0, 0)$ has *at most a simple pole* at $z = \rho_k$.

Step 5: Contradiction with the span hypothesis. Suppose, contrary to (28), that $\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0)$ lies in $\text{Range}(D_\theta \mathcal{S}(\vartheta_0, 0))$. Then there exist coefficients β_j such that, in $L^2([-\pi, \pi])$,

$$\frac{\partial}{\partial \rho_k}\text{vec}(S_y(\omega; \vartheta_0, 0)) = \sum_j \beta_j \frac{\partial}{\partial \theta_j}\text{vec}(S_y(\omega; \vartheta_0, 0)).$$

By Step 1, the corresponding identity holds as an identity between rational functions for the analytic extensions:

$$\frac{\partial}{\partial \rho_k}\text{vec}(\mathscr{S}(z; \vartheta_0, 0)) = \sum_j \beta_j \frac{\partial}{\partial \theta_j}\text{vec}(\mathscr{S}(z; \vartheta_0, 0)).$$

But Step 3 shows the left-hand side has a second-order pole at $z = \rho_k$ in at least one entry, whereas Step 4 shows the right-hand side has at most a simple pole at $z = \rho_k$ in every entry. A rational identity cannot hold if pole orders differ. Hence there is a contradiction.

Therefore $\frac{\partial}{\partial \rho_k}\mathcal{S}(\vartheta_0, 0) \notin \text{Range}(D_\theta \mathcal{S}(\vartheta_0, 0))$, proving (28).

Proof of Proposition 1

Equip $L^2([-\pi,\pi];\mathbb{C}^{n_y^2})$ with the inner product

$$\langle f, g\rangle := \int_{-\pi}^{\pi} f(\omega)^* g(\omega)\, d\omega, \qquad \|f\|_{L^2}^2 = \langle f, f\rangle,$$

and let J^* denote the adjoint operator induced by this inner product.

Because $\mathcal{S}$ is Fréchet differentiable at $(\vartheta_0, 0)$, we have the first-order expansions

$$\mathcal{S}(\rho_0, \psi_0, \theta_0) = \mathcal{S}(\rho_0, \psi_0, 0) + J_\theta \theta_0 + o(\|\theta_0\|),$$

and, for $(\rho, \psi) = (\rho_0, \psi_0) + \delta$,

$$\mathcal{S}(\rho, \psi, 0) = \mathcal{S}(\rho_0, \psi_0, 0) + J_{(\rho,\psi)}\delta + o(\|\delta\|).$$

Subtracting yields

$$\mathcal{S}(\rho_0, \psi_0, \theta_0) - \mathcal{S}(\rho, \psi, 0) = J_\theta \theta_0 - J_{(\rho,\psi)}\delta + o(\|\theta_0\| + \|\delta\|).$$

Therefore, up to second-order terms, the restricted problem (29) is approximated by the least-squares projection problem

$$\min_{\delta} \ \left\| J_\theta \theta_0 - J_{(\rho,\psi)}\delta \right\|_{L^2}^2 .$$

The normal equations for this Hilbert-space least-squares problem are

$$J_{(\rho,\psi)}^* J_{(\rho,\psi)}\, \delta = J_{(\rho,\psi)}^* J_\theta\, \theta_0.$$

Under the full-column-rank condition, $J_{(\rho,\psi)}^* J_{(\rho,\psi)}$ is invertible, and the unique solution is

$$\delta = \left(J_{(\rho,\psi)}^* J_{(\rho,\psi)} \right)^{-1} J_{(\rho,\psi)}^* J_\theta\, \theta_0.$$

Substituting $(\rho^*, \psi^*) = (\rho_0, \psi_0) + \delta$ gives (30). If $J_\theta \theta_0$ has a nonzero projection onto $\mathrm{Range}(J_{(\rho,\psi)})$, then $\delta \neq 0$ and the restricted parameters shift at first order, implying $\rho^* \neq \rho_0$ generically.

Proof of Proposition 2

Equip $L^2([-\pi,\pi];\mathbb{C}^{n_y^2})$ with its standard inner product and let * denote the adjoint operator. By Fréchet differentiability at $(\vartheta_0,\theta_{1,0},0)$, we have the expansions

$$\mathcal{S}(\vartheta_0,\theta_{1,0},\theta_{2,0}) = \mathcal{S}(\vartheta_0,\theta_{1,0},0) + J_2\theta_{2,0} + o(\|\theta_{2,0}\|),$$

and for $(\rho,\psi,\theta_1) = (\rho_0,\psi_0,\theta_{1,0}) + \delta$,

$$\mathcal{S}(\rho,\psi,\theta_1,0) = \mathcal{S}(\vartheta_0,\theta_{1,0},0) + J_1\delta + o(\|\delta\|).$$

