Higher-Order Perturbation in Sequence Space: the Certainty Correspondence

Adrien Auclert* Rodolfo Rigato† Matthew Rognlie‡ Ludwig Straub§

July 2026

Abstract

We show that higher-order perturbation solutions to rational expectations models exhibit a *certainty correspondence* property: they can be obtained from the derivatives of the model's perfect foresight solution. This extends the well-known certainty equivalence result beyond first order. Our approach gives the same answer as state-space perturbation and gets very close to the global solution when the latter is feasible, but for models with large idiosyncratic state spaces, it is significantly more practical than both. We apply our method to solve a HANK model and a menu cost model to third order in aggregate risk. We find significant time variation in the fiscal multiplier and the responsiveness of inflation to costs, including due to fluctuations in the effects of perceived aggregate risk. Third-order perturbation appears to capture the relevant nonlinearities for business cycle applications of heterogeneous-agent models.

*Stanford University, CEPR and NBER. Email: aauclert@stanford.edu.

†European Central Bank. Email: rodolfo.dinis_rigato@ecb.europa.eu.

‡Northwestern University and NBER. Email: matthew.rognlie@northwestern.edu.

§Harvard University, CEPR and NBER. Email: ludwigstraub@g.harvard.edu.

This research is supported by the National Science Foundation grant number SES-234393 and the Harvard Chae Initiative. The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank.

1 Introduction

Heterogeneous-agent models with aggregate risk are notoriously difficult to solve in general equilibrium. A recent literature (Boppart, Krusell and Mitman 2018, Auclert, Bardóczy, Rognlie and Straub 2021) has made progress by exploiting *certainty equivalence*: the fact that to first order in shock size, aggregate risk is neutral, and future variables can be replaced by their expectations. A consequence is that to first order, the impulse response to an "MIT shock"—a one-time, unexpected shock, with perfect foresight thereafter—is the same as in a stochastic model. Such impulse responses can be obtained with sequence-space methods, sidestepping the curse of dimensionality that often arises in the state space, and enabling faster and easier computation.

The restriction to first-order aggregate risk, however, is currently a major drawback of the sequence-space approach. In a first-order approximation, there is no role for size- or history-dependent effects, nor is there any precautionary response to risk. It is impossible to study how aggregate shocks, or policies responding to them, affect welfare.

In this paper, we overcome these limitations by extending sequence-space methods to higher order. We show that certainty equivalence can be generalized to what we call the *certainty correspondence*: a systematic relationship between MIT-shock impulse responses and the higher-order perturbation solution. We derive this explicitly for the second- and third-order cases, showing how each term can be obtained via the certainty correspondence, and apply these methods numerically to obtain the full third-order solution of a HANK model and a menu cost model. Our perturbation solutions coincide with the state-space approach with pruning (Andreasen, Fernández-Villaverde and Rubio-Ramírez 2018), and can be easily used to derive generalized impulse response functions and moments of the stochastic solution. The new approach is straightforward to implement for anyone who can calculate nonlinear impulses to MIT shocks, and we provide additional tools to speed up general equilibrium. It is also very accurate when compared to alternative methods, when those are available.

The foundation of our approach is the nonlinear sequence-space solution $X_t = \mathcal{X}(\sigma, \epsilon_t, \epsilon_{t-1}, \ldots)$, which expresses the path of an economy as a function of anticipated risk σ and the history $\epsilon_t, \epsilon_{t-1}, \ldots$ of realized shocks.[1] Whenever a locally stable state-space solution $X_t = g(X_{t-1}, \epsilon_t, \sigma)$ exists, a sequence-space solution also exists, and we can use perturbation to approximate it around the deterministic steady state. To first order, this is exactly the approach advocated by Boppart et al. (2018): since $\mathcal{X}_\sigma = 0$ (certainty equivalence), the first-order solution is determined by the derivatives $\mathcal{X}_j$ with respect to past shocks ϵ_{t-j}, which can be identified with the first-order impulse response to an MIT shock.

Moving to second order, certainty equivalence no longer holds, since there is a constant risk term $\mathcal{X}_{\sigma\sigma}$. All other terms in the second-order expansion, however, are independent of risk σ and can be obtained by differentiating impulse responses to MIT shocks. We show that, remarkably,

[1] Lan and Meyer-Gohde (2013b) also study such a solution, which they call a "nonlinear moving average", and Lombardo and Uhlig (2018) study a similar function where the history is finite—starting at date 0—rather than infinite. We discuss the relation to these papers further in the related literature section.

the risk term $\mathcal{X}_{\sigma\sigma}$ can *also* be obtained from the information in MIT-shock solutions: it is the effect from perturbing steady-state expectations by the size-dependence term $\mathcal{X}_{00}$. This is the second-order certainty correspondence.

With the second-order sequence-space solution in hand, we can read off various concepts of interest. For instance, $\mathcal{X}_{\sigma\sigma}$ gives the risky steady state: the steady state when risk is anticipated, but no shocks are ever realized. Summing this with the size-dependent terms, $\mathcal{X}_{00} + \mathcal{X}_{11} + \ldots$, gives the second-order ergodic mean of $X_t - X$. Second-order impulse responses are almost identical to their MIT shock counterparts, but with a correction reflecting expected size dependence.

We follow a similar approach for the third-order perturbation solution. Most terms in the third-order expansion of $\mathcal{X}$ are independent of risk and can be obtained from MIT shocks. The risk-dependent terms have the form $\mathcal{X}_{\sigma\sigma j}$, and capture how risk alters the transmission from ϵ_{t-j} to X_t. In part, these terms reflect the economy starting at the risky steady state $\mathcal{X}_{\sigma\sigma}$; this contribution is easy to obtain by shifting the steady state at which first-order impulses are calculated. The risk-dependent terms also reflect interactions between size and history dependence—which can again be obtained from MIT shocks, completing the third-order certainty correspondence.

As previously shown by Lombardo and Uhlig (2018), our sequence-space perturbation solution agrees with the pruned state-space solution. To obtain approximate moments of the stochastic solution, the conventional approach is to directly evaluate these moments on the perturbation approximation, as in Andreasen et al. (2018). We propose a different approach: a formal perturbation of the nonlinear moments themselves. This yields parsimonious formulas for each moment: for instance, evaluating covariances at any horizon using the first-order solution as in Auclert et al. (2021) is accurate to third order, and augmenting this with some terms from the second- and third-order solutions increases accuracy to fifth order.