Thus the objective in (31) is locally equivalent to the Hilbert-space least-squares problem

$$\min_{\delta} \ \|J_2\theta_{2,0} - J_1\delta\|_{L^2}^2,$$

whose normal equations are $J_1^*J_1\delta = J_1^*J_2\theta_{2,0}$. Invertibility of $J_1^*J_1$ yields

$$\delta = (J_1^*J_1)^{-1}J_1^*J_2\theta_{2,0},$$

which implies (32). The bias condition for θ_1 follows from reading off the θ_1-block of δ. The final statement follows because Theorem 1 relies on pole-order differences between ρ-derivatives and belief-derivatives, whereas θ_1- and θ_2-derivatives share the same pole-order class.

Proof of Proposition 3

(1) Since $\Psi^S(\rho_0,\theta^{\mathrm{OPB}}) = \Psi^A(\rho_0) + \sum_{m=1}^{n_\rho}\theta_m^{\mathrm{OPB}}E_{mm}$, differentiation yields $\partial\Psi^S/\partial\theta_k^{\mathrm{OPB}} = E_{kk}$ and thus $\Delta_k^{\mathrm{OPB}} = E_{kk}$.

(2) By definition,

$$\Omega(\vartheta,\theta^{\mathrm{OPB}}) = A_{+1}P_\zeta^S(\vartheta,\theta^{\mathrm{OPB}})\big(\Psi^S(\rho,\theta^{\mathrm{OPB}}) - \Psi^A(\rho)\big).$$

Applying the product rule at the evaluation point $(\vartheta_0,0)$ gives

$$\frac{\partial\Omega(\vartheta_0,\theta^{\mathrm{OPB}})}{\partial\theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0} = A_{+1}\,\frac{\partial P_\zeta^S(\vartheta_0,\theta^{\mathrm{OPB}})}{\partial\theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0}\big(\Psi^S(\rho_0,0)-\Psi^A(\rho_0)\big)+A_{+1}P_\zeta^S(\vartheta_0,0)\,\frac{\partial\big(\Psi^S(\rho_0,\theta^{\mathrm{OPB}})-\Psi^A(\rho_0)\big)}{\partial\theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0}$$

Since $\Psi^S(\rho_0, 0) - \Psi^A(\rho_0) = 0$, the first term vanishes. Moreover,

$$\frac{\partial\big(\Psi^S(\rho_0, \theta^{\mathrm{OPB}}) - \Psi^A(\rho_0)\big)}{\partial \theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0} = \frac{\partial \Psi^S(\rho_0, \theta^{\mathrm{OPB}})}{\partial \theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0} = E_{kk}.$$

Therefore,

$$\frac{\partial \Omega(\vartheta_0, \theta^{\mathrm{OPB}})}{\partial \theta_k^{\mathrm{OPB}}}\bigg|_{\theta^{\mathrm{OPB}}=0} = A_{+1} P_\zeta^S(\vartheta_0, 0)\, E_{kk} = A_{+1} P_\zeta^A\, E_{kk}.$$

(3) The second-order pole for $\partial \mathscr{S}(z; \vartheta_0, 0)/\partial \rho_k$ follows from Theorem 1. It remains to show that $\partial \mathscr{S}(z; \vartheta_0, 0)/\partial \theta_k^{\mathrm{OPB}}$ has at most a simple pole at $z = \rho_{0,k}$.

By Lemma 1, the stacked transition matrix has the form

$$F(\vartheta, \theta^{\mathrm{OPB}}) = \begin{pmatrix} F_{xx}(\vartheta, \theta^{\mathrm{OPB}}) & F_{x\zeta}(\vartheta, \theta^{\mathrm{OPB}}) \\ 0 & \Psi^A(\rho) \end{pmatrix},$$

so the exogenous block $\Psi^A(\rho)$ is independent of θ^{OPB}. Hence

$$\frac{\partial F(\vartheta_0, \theta^{\mathrm{OPB}})}{\partial \theta_k^{\mathrm{OPB}}} = \begin{pmatrix} \frac{\partial F_{xx}}{\partial \theta_k^{\mathrm{OPB}}} & \frac{\partial F_{x\zeta}}{\partial \theta_k^{\mathrm{OPB}}} \\ 0 & 0 \end{pmatrix},$$

i.e. the lower-right block is zero. Using the standard identity for derivatives of resolvents,

$$\frac{\partial}{\partial \theta_k^{\mathrm{OPB}}}(I - Fz^{-1})^{-1} = (I - Fz^{-1})^{-1}\Big(\frac{\partial F}{\partial \theta_k^{\mathrm{OPB}}}\Big) z^{-1} (I - Fz^{-1})^{-1},$$