The only requirement of our approach is the ability to compute nonlinear responses to MIT shocks, and to repeatedly differentiate them around the steady state—possibly via either numerical or automatic differentiation. Although we do not require that these responses are obtained with any particular solution method, one can exploit analytical results to improve upon a naive approach. For heterogeneous-agent models, nonlinear impulse responses are typically obtained by reducing the problem to a limited number of unknown sequences, then iterating until convergence with a variant of Newton's method. This iterative process, however, is unnecessary for a perturbation solution. Instead, we show that the nth-order solution can be obtained by evaluating equilibrium conditions on the $n-1$th-order solution, finding the nth-order error, and then applying the same inverse Jacobian used to solve the first-order system. This "one-shot principle" makes higher-order solutions much easier to obtain.

Throughout the paper, we apply our methods to a number of models. Our primary application is a standard though "smooth" heterogeneous-agent model, first in partial and then in general equilibrium.[2] In partial equilibrium, we study the response to income shocks—finding that,

[2]The "smooth" model is an otherwise standard incomplete market (SIM) model with additional i.i.d. shocks to income drawn from a smooth density. As we discuss, this feature of the model is important to make the model differentiable to the second and higher order, overcoming the difficulties with the non-smooth model highlighted by

as expected from the concavity of the consumption function, the second-order term for contemporaneous consumption is negative. Interestingly, however, we find that the second-order terms for consumption in subsequent periods are *positive*—so that, as we scale up an income shock, the delayed effect on consumption increases more than proportionally. This result extends to a general equilibrium HANK setting where income fluctuates due to deficit-financed transfers. As argued in Auclert, Rognlie and Straub (2023), these intertemporal spillovers may be partly responsible for sustained high demand in the US after the 2020–21 fiscal interventions; the fact that these intertemporal spillovers are even stronger for large shocks strengthens this argument, given the unprecedented size of these interventions. We also study how the impact transfer multiplier varies with the level of current economic activity, finding that the multiplier can be four times larger in recessions, when the economy is depressed, balance sheets are weak, and MPCs are high, consistent with the empirical literature on the state dependence of fiscal multipliers (eg Auerbach and Gorodnichenko 2012).

Our second main application is a menu cost model with idiosyncratic shocks in the spirit of Golosov and Lucas (2007). For that model, we study how the responsiveness of inflation to nominal marginal cost is affected by the current level of inflation. In line with other recent studies that use similar models (e.g. Blanco, Boar, Jones and Midrigan 2024), we find a significantly higher responsiveness of inflation to costs in times of high inflation.[3]

We conduct extensive accuracy checks throughout the paper. First, we test our third-order sequence-space perturbation routines on a small state space by showing that, in the neoclassical growth model, they deliver exactly the same solution as third-order state-space perturbation. Second, we compare our partial equilibrium heterogeneous-agent model solution to that obtained from a global solution, which is attainable via a Bellman approach.[4] We find that third order perturbation gets very close to the global solution. Importantly, we also find that the impulse response to the third order is closer to the global solution than the perfect-foresight solution is, because the third-order solution captures the effect of aggregate precautionary savings on MPCs, which the perfect-foresight solution misses. Finally, we show that, for the part of the solution that ignores aggregate precautionary effects, the third order solution is extremely close to the nonlinear perfect-foresight solution for shocks of plausible magnitude, persistence, and distribution.

Related literature. The closest existing work to ours is Lan and Meyer-Gohde (2013b), which studies a nonlinear sequence-space solution of the same form (a "nonlinear moving average") and its perturbation solution up to third order, which is known as a Volterra expansion. Beyond this similar core, however, the two papers diverge: Lan and Meyer-Gohde use techniques that parallel the state-space solution,[5] while we introduce the certainty correspondence. As we discuss

Bhandari, Bourany, Evans and Golosov (2023).

[3]We also verify that our formulas and our implementation of them are correct by checking that they produce the same solution as state-space perturbation up to third order on a neoclassical growth model.

[4]We thank Ben Moll for this suggestion.

[5]As they note in footnote 21: "Thus, our nonlinear moving average solution parallels nonlinear state space solutions in a manner analogous to the linear case, where the recursion is in the coefficients as opposed to the variables

further in section 2.4, the certainty correspondence does not hold in their framework, because of a difference in where expectations are taken.

Another closely related paper is Lombardo and Uhlig (2018), which uses a nonlinear sequence-space solution to explain why the state-space perturbation solution should be pruned: namely, because this makes it equivalent to the perturbed sequence-space solution.[6] We re-derive this important result in our setting. Unlike our paper and Lan and Meyer-Gohde, however, Lombardo and Uhlig only track the finite history since time 0, which makes their sequence-space solution vary with t.

Our paper also connects to the broader state-space perturbation literature, including the seminal work of Schmitt-Grohé and Uribe (2004) and Kim, Kim, Schaumburg and Sims (2008), and the pruned state-space system of Andreasen et al. (2018). Several recent papers, including Bilal (2023), Reiter (2023), and Bayer, Luetticke, Weiss and Winkelmann (2026b), have extended state-space perturbation to the heterogeneous-agent context. For rich enough heterogeneity, however, this can require significant "model reduction"—as the size of a second-order state-space solution is cubic in the dimension of the state, and becomes prohibitively large and costly when this dimension is high. The recent work of Bhandari et al. (2023) provides a hybrid of state- and sequence-space approaches, and derives a second-order Volterra expansion for heterogeneous-agent models. To our knowledge, our work is the first to point out the certainty correspondence result, and also to obtain a third-order solution, without model reduction, in the presence of substantial heterogeneity.[7]

Finally, our paper is connected to a literature that seeks "global" solutions to heterogeneous-agent models to study the effect of aggregate risk beyond first order. This literature uses a variety of approaches, such as neural networks (Kase, Melosi and Rottner 2022), deep learning (Han, Yang et al. 2021, Azinovic-Yang and Žemlička 2025, Druedahl, Huleux and Røpke 2026), structural reinforcement learning (Yang, Wang, Schaab and Moll 2025), and alternative simulation-based methods (Lee 2025). We view this literature as highly complementary. Our test against a global partial equilibrium solution performs well. However, global solutions may be useful for general equilibrium models with sufficiently extreme nonlinearities, though they often require additional assumptions and substantial computational power—while our third-order solution uses an unreduced HANK model, which we demonstrate can be solved much more practically in the sequence-space than the state-space. The extent to which global solutions differ from our third-order solution for typical heterogeneous-agent models remains an intriguing open question.

Bianchi and Kaplan (2026) have recently argued that third-order perturbation solutions are unlikely to be reliable in heterogeneous-agent models with high MPCs. We show that this is much too strong a statement. First, we show in appendix B why it is critical to do third-order

themselves."

[6] See also related work in Lan and Meyer-Gohde (2013a).