one might worry that the product of two resolvents could generate squared exogenous terms. However, because $\partial F/\partial \theta_k^{\mathrm{OPB}}$ has a zero lower-right block, a direct block multiplication together with the block inverse representation in Lemma 2 implies that the exogenous resolvent $(I - \Psi^A z^{-1})^{-1}$ can appear in each entry of $\frac{\partial}{\partial \theta_k^{\mathrm{OPB}}}(I - Fz^{-1})^{-1}$ at most to the first power, never squared. The same reasoning applies to the right resolvent $(I - F'z)^{-1}$.

Furthermore, the product rule applied to the complex extension

$$\mathscr{S}(z; \vartheta, \theta^{\mathrm{OPB}}) = H(I - F(\vartheta, \theta^{\mathrm{OPB}})z^{-1})^{-1} G(\vartheta, \theta^{\mathrm{OPB}}) \Sigma_\varepsilon G(\vartheta, \theta^{\mathrm{OPB}})' (I - F(\vartheta, \theta^{\mathrm{OPB}})' z)^{-1} H' + \Sigma_\eta,$$

also differentiates the impact matrices $G(\vartheta, \theta^{\mathrm{OPB}})$. However, because these matrices are static, their derivatives introduce no new dynamic poles. Thus, differentiation with respect to θ_k^{OPB} cannot create $(z - \rho_{0,k})^{-2}$ terms and yields at most a simple pole at $z = \rho_{0,k}$. This proves the

pole-order contrast stated in part (3), and hence the non-confounding conclusion.

Proof of Proposition 4

(1) $\Psi^S(\rho_0, \theta^{\mathrm{DE}})$ differs from $\Psi^A(\rho_0)$ only through the upper-right block $\Theta^{\mathrm{DE}} = \mathrm{diag}(\theta_1^{\mathrm{DE}}, \ldots, \theta_{n_\rho}^{\mathrm{DE}})$. Hence $\partial\Psi^S(\rho_0, \theta^{\mathrm{DE}})/\partial\theta_k^{\mathrm{DE}} = E_{k,n_\rho+k}$ and thus $\Delta_k^{\mathrm{DE}} = E_{k,n_\rho+k}$.

(2) By definition,

$$\Omega(\vartheta, \theta^{\mathrm{DE}}) = A_{+1} P_\zeta^S(\vartheta, \theta^{\mathrm{DE}}) \big(\Psi^S(\rho, \theta^{\mathrm{DE}}) - \Psi^A(\rho)\big).$$

Applying the product rule at the evaluation point $(\vartheta_0, 0)$ gives

$$\left.\frac{\partial\Omega(\vartheta_0, \theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0} = A_{+1} \left.\frac{\partial P_\zeta^S(\vartheta_0, \theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0} \big(\Psi^S(\rho_0, 0) - \Psi^A(\rho_0)\big) + A_{+1} P_\zeta^S(\vartheta_0, 0) \left.\frac{\partial\big(\Psi^S(\rho_0, \theta^{\mathrm{DE}}) - \Psi^A(\rho_0)\big)}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0}.$$

Since $\Psi^S(\rho_0, 0) - \Psi^A(\rho_0) = 0$, the first term vanishes. Moreover,

$$\left.\frac{\partial\big(\Psi^S(\rho_0, \theta^{\mathrm{DE}}) - \Psi^A(\rho_0)\big)}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0} = \left.\frac{\partial\Psi^S(\rho_0, \theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0} = E_{k,n_\rho+k}.$$

Therefore,

$$\left.\frac{\partial\Omega(\vartheta_0, \theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}}\right|_{\theta^{\mathrm{DE}}=0} = A_{+1} P_\zeta^S(\vartheta_0, 0)\, E_{k,n_\rho+k} = A_{+1} P_\zeta^A\, E_{k,n_\rho+k}.$$

(3) The second-order pole for $\partial\mathscr{S}(z; \vartheta_0, 0)/\partial\rho_k$ follows from Theorem 1. It remains to show that $\partial\mathscr{S}(z; \vartheta_0, 0)/\partial\theta_k^{\mathrm{DE}}$ has at most a simple pole at $z = \rho_{0,k}$.

By Lemma 1, the stacked transition matrix has the form

$$F(\vartheta, \theta^{\mathrm{DE}}) = \begin{pmatrix} F_{xx}(\vartheta, \theta^{\mathrm{DE}}) & F_{x\zeta}(\vartheta, \theta^{\mathrm{DE}}) \\ 0 & \Psi^A(\rho) \end{pmatrix},$$

so the exogenous block $\Psi^A(\rho)$ is independent of θ^{DE}. Hence

$$\frac{\partial F(\vartheta_0, \theta^{\mathrm{DE}})}{\partial\theta_k^{\mathrm{DE}}} = \begin{pmatrix} \frac{\partial F_{xx}}{\partial\theta_k^{\mathrm{DE}}} & \frac{\partial F_{x\zeta}}{\partial\theta_k^{\mathrm{DE}}} \\ 0 & 0 \end{pmatrix},$$

i.e. the lower-right block is zero. Using the standard identity for derivatives of resolvents,

$$\frac{\partial}{\partial \theta_k^{\mathrm{DE}}}(I - Fz^{-1})^{-1} = (I - Fz^{-1})^{-1}\Big(\frac{\partial F}{\partial \theta_k^{\mathrm{DE}}}\Big)z^{-1}(I - Fz^{-1})^{-1},$$

one might worry that the product of two resolvents could generate squared exogenous terms. However, because $\partial F/\partial \theta_k^{\mathrm{DE}}$ has a zero lower-right block, a direct block multiplication together with the block inverse representation in Lemma 2 implies that the exogenous resolvent $(I - \Psi^A z^{-1})^{-1}$ can appear in each entry of $\frac{\partial}{\partial \theta_k^{\mathrm{DE}}}(I - Fz^{-1})^{-1}$ at most to the first power, never squared. The same reasoning applies to the right resolvent $(I - F'z)^{-1}$.

Furthermore, the product rule applied to the complex extension

$$\mathscr{S}(z; \vartheta, \theta^{\mathrm{DE}}) = H(I - F(\vartheta, \theta^{\mathrm{DE}})z^{-1})^{-1}G(\vartheta, \theta^{\mathrm{DE}})\Sigma_\varepsilon G(\vartheta, \theta^{\mathrm{DE}})'(I - F(\vartheta, \theta^{\mathrm{DE}})'z)^{-1}H' + \Sigma_\eta,$$

also differentiates the static impact matrices $G(\vartheta, \theta^{\mathrm{DE}})$, which introduces no new dynamic poles. Thus, differentiation with respect to θ_k^{DE} cannot create $(z - \rho_{0,k})^{-2}$ terms and yields at most a simple pole at $z = \rho_{0,k}$. This proves the pole-order contrast stated in part (3), and confirms the non-confounding conclusion.

Proof of Corollary 2

Substitute $\nu_{k,t} = u_{k,t} - \rho_k u_{k,t-1}$ into (36) and collect terms to obtain (37). Part (2) follows because under the restriction $b_{1,k} \equiv 0$ only $b_{0,k}$ matters, and $b_{0,k} - \rho_k = \theta_k^{\mathrm{OPB}} + \theta_k^{\mathrm{DE}}$. Part (3) follows by solving (37) for $(\theta_k^{\mathrm{OPB}}, \theta_k^{\mathrm{DE}})$ when $\rho_k \neq 0$.

B Additional distortion templates and a derivation workflow

This appendix provides worked templates illustrating how to derive distortion-specific implications within our framework. The logic is uniform: specify a coherent perceived transition Ψ^S, compute tangents Δ_i, map them into wedge directions $A_{+1}P_\zeta^A\Delta_i$, and interpret identification using pole-order separation for ρ versus θ (Theorem 1) and derivative-span overlap for θ_1 versus θ_2 (Proposition 2).

Narrative confusion as a confusion matrix (mixing across fundamentals).

Consider two fundamentals (e.g. demand and supply drivers), $\zeta_t = (d_t, s_t)'$, with actual diagonal AR(1) transition

$$\zeta_{t+1} = \Psi^A(\rho)\zeta_t + \varepsilon_{t+1}, \qquad \Psi^A(\rho) = \mathrm{diag}(\rho_d, \rho_s), \qquad |\rho_d| < 1,\ |\rho_s| < 1.$$

A natural narrative/confusion distortion (Andre, Haaland, Roth, Wiederholt, and Wohlfart (forthcoming)) is that agents apply a *confusion matrix* to the current state before propagating it with the true persistence rates:

$$E_t^S[\zeta_{t+1}] = \Psi^A(\rho)\, C(\lambda)\, \zeta_t, \qquad C(\lambda) = \begin{pmatrix} 1-\lambda_{sd} & \lambda_{sd} \\ \lambda_{ds} & 1-\lambda_{ds} \end{pmatrix}, \qquad \lambda_{sd}, \lambda_{ds} \in [0,1].$$

Equivalently,

$$\Psi^S(\rho, \lambda) = \Psi^A(\rho)\, C(\lambda),$$

which cleanly separates *physical persistence* ρ from *belief mixing* λ.