[7] Auclert, Rognlie, Straub and Tapak (2024b) use a second-order perturbation approach to solve for zero-th order portfolios and steady-state risk premia in a model with portfolio choice. The methods in that paper could be usefully combined with those from this paper to solve for time-varying risk premia.

perturbation on a model that admits a smooth density, which a model with only discrete income shocks does not have. Bayer, Briglia, Luetticke, Weiss and Winkelmann (2026a) make a similar point, and propose an alternative solution based on the method in Bayer et al. (2026b). Second, we show in appendix C that a third order solution performs extremely well against a nonlinear solution except if the shocks are extremely large, lump-sum, one-time, and if only their impact effect is considered.

2 The certainty correspondence

2.1 General setup

We consider a model driven by an exogenous shock process ϵ_t with i.i.d. innovations.[8] In the model, the endogenous variable $X_t \in \mathbb{R}^n$ is determined at t to solve

$$F\left(X_{t-1}, X_t, \mathbb{E}_t\left[X_{t+1}\right], \epsilon_t\right) = 0, \tag{1}$$

with F mapping $\left(\mathbb{R}^n\right)^3 \times \mathbb{R}$ to $\mathbb{R}^n$. Our results apply for any n, but they are especially useful when n is large—for instance, in discretized heterogeneous-agent models, where typically $n \geq 1000$.

The expectations in (1) are taken assuming that $\epsilon_t = \sigma\bar{\epsilon}_t$, where $\bar{\epsilon}_t$ are drawn i.i.d. from a distribution that is symmetric around zero, has unit variance, and has bounded support, and σ is a scalar that indexes the standard deviation of shocks. The actual, realized ϵ_t are given by $\epsilon_t = \sigma_R\bar{\epsilon}_t$. Our objective is to solve the rational expectations case with $\sigma = \sigma_R$, but our "certainty correspondence" result will connect this case to perfect-foresight or "MIT shock" impulse responses, where $\sigma = 0$ but $\sigma_R \neq 0$.

A *deterministic steady state* of the model is a constant vector $X \in \mathbb{R}^n$ that solves

$$F\left(X, X, X, 0\right) = 0.$$

A *nonlinear sequence-space solution* of the model around X is a continuous function $\mathcal{X}$ satisfying $\mathcal{X}(0, 0, 0, \ldots) = X$ such that

$$X_t = \mathcal{X}\left(\sigma, \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right) \tag{2}$$

solves equation (1) for all σ and $\{\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\}$ in some neighborhood of 0. This solution expresses the path of X_t as a function of anticipated risk and the history of realized shocks.[9]

A *nonlinear state-space solution* of the model around X is a continuous function g satisfying $g(X, 0, 0) = X$ such that

$$X_t = g\left(X_{t-1}, \epsilon_t, \sigma\right) \tag{3}$$

[8] We assume a single exogenous shock for notational convenience. Results generalize in a straightforward way to multiple exogenous shocks; see appendix A.10 for details.

[9] Note that σ_R does not directly appear in (2): given σ, only the realized sequence of ϵ_t matters for X_t. This property is also satisfied by the state-space solution (3). Continuity is defined using the ℓ^∞ norm.

solves equation (1) for all X_{t-1}, ϵ_t, and σ in some neighborhood of $(X, 0, 0)$. In this more common solution concept, the history of past shocks is summarized by the lagged state X_{t-1}. We say that g is *locally stable* if it is differentiable around $(X, 0, 0)$, and its derivative g_X there has all eigenvalues inside the unit circle.

We next make the following assumption, which parallels the usual conditions assumed in the linear rational expectations literature (Blanchard and Kahn 1980, Klein 2000). Here F_1, F_2, F_3 denote the partial derivatives of F at the steady state with respect to its first three arguments.[10]

Assumption 1. *F is continuously differentiable in a neighborhood of the deterministic steady state, and:*

- $\det(F_1 + \lambda F_2 + \lambda^2 F_3) = 0$ *has n roots $|\lambda| < 1$ inside the unit circle, counted including algebraic multiplicity, with no roots $|\lambda| = 1$ on the unit circle.*
- *The rank condition (A.1) is satisfied.*

This assumption guarantees the existence and uniqueness of *both* a sequence-space solution and a locally stable state-space solution.[11]

Proposition 1. *Assume that assumption 1 holds. Then both a nonlinear sequence-space solution $\mathcal{X}$ and a locally stable nonlinear state-space solution g exist and are unique in a neighborhood of the steady state. Furthermore, they are equivalent to each other in the sense that*

$$\mathcal{X}\left(\sigma, \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots\right) = g\left(\mathcal{X}\left(\sigma, \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \ldots\right), \epsilon_t, \sigma\right) \tag{4}$$

If F is $\mathcal{C}^r$, then g and $\mathcal{X}$ are also $\mathcal{C}^r$.

In the state space, "global" solution methods attempt to solve for the entire function g. Since this can be difficult, a popular alternative, known as state-space perturbation, is to solve instead for a Taylor expansion of the function g in (3) around the deterministic steady state $(X, 0, 0)$. (See, e.g., Schmitt-Grohé and Uribe 2004, Fernández-Villaverde, Rubio-Ramírez and Schorfheide 2016.) Our sequence-space perturbation approach parallels state-space perturbation: since solving for the function $\mathcal{X}$ directly is complicated, we will solve for a Taylor expansion of the $\mathcal{X}$ function in (2) around the deterministic steady state $(0, 0, 0, 0, \ldots)$.[12]

Proposition 1 implies that state-space and sequence-space perturbation solutions are closely connected. We review this connection, which has previously been established by Lan and Meyer-Gohde (2013b) and Lombardo and Uhlig (2018), in appendix A.9, where we use it to validate our

[10]We show in appendix A.2 that assuming continuous differentiability of F, the rest of this assumption is equivalent to the invertibility of the sequence-space Jacobian system formed by the derivatives of F, which can be tested following Auclert, Majic, Rognlie and Straub (2025).

[11]The state-space solution has a nonlinear vector autoregressive (VAR) form while the sequence-space solution has a nonlinear vector moving average (VMA) form. It is a well-known result from time series (eg Hamilton 1994, chapter 10) that, under a stability condition, linear VAR and VMA are equivalent representations of the same process. Proposition 1 shows that this equivalence also holds for nonlinear solutions of the model in (1).

[12]Azinovic-Yang and Žemlička (2025) use deep learning to solve globally for a truncated version of $\mathcal{X}$.

results. Our main theoretical contribution in this paper will be to further connect the sequence-space perturbation solution to the perfect-foresight solution, $\mathcal{X}(0, \epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots)$.

We finally note that since $\bar{\epsilon}_t$ is distributed symmetrically around zero, the distributions of $\sigma\bar{\epsilon}_t$ and $-\sigma\bar{\epsilon}_t$ are identical. This implies the following lemma, which we will use frequently.

Lemma 1 (Symmetry in σ)**.** Both g and $\mathcal{X}$ are even functions of σ. Hence every derivative containing an odd number of differentiations with respect to σ vanishes at $\sigma = 0$, whenever that derivative exists.