The tangents at $\lambda = 0$ are immediate because $C(0) = I$ and

$$\left.\frac{\partial C}{\partial \lambda_{sd}}\right|_0 = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \left.\frac{\partial C}{\partial \lambda_{ds}}\right|_0 = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}.$$

Therefore

$$\Delta_{\lambda_{sd}} = \left.\frac{\partial \Psi^S}{\partial \lambda_{sd}}\right|_0 = \Psi^A(\rho) \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -\rho_d & \rho_d \\ 0 & 0 \end{pmatrix}, \qquad \Delta_{\lambda_{ds}} = \begin{pmatrix} 0 & 0 \\ \rho_s & -\rho_s \end{pmatrix}.$$

The corresponding first-order wedge directions are $A_{+1}P_\zeta^A\Delta_{\lambda_{sd}}$ and $A_{+1}P_\zeta^A\Delta_{\lambda_{ds}}$. Since $\Psi^A(\rho)$ is unchanged, such narrative confusion cannot move exogenous pole locations and therefore can locally be separately identified from ρ under Theorem 1, but it can be confounded with other belief channels that generate similar off-diagonal wedge directions.

Narrative shocks via spurious predictors.

To capture narratives of the "this time is different" / "price gouging" type in the sense of Shiller (2019), augment the state as $\zeta_t = (u_t, \xi_t)'$, where u_t is a true fundamental and ξ_t is a *spurious signal* observed by agents. The defining feature is that ξ_t is *not even predictive* for

u_t under the actual DGP. A simple representation is block-diagonal actual dynamics:

$$\zeta_{t+1} = \Psi^A(\rho)\zeta_t + \varepsilon_{t+1}, \qquad \Psi^A(\rho) = \begin{pmatrix} \rho_u & 0 \\ 0 & \rho_\xi \end{pmatrix}, \qquad \varepsilon_{t+1} = (\varepsilon^u_{t+1}, \varepsilon^\xi_{t+1})',$$

with $\varepsilon^u_{t+1} \perp \varepsilon^\xi_{t+1}$ and no cross-predictability. A narrative distortion is that agents perceive ξ_t as predictive for u_{t+1}:

$$E^S_t[u_{t+1}] = \rho_u u_t + \theta^\xi \xi_t, \qquad E^S_t[\xi_{t+1}] = \rho_\xi \xi_t,$$

so that Ψ^S differs from Ψ^A only through the off-diagonal block $\Psi^S_{u\xi} = \theta^\xi$. The tangent direction at $\theta^\xi = 0$ is supported exactly on that block, and the induced wedge direction follows from $A_{+1}P^A_\zeta \Delta_{\theta^\xi}$. This makes the empirical message transparent: without either observables that load meaningfully on ξ_t (or a proxy) *or* belief data that discipline forecasts conditional on ξ_t, the narrative channel can be absorbed by other belief channels that shift perceived persistence of u_t.

Cognitive discounting (damping / limited attention).

A canonical reduced-form version of cognitive discounting (limited attention) in the sense of Gabaix (2016, 2020) scales perceived propagation:

$$\Psi^S(\kappa) = \mathrm{diag}(\kappa_1\rho_1, \ldots, \kappa_{n_\rho}\rho_{n_\rho}), \qquad \kappa_k \in [0,1],$$

nested at $\kappa_k = 1$ (full attention). The tangents at $\kappa = \mathbf{1}$ are

$$\Delta_{\kappa_k} = \left.\frac{\partial \Psi^S(\kappa)}{\partial \kappa_k}\right|_{\kappa=\mathbf{1}} = \rho_k E_{kk},$$

and the corresponding wedge directions are $A_{+1}P^A_\zeta \rho_k E_{kk}$. Damping therefore shares a "diagonal" perceived-transition structure with persistence-type distortions such as OPB. As a result, separating damping from other diagonal distortions is not a pole-location problem (poles are still anchored by Ψ^A), but a within-θ confounding problem governed by derivative-span overlap as in Proposition 2.

Takeaway.

Across these examples, the workflow is the same: specify Ψ^S, compute tangents Δ_i, trans-

late them into wedge directions, and interpret identification through pole-order separation for ρ versus θ and through overlap across wedge/Jacobian directions for θ_1 versus θ_2.