2.2 Examples

We will illustrate the results in this section using three simple model examples. Section 4 discusses more complex models.

Example 1: The Neoclassical Growth Model. Our first example is a neoclassical growth model with shocks, as in e.g. Schmitt-Grohé and Uribe (2004). We will use this classic model to verify that our methods deliver the same results as standard state-space perturbation. In this model, a planner maximizes the utility $\mathbb{E}\sum \beta^t u(C_t)$ of a representative agent, with CRRA period utility $u(C) = C^{1-\gamma}/(1-\gamma)$, subject to the feasibility constraint $C_t + K_t - (1-\delta)K_{t-1} = e^{a_t}K_{t-1}^{\alpha}$, and an AR(1) shock process for log TFP $a_t = \rho_a a_{t-1} + \epsilon_t$. The optimum in this model is characterized by an Euler equation $C_t^{-\gamma} = \beta\mathbb{E}_t\left(1 - \delta + \alpha e^{a_{t+1}}K_t^{\alpha-1}\right)C_{t+1}^{-\gamma}$.

To put this in the notation of (1), we define $\lambda_t \equiv \left(1 - \delta + \alpha e^{a_t}K_{t-1}^{\alpha-1}\right)C_t^{-\gamma}$ to be the shadow value of capital at t, and define $X_t \equiv (a_t, \lambda_t, K_t, C_t)$. Then we can stack

$$F\left(X_{t-1}, X_t, \mathbb{E}_t\left[X_{t+1}\right], \epsilon_t\right) \equiv \begin{bmatrix} \lambda_t - \left(1 - \delta + \alpha e^{a_{t+1}}K_t^{\alpha-1}\right)C_{t+1}^{-\gamma} \\ C_t^{-\gamma} - \beta\mathbb{E}_t\left[\lambda_{t+1}\right] \\ C_t + K_t - (1-\delta)K_{t-1} - e^{a_t}K_{t-1}^{\alpha} \\ a_t - \rho_a a_{t-1} - \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5}$$

Fos simulations, we use a quarterly calibration with $\gamma = 2$, $\alpha = 0.36$, $\delta = 0.025$, and $\beta = 0.995$, and a log TFP process with persistence $\rho_a = 0.9$ and innovation standard deviation $\sigma = 0.1$.

Example 2: The Smooth Standard Incomplete Markets (SIM) Model. We next consider a partial equilibrium heterogeneous-agent model in which consumers face idiosyncratic income risk and a borrowing constraint. This model is convenient because it can be solved globally in the state space, and so we can verify the accuracy of our perturbation against this global solution. The Bellman equation for a consumer with assets a, persistent income state e, and discount factor β is

$$\begin{aligned} V_t(a, e, \beta) \quad &= \quad \mathbb{E}_\xi\left[\max_{c,a'} u(c) + \beta\mathbb{E}_t[V_{t+1}(a', e', \beta')|e]\right] \\ \text{s.t.} \qquad & a' + c = (1+r)a + Z_t\xi e + Tr; \qquad a' \geq \underline{a} \end{aligned} \tag{6}$$

where $u(\cdot)$ continues to be CRRA utility, r is the interest rate, Tr is a constant transfer, and non-transfer income $Z_t \xi e$ is made up of three parts: an aggregate component Z_t, which follows an AR(1) process, $Z_t - Z_{ss} = \rho_Z (Z_{t-1} - Z_{ss}) + \epsilon_t$ with mean Z_{ss}, a persistent idiosyncratic component e, which follows a finite Markov chain with a fixed transition matrix Π^e, and an independent transitory idiosyncratic component ξ, which is i.i.d. smoothly distributed with density $f(\cdot)$ realized at the beginning of the period, with $\mathbb{E}[\xi] = \mathbb{E}[e] = 1$. Patience β evolves independently of income according to a finite Markov chain with a fixed transition matrix Π^β, with the overall transition matrix between discrete states (e, β) denoted by $\Pi = \Pi^e \otimes \Pi^\beta$.

Appendix A.4 shows that this model can be put into the form in (1) by stacking the marginal value function at each discretization point into a vector $\boldsymbol{v}_t$, the distribution at each discretization point into a vector $\boldsymbol{\phi}_t$, and then forming the large vector $X_t \equiv (\boldsymbol{v}_t, \boldsymbol{\phi}_t, C_t, Z_t)$, where C_t is aggregate consumption (the integral of the consumption policy across the distribution) such that

$$F\left(X_{t-1}, X_t, \mathbb{E}_t\left[X_{t+1}\right], \epsilon_t\right) \equiv \begin{bmatrix} \boldsymbol{v}_t - \mathcal{V}(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1}) \\ \boldsymbol{\phi}_t - \Lambda(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1}) \\ C_t - \mathcal{C}(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1}) \\ Z_t - Z_{ss} - \rho_Z (Z_{t-1} - Z_{ss}) - \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

for some functions $\mathcal{V}$, Λ, and $\mathcal{C}$ described in the appendix.

Appendix B provides more details for both the underlying smooth model and its computation. The appendix emphasizes the importance of the smooth ξ shock in making the overall model smooth: here, $V_{a,t}$ and Φ_t are both infinitely differentiable functions of a, and C_t is an infinitely differentiable function of Z_t. By contrast, the model with a discretized ξ shock is not second-order differentiable. As emphasized by Bhandari et al. (2023) and illustrated by Bianchi and Kaplan (2026), using a non-smooth model complicates the implementation of higher-order perturbation.

Our calibration, which is at a quarterly frequency, mostly follows Auclert, Rognlie and Straub (2024a), except for the addition of the smooth i.i.d. income shocks. We describe this calibration in appendix A.4, and summarize our calibration parameters in table A.4.

Example 3: The Smooth Intertemporal Keynesian Cross (IKC) Model. Finally, we develop a New Keynesian general equilibrium version of Example 2, along the lines of the "intertemporal Keynesian cross" model in Auclert et al. (2024a). This model constitutes a natural use case for our method: it requires solving endogenously for the market-clearing level of output at each point in time, so it does not admit a natural global solution, and it features high n, so it also cannot easily be handled with state-space perturbation.

The underlying partial equilibrium problem is the same as in Example 2: households still solve the Bellman equation (6), except that now $Z_t = (1 - \tau_t) Y_t$ is equal to production net of taxes. Production is from labor, $Y_t = N_t$, nominal prices are flexible, nominal wages are sticky, and monetary policy maintains the ex-ante real interest rate equal to r at all times. The government's

Parameter	Description	Value
Standard incomplete markets model (partial equilibrium)		
r	Real interest rate	0.005
ρ_e	Persistence of persistent component	0.9767
σ_e	Innovation s.d. of persistent component	0.1964
σ_ξ	Innovation s.d. of transitory component	0.0964
β_L	Low type discount factor	0.95996
β_H	High type discount factor	0.99049
Z	Aggregate non-transfer income	0.67
Tr	Transfer income	0.14
A	Aggregate assets	4
$\partial C_0/\partial Z_0$	Aggregate MPC	0.2
ρ_Z	Aggregate income persistence	0
σ	Aggregate income innovation s.d.	0.05
IKC model (general equilibrium)		
Y	GDP	1
B	Government bonds	4
G	Government spending	0.17
τ	Marginal tax rate	0.33
ρ_B	Bonds persistence	0.95
σ	Bonds innovation s.d.	0.05

Table 1: Baseline calibration of heterogeneous-agent models

budget constraint is

$$B_t + Tr + G = (1+r)B_{t-1} + \tau_t Y_t$$

The level of bonds B_t is stochastic, and follows an AR(1) process $B_t - B = \rho_B(B_{t-1} - B) + \sigma\bar{\epsilon}_t$, with the government adjusting the tax rate τ_t each period to meet this bond target. In equilibrium, the goods market must clear, so $C_t + G_t = Y_t$, and the asset market must clear, so $A_t = B_t$.

Appendix A.4 describes the calibration of this model—which is summarized in table 1—and shows that it can be put in the form (1), described by equation (A.19). The important difference relative to (7) is the addition of an aggregate market clearing condition. This reflects a key difficulty of general equilibrium: due to the endogeneity (here) of aggregate income Y_t, the model can no longer be easily solved globally in the state space.

Summary. It is important to note that in all of (5), (7) and (A.19), all expectations at time t show up only through $\mathbb{E}_t[X_{t+1}]$, in line with our form for F: in (5), through $\mathbb{E}_t\lambda_{t+1}$, and in (7) and (A.19), through $\mathbb{E}_t v_{t+1}$. It is usually possible to cast models in this form, which emerges naturally in dynamic programming—where the future appears through the expectation of a value function or its derivative. This feature of F will be crucial to deriving the higher-order certainty correspondence.

2.3 First-order perturbation and certainty equivalence

We begin with a first-order Taylor approximation of the sequence-space solution $\mathcal{X}$ in (2) around the deterministic steady state $(\sigma, \epsilon_t, \epsilon_{t-1} \ldots) = (0, 0, 0, \ldots)$ (or $\mathbf{0}$ for short). We write $\mathcal{X}_\sigma \equiv \frac{\partial \mathcal{X}}{\partial \sigma}(\mathbf{0})$ and $\mathcal{X}_j \equiv \frac{\partial \mathcal{X}}{\partial \epsilon_{-j}}(\mathbf{0})$. Assuming differentiability, and noting that $\mathcal{X}_\sigma = 0$ by lemma 1, this first-order Taylor approximation is

$$X_t \approx X + \sum_{j=0}^{\infty} \mathcal{X}_j \epsilon_{t-j} \tag{8}$$

Hence, to first order, $X_t - X$ is $MA(\infty)$ in ϵ_t. An important feature of (8) is that σ does not appear, so that anticipated risk does not affect the first-order solution.

We next define the sequence

$$X_j^{MIT}(\nu) \equiv \mathcal{X}(0, 0, \ldots, 0, \underbrace{\nu}_{j}, 0 \ldots) \tag{9}$$

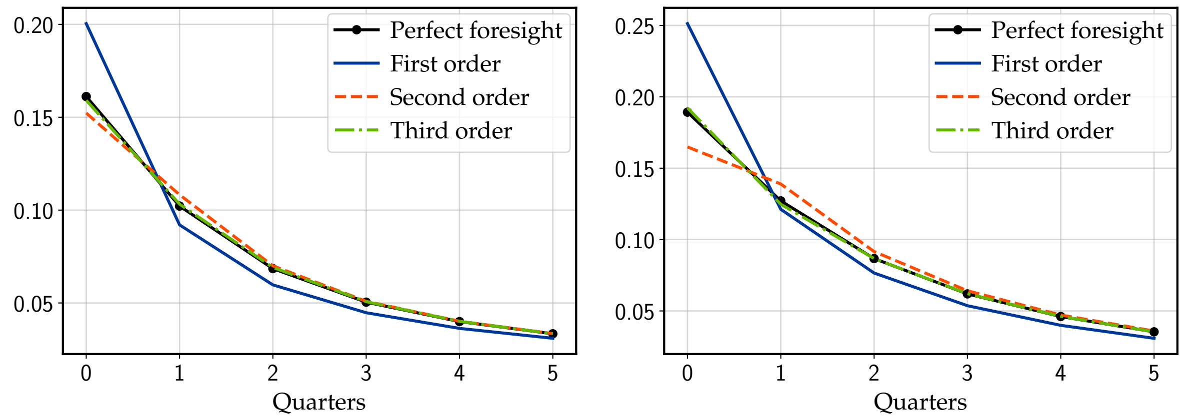
as the impulse response to an *MIT shock*: a one-time, unanticipated shock of size ν at date 0, starting from the steady state, and with perfect foresight in every period j thereafter. Because $\mathcal{X}_j$ is evaluated at the deterministic steady state, we have:

$$\mathcal{X}_j = (X_j^{MIT})'(0) \tag{10}$$

Hence, the $MA(\infty)$ coefficients in (8) can be obtained as the local impulse response to an MIT shock (see, e.g. Boppart et al. 2018). This is an especially helpful result for heterogeneous-agent models, since there are well-known procedures to calculate this MIT shock impulse response in these models, both in partial equilibrium where a simple backward-forward procedure suffices, or in general equilibrium where one can use the sequence-space Jacobian method (Auclert et al. 2021). We summarize this result as:

Proposition 2 (Certainty equivalence)**.** *The first-order perturbation of X_t in the sequence space is given by the linear MA(∞) process in (8). The constant term is the deterministic steady state X, and the coefficients on past innovations are given by (10), so they can be obtained from linearized perfect-foresight impulse responses.*

The blue lines in figure 1 visualize the first-order impulse response of consumption $\mathcal{X}_j$ to a one-time innovation to ϵ in our partial equilibrium model (left) and our general equilibrium model (right). In the PE model, a unit shock to ϵ increases income Z by 1 on impact, and since Z is i.i.d, households then expect it back to its mean. As a result, the left panel traces out the intertemporal marginal propensities to consume $\partial C_j / \partial Z_0$, with the impact effect calibrated to 0.2, and spending remaining elevated in the following quarters. In the GE model, a unit shock to ϵ lower taxes T_t by 1 on impact and raise them slightly going forward as bonds go back to the steady state at rate ρ_B. This triggers an intertemporal Keynesian cross mechanism (Auclert et al. 2024a), with equilibrium spending and income following the same pattern as iMPCs, but with a larger impact effect and



(a) PE (intertemporal MPCs)

(b) GE impulse response

Figure 1: Impulse response of consumption response over time, PE and GE

slightly more persistence.