C Supplementary Smets–Wouters Results

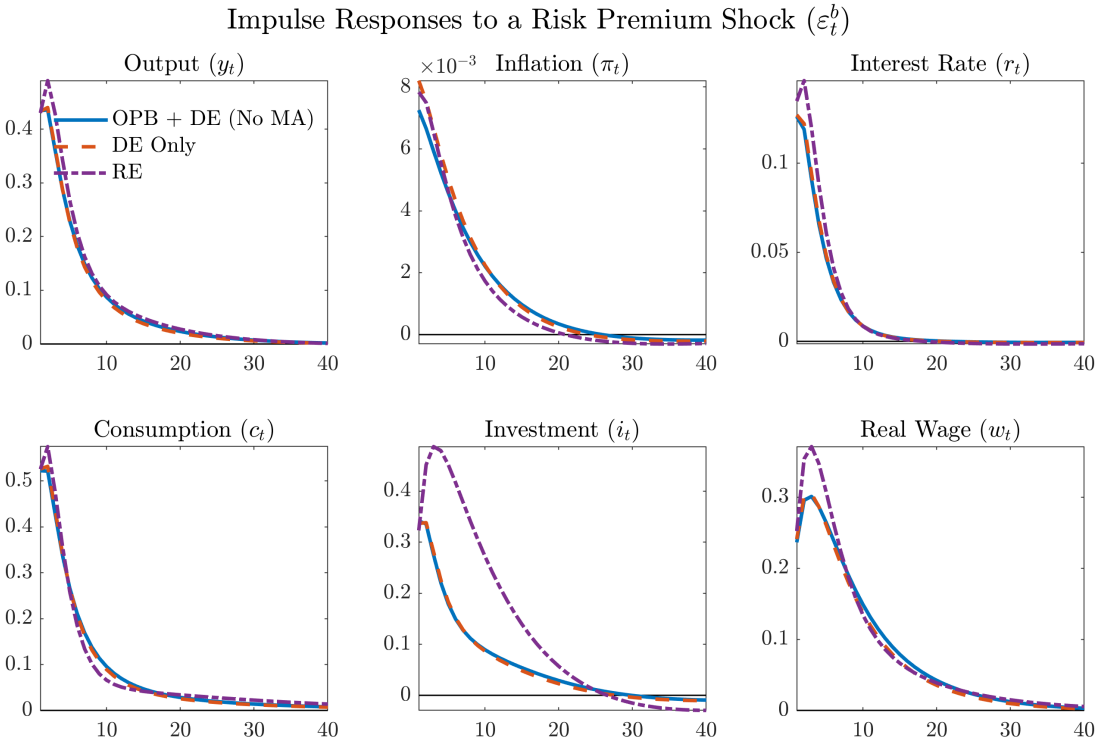


Figure 4: Impulse responses to a risk premium shock (ε_t^b).

Notes: See notes to Figure 1.