The dark-dotted lines show the perfect foresight impulse response to an innovation of size $\nu = 0.10$ (10% of quarterly GDP, or 2σ), normalized by the size of the shock, $(C_j^{MIT}(\nu) - C)/\nu$. The normalized impact effect is lower, but the effect is also more persistent, in both partial and general equilibrium. This is a general feature of nonlinear iMPCs that we will come back to in the next section. For a shock of this magnitude, the difference between the first-order and the nonlinear perfect-foresight solution are significant, as pointed out by Bianchi and Kaplan (2026).

2.4 Second-order perturbation and certainty correspondence

We continue with a second-order Taylor approximation of $\mathcal{X}$ around the deterministic steady state. We now write $\mathcal{X}_{jk} \equiv \frac{\partial^2 \mathcal{X}}{\partial \epsilon_{-j} \partial \epsilon_{-k}}(\mathbf{0})$, and similarly for derivatives with respect to σ. Assuming second differentiability, and noting that lemma 1 also implies that any mixed derivatives $\mathcal{X}_{\sigma j}$ are zero, a second-order Taylor approximation of X_t, also known as the second-order Volterra expansion, can be written as

$$X_t \approx X + \sum_{j=0}^{\infty} \mathcal{X}_j \epsilon_{t-j} + \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sum_{j,k\geq 0} \mathcal{X}_{jk}\epsilon_{t-j}\epsilon_{t-k} \tag{11}$$

Hence, to second-order, up to constants, X_t is a quadratic infinite moving-average process in ϵ_t.

Using symmetry of mixed partials, the new second-order terms in (11) can be usefully grouped as follows:

$$\underbrace{\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2}_{\text{Risk}} + \underbrace{\frac{1}{2}\sum_{j=0}^{\infty}\mathcal{X}_{jj}\epsilon_{t-j}^2}_{\text{Size}} + \underbrace{\sum_{j=0}^{\infty}\sum_{k>j}^{\infty}\mathcal{X}_{jk}\epsilon_{t-j}\epsilon_{t-k}}_{\text{History}} \tag{12}$$

Now, anticipation of future risk shows up as a constant "risk" term.

Notions of steady states. In the presence of aggregate risk, there are two natural notions of "steady state". The *risky steady state* (Coeurdacier, Rey and Winant 2011) arises when agents continually anticipate aggregate risk, even though realized risk is $\sigma_R = 0$, so that $\epsilon_t = 0$ at all dates. Given (11), the system is then perpetually at $X + \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2$. The *ergodic mean* is the average value of the system when risk realizes at its expected value $\sigma_R = \sigma$. This mean can be immediately derived from (11) as $\mathbb{E}[X_t] = X + \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sum_{j=0}^{\infty}\mathcal{X}_{jj}\sigma^2$. While these notions all collapse to the deterministic steady state X in the first-order solution (8), they generally differ from it in the second-order solution.

Certainty correspondence. For the size and history terms, since the derivatives $\mathcal{X}_{jk}$ are evaluated at the deterministic steadys state, we can use the same connection to perfect foresight that we used for the first order term. First, we have:

$$\mathcal{X}_{jj} = (X_j^{MIT})''(0) \tag{13}$$

which connects the size effect in the quadratic MA expression for X_t to the size effect in the MIT shock impulse response. Further, for $l > 0$ we can define the sequence

$$X_j^{MIT,l}(\epsilon, \nu) \equiv \mathcal{X}(0, 0, \ldots 0, \underbrace{\nu}_{j}, 0, \ldots, 0, \underbrace{\epsilon}_{j+l}, 0 \ldots) \tag{14}$$

as the impulse response to a sequence of two consecutive MIT shocks: a first unexpected shock of size ϵ occurring at date $-l$, and a second unexpected shock of size ν occurring at date 0. Then we have that, for $k > j$,

$$\mathcal{X}_{jk} = \frac{\partial^2 X_j^{MIT,k-j}}{\partial\nu\partial\epsilon}(0,0) \tag{15}$$

which connects the history effect in the second-order Volterra expansion of X_t to the local interaction term of the impulse response to two consecutive MIT shocks, $k - j$ periods apart.

To complete the certainty correspondence result, we now show how to also obtain the risk term using only the perfect foresight solution. To do this, we introduce the concept of a *perturbed-expectations steady state*, which for any δ solves the equation

$$F\left(X^{pss}(\delta), X^{pss}(\delta), X^{pss}(\delta) + \mathcal{X}_{00}\delta, 0\right) = 0, \tag{16}$$

Observe that, by (13), $X^{pss}(\delta)$ can be solved just like the deterministic steady state, provided one replaces, everywhere they appear, occurrences of $\mathbb{E}_t\left[X_{t+1}\right]$ by the (yet unknown) steady state plus the immediate size dependence $(X_0^{MIT})''(0)$, scaled by δ. Then, appendix A.6 shows that:

$$\mathcal{X}_{\sigma\sigma} = (X^{pss})'(0) \tag{17}$$

The intuition is that we can solve for the risky steady state $X^* = X + \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2$ by perturbing expectations. Indeed, in the risky steady state $X_{t-1} = X_t = X^*$ every period but $\mathbb{E}_t[X_{t+1}] = X^* + \frac{1}{2}\mathcal{X}_{00}\sigma^2$ due to the expectation of size dependence. Hence, we can find $\mathcal{X}_{\sigma\sigma}$ by adding a constant perturbation to expectations proportional to $\mathcal{X}_{00}$ and re-equilibrating using steady-state routines, as (16) and (17) do. This leads us to our core result:

Proposition 3 (Second-order certainty correspondence)**.** *The second-order perturbation of X_t in the sequence space is given by the quadratic MA(∞) in (11). The new coefficients $\mathcal{X}_{jj}$, $\mathcal{X}_{jk}$ and $\mathcal{X}_{\sigma\sigma}$ are respectively given by equations (13), (15), and (17), so they can be obtained from second-order perfect-foresight impulse responses.*

Relation to pruned state-space and testing. As shown by Lombardo and Uhlig (2018), the second-order sequence space perturbation in (11) corresponds exactly to the so-called second-order pruned state space solution (Andreasen et al. 2018)—indeed, Lombardo and Uhlig (2018) derive a theory of pruning to any order based on this equivalence. In appendix A.9, we verify the accuracy of our routines by checking that this result holds. Specifically, we make sure that in our calibrated neoclassical growth model, the pruned state-space solution obtained using standard methods numerically coincides with our sequence-space solution.

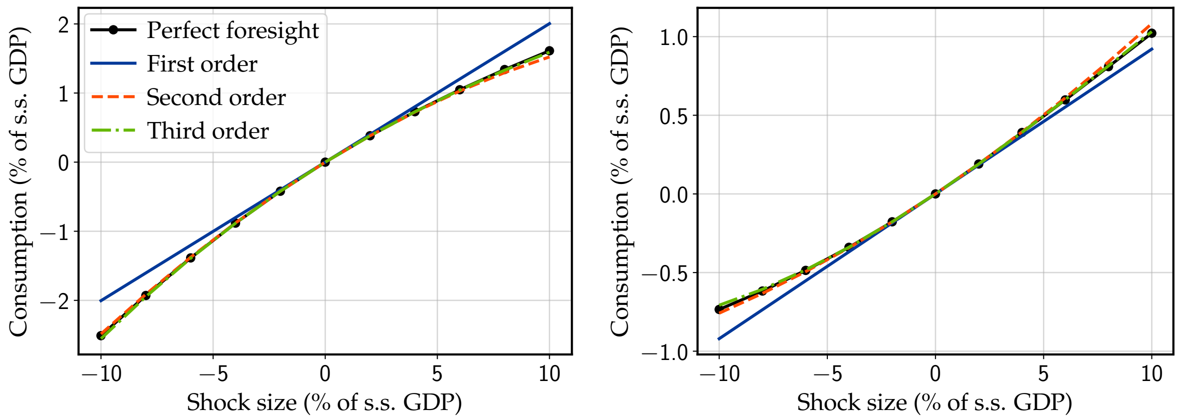
Relation to Lan and Meyer-Gohde (2013b). Lan and Meyer-Gohde (2013b) also pursue perturbation in the sequence space, and introduce the Volterra expansion in (11). Their paper obtains expressions for the coefficients in this expansion in terms of the derivatives of the F function, paralleling the usual approach to perturbation in the state space. By contrast, our result in proposition 3 shows that we can solve for these coefficients using perfect-foresight impulse responses, bypassing the need to explicitly obtain and manipulate derivatives of F. This is a significant improvement because obtaining these derivatives can be very complicated, especially in the case of heterogeneous-agent models.

Furthermore, certainty correspondence does not actually hold in Lan and Meyer-Gohde (2013b). This is because, as in much of the literature, Lan and Meyer-Gohde (2013b) write their model in the form

$$\mathbb{E}_t \left[F\left(X_{t-1}, X_t, X_{t+1}, \epsilon_t \right) \right] = 0. \tag{18}$$

Compared to (1), here the expectation operator on the outside of the F function. While (18) may appear to be strictly more general, dynamic programming usually enables us to rewrite the variables in X_t so that expectations only appear through the mean $\mathbb{E}_t[X_{t+1}]$, as argued in section 2.2. This reformulation is key to enabling us to obtain the $\mathcal{X}_{\sigma\sigma}$ term solely via perfect foresight in equation (17). Otherwise, as appendix A.6 shows, there is another term, reflecting the interaction of risk in X_{t+1} with the nonlinearity of F.

Application to heterogeneous-agent models. The orange dashed lines in figure 1 display the second-order normalized impulse response (the time path of $\mathcal{X}_j + \frac{1}{2}\mathcal{X}_{jj}\nu$ for aggregate consump-



(a) Response at $t = 0$

(b) Response at $t = 1$

Figure 2: Size dependence in the response of consumption to income shocks (PE model)

tion) for a shock of size $\nu = 10\%$ of quarterly GDP, in partial and general equilibrium.[13] This second-order impulse captures a key feature of the nonlinear solution: in both models, the impact effect is lower than the first order impact due to the concavity of the consumption function ($\mathcal{X}_{00} < 0$), but the effect in later periods is higher, reflecting convexity in later periods ($\mathcal{X}_{jj} > 0$ for $j > 0$). This is a manifestation of the intertemporal budget constraint of agents, which holds irrespective of shock size: any additional savings from period 0 must later be spent down as agents return to their buffer stock of assets.

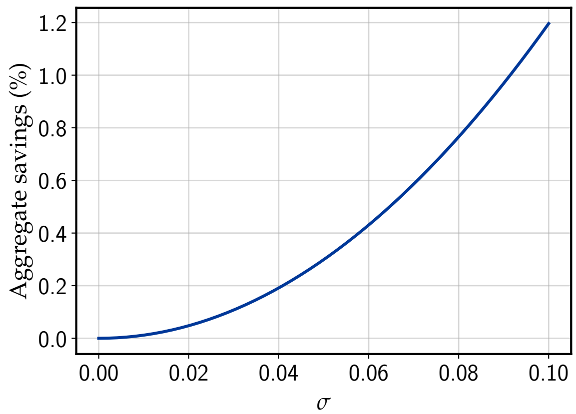
Figure 2 visualizes these results in a different way, by considering how the level of the consumption response net of the risky steady state varies with shock size on impact ($j = 0$, left panels) and at date 1 ($j = 1$, right panels) in the PE model (the results in the GE model are very similar and displayed in appendix figure A.2). Here again, the blue line represents the linear approximation, and the dashed orange line the second order approximation: this has a concave shape at date 0, and a convex shape at date 1, for the reason just discussed. Compared to the nonlinear perfect foresight solution $C_j^{MIT}(\nu) - C$ in the black line, for the range of shocks considered here (aggregate income shocks of up to 10% of quarterly GDP), the second order approximation does an excellent job, although the fit slightly deteriorates for very large shocks.

We next turn to the "risk" effect $\mathcal{X}_{\sigma\sigma}$. While aggregate uncertainty affects all the elements of X, we focus on the effect of σ on aggregate asset accumulation in the risky steady state in the partial equilibrium model,[14] with corresponding results for the general equilibrium model in appendix figure A.3. Certainty correspondence allows us to obtain this in a straightforward way: we obtain the impact size dependence $\mathcal{X}_{00}$ in the marginal value function v_t, since this is the only term entering with expectations in (7), and then solve for the risky steady state $\mathcal{X}_{\sigma\sigma}$ using (17).

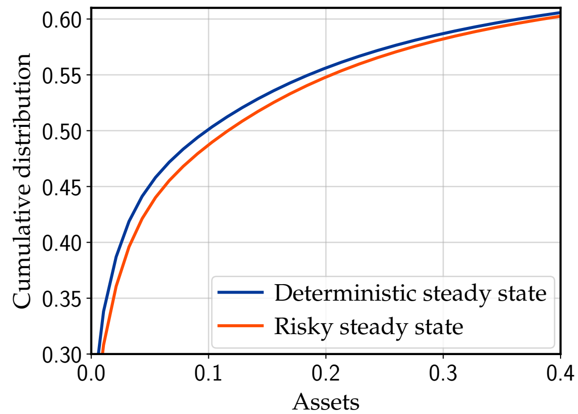
Figure 3 visualizes the outcome of this procedure. The left panel plots the effect of aggregate

[13]This is also the normalized 2nd-order generalized impulse response starting from the steady state (see section 2.6).