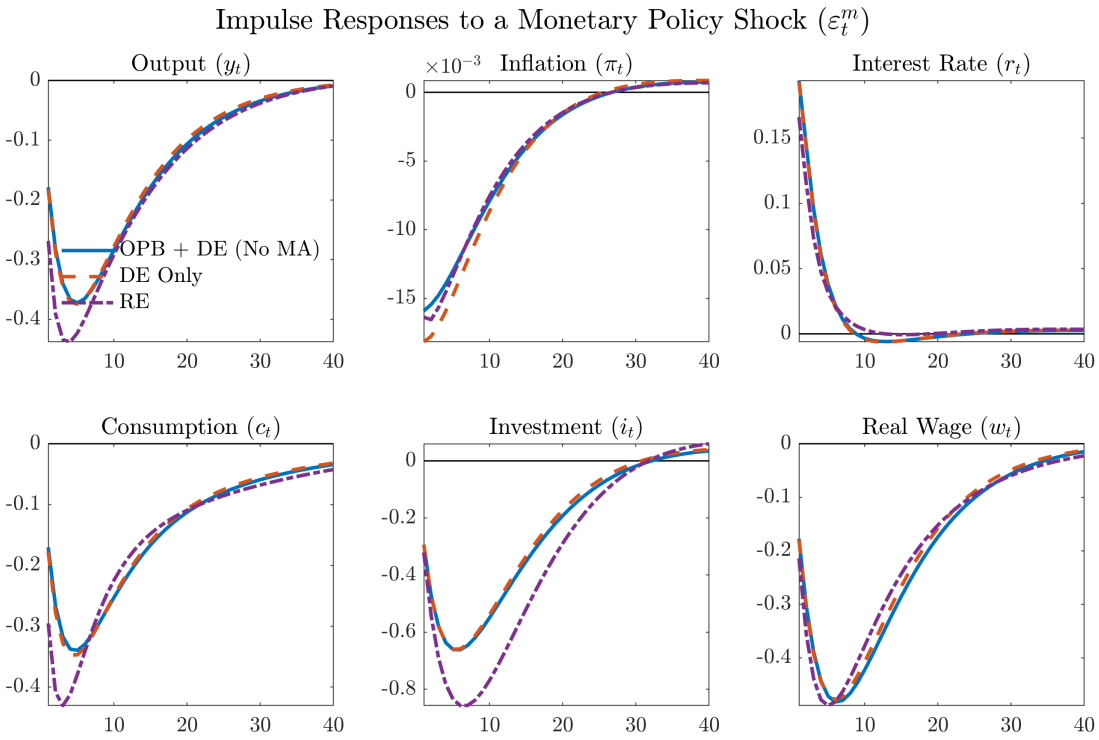


Figure 5: Impulse responses to a monetary policy shock (ε_t^m).

Notes: See notes to Figure 1.

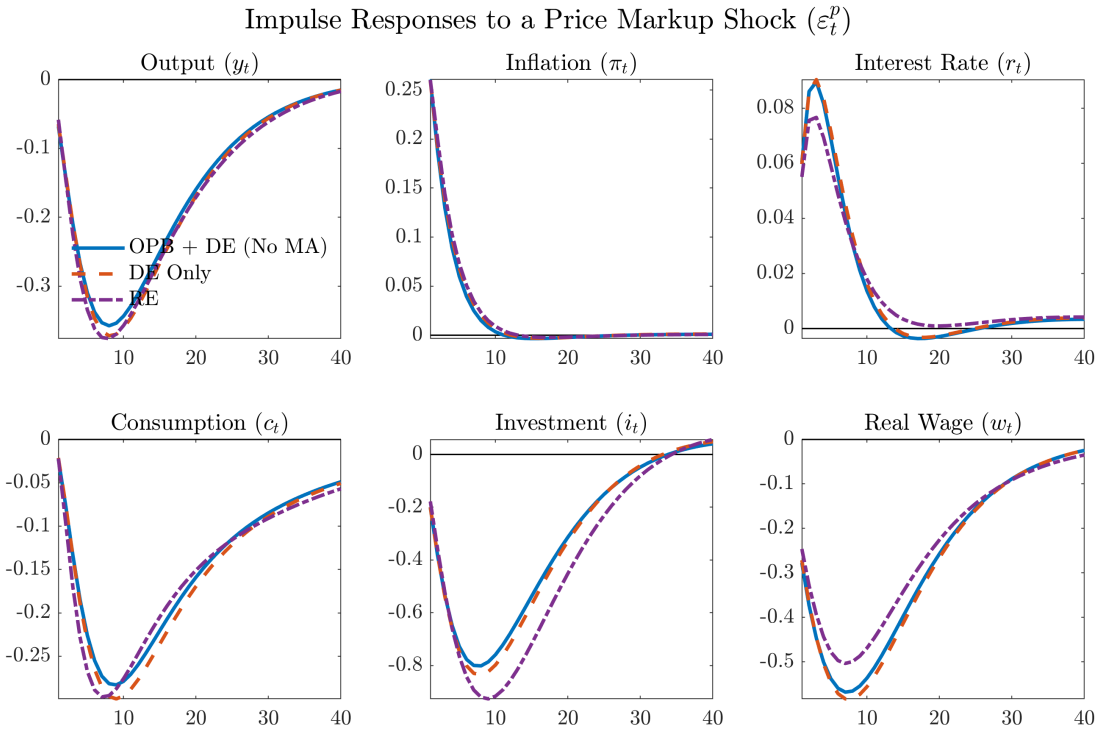


Figure 6: Impulse responses to a price markup shock (ε_t^p).

Notes: See notes to Figure 1.

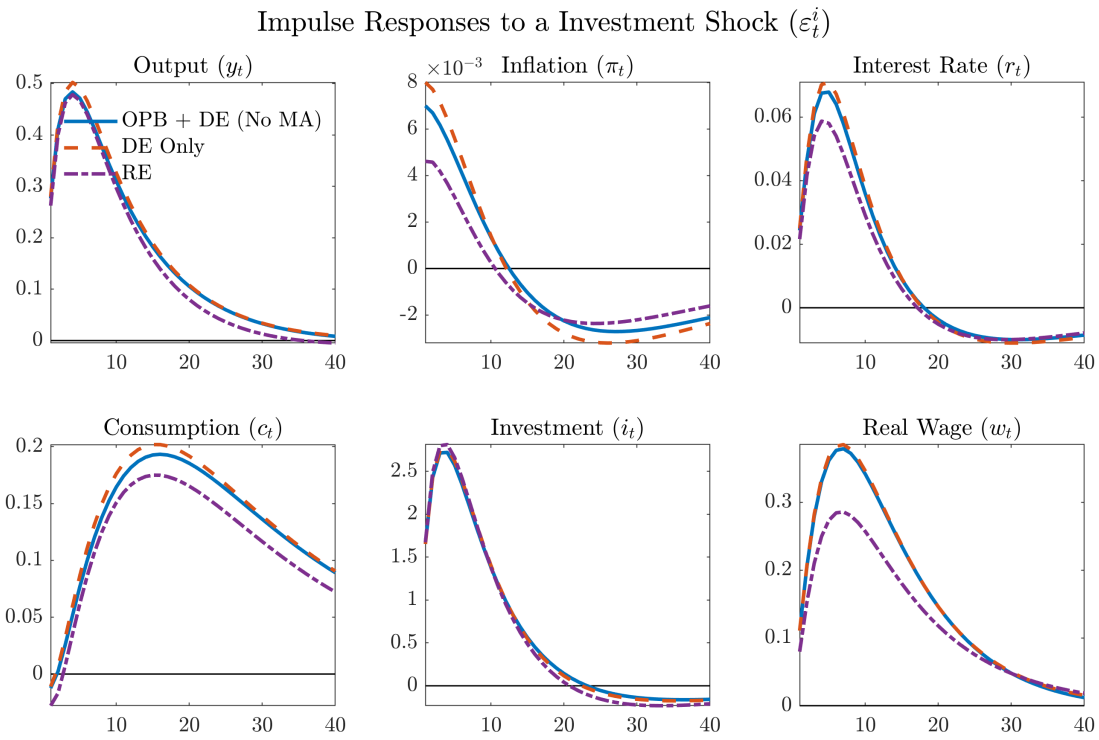


Figure 7: Impulse responses to an investment-specific technology shock (ε_t^i).

Notes: See notes to Figure 1.

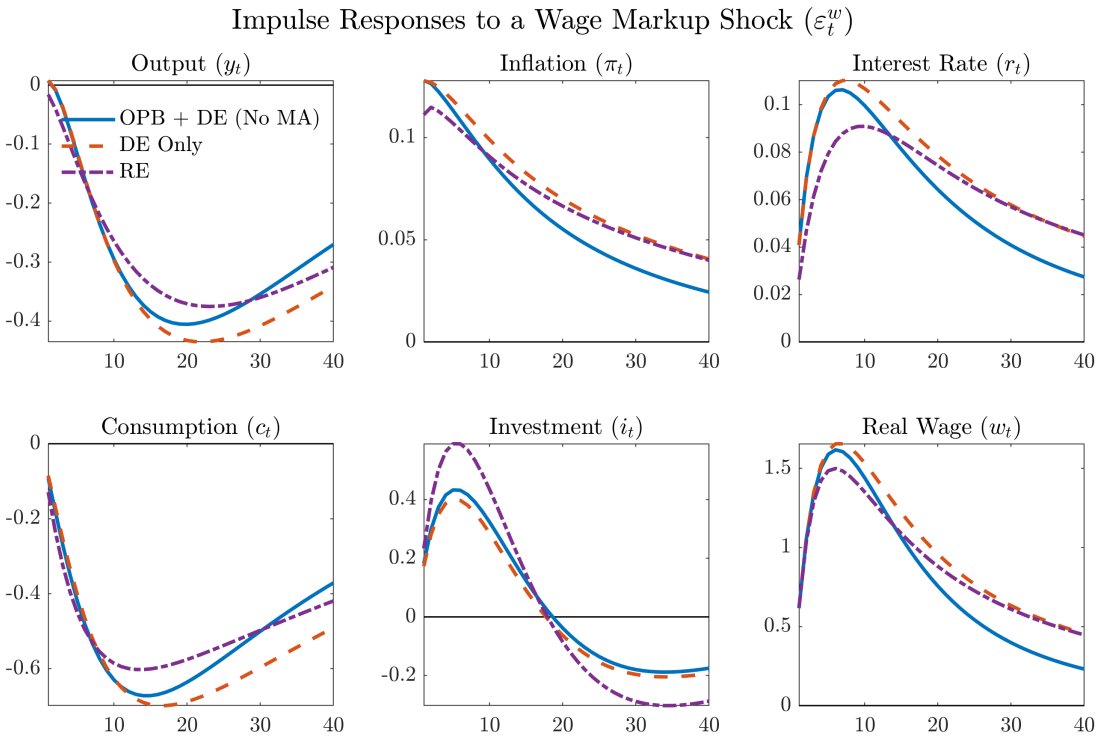


Figure 8: Impulse responses to a wage markup shock (ε_t^w).

Notes: See notes to Figure 1.