[14]Note that this is closely related to the effect on aggregate consumption, since the country budget constraint implies $C = Y + rA$.



(a) Aggregate assets



(b) Wealth distribution

Figure 3: The risky steady state: aggregate savings and aggregate income uncertainty (PE model)

risk on risky steady-state aggregate assets as a function of σ, in other words the term $\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2$ corresponding to total assets. The right panel plots the cumulative distribution function of assets, both at the deterministic steady state and at the risky steady state with $\sigma = 0.05$. Perceived aggregate risk makes agents save more at every point in the distribution, resulting in a rightward shift of the cdf and a significant increase in aggregate precautionary savings. Notably, this lowers the mass of agents at the borrowing constraint by a significant amount. In turn, this implies that MPCs are significantly lower in the risky steady state, compared to the deterministic steady state. This has important implications for the third order perturbation.

The left panel of Figure 4 concludes this section by visualizing the three new types of terms from the second-order perturbation as impulse responses to consumption. The dashed line shows the constant, positive effect on risky-steady state consumption from higher asset accumulation. The blue line shows the contemporaneous size effect $\frac{1}{2}\mathcal{X}_{jj}$, which is negative and positive, reflecting the concave-convex pattern from Figure 2. The orange line shows the lagged size effect $\frac{1}{2}\mathcal{X}_{j-1,j-1}$, which is positive everywhere with a decaying pattern, because large past shocks lead to additional asset accumulation, which is then spent down. Finally, the green line shows the history effect $\mathcal{X}_{j,j-1}$, which has the same pattern as the contemporaneous size effect: a past shock strengthens balance sheets and mitigates the impact consumption effect of a new shock, leading to more saving and therefore higher spending later on. One benefit of our perturbation approach is that all these terms can be separately calculated, rather than lumped together as a global solution would.

2.5 Third-order perturbation and certainty correspondence

We proceed with a third-order Taylor approximation of $\mathcal{X}$ around the deterministic steady state. We continue to write $\mathcal{X}_{jkk'} \equiv \frac{\partial^3 \mathcal{X}}{\partial\epsilon_{-j}\partial\epsilon_{-k}\partial\epsilon_{-k'}}(\mathbf{0})$, and similarly for derivatives with respect to σ. Assuming third differentiability, and using lemma 1 again to observe that all mixed derivatives $\mathcal{X}_{\sigma jk}$

are zero, a third-order Taylor approximation (Volterra expansion) of X_t can now be written

$$X_t \approx X_t^{(2)} + \frac{1}{2}\sum_{j=0}^{\infty} \mathcal{X}_{\sigma\sigma j}\epsilon_{t-j}\sigma^2 + \frac{1}{6}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty} \mathcal{X}_{jkk'}\epsilon_{t-j}\epsilon_{t-k}\epsilon_{t-k'} \tag{19}$$

where $X_t^{(2)}$ is the second-order Volterra expansion in equation (11). Hence, to third order, up to constants, X_t is a cubic infinite moving-average process in ϵ_t.

The new term on the right in (19) corresponds to all triple interactions of shocks, and using the symmetry of mixed partials again, can further be rewritten as

$$\frac{1}{6}\sum_{j=0}^{\infty} \mathcal{X}_{jjj}\epsilon_{t-j}^3 + \frac{1}{2}\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} \mathcal{X}_{jjk}\epsilon_{t-j}^2\epsilon_{t-k} + \frac{1}{2}\sum_{j=0}^{\infty}\sum_{k>j}^{\infty} \mathcal{X}_{jkk}\epsilon_{t-j}\epsilon_{t-k}^2 + \sum_{j=0}^{\infty}\sum_{k>j}^{\infty}\sum_{k'>k}^{\infty} \mathcal{X}_{jkk'}\epsilon_{t-j}\epsilon_{t-k}\epsilon_{t-k'}$$

These correspond, respectively, to cubic size dependence, the interaction of quadratic size dependence with history, the quadratic size history effect, and the history effect from the interaction of two past shocks. Just like with second-order certainty correspondence, these terms are all straightforwardly related to the corresponding MIT shock solution. Indeed, we have:

$$\mathcal{X}_{jjj} = (X_j^{MIT})'''(0) \tag{20}$$

and for $k > j$ we have

$$\mathcal{X}_{jjk} = \frac{\partial^3 X_j^{MIT,k-j}}{\partial^2\nu\partial\epsilon}(0,0) \qquad \text{and} \quad \mathcal{X}_{jkk} = \frac{\partial^3 X_j^{MIT,k-j}}{\partial\nu\partial^2\epsilon}(0,0) \tag{21}$$

while, for $l > 0$ and $l' > 0$ we can define the sequence

$$X_j^{MIT,l,l'}(\epsilon',\epsilon,\nu) \equiv \mathcal{X}(0,0,\ldots 0,\underbrace{\nu}_{j},0,\ldots 0,\underbrace{\epsilon}_{j+l},0,\ldots,0,\underbrace{\epsilon'}_{j+l+l'},0\ldots) \tag{22}$$

and then we have, for $k' > k > j$,

$$\mathcal{X}_{jkk'} = \frac{\partial^3 X_j^{MIT,k-j,k'-k}}{\partial\nu\partial\epsilon\partial\epsilon'}(0,0,0) \tag{23}$$

Equations (20), (21) and (23) show that all triple shock interaction terms in the third order Volterra expansion can be obtained via perfect-foresight impulse responses.

The additional new term in the third order sequence space perturbation (19) is the term from the $\epsilon_{t-j}\sigma^2$ interactions. This term quantifies both how the presence of aggregate risk affects the impulse response to shocks, and how the occurrence of past shocks affects the precautionary motive. Appendix A.7 shows that the overall effect is in fact the sum

$$\mathcal{X}_{\sigma\sigma j} = \left(\mathcal{X}_{\sigma\sigma j}\right)^{rss} + \left(\mathcal{X}_{\sigma\sigma j}\right)^{tvp}. \tag{24}$$

where *rss* stands for the risky steady state, and *tvp* for time-varying precautionary.

Given our routines to calculate the risky steady state, the *rss* term in (24) is simple to calculate using a perfect-foresight impulse response: obtain the first-order impulse response to a shock starting from the risky steady state distribution rather than the deterministic steady state, with all expectations modified using expected size dependence as in the risky steady state calculation. The *tvp* term, on the other hand, requires us to modify our concept of perturbed expectations to accommodate time variation in these perturbations. To solve for it, we consider the deterministic *perturbed-expectations transition* $\{X_j^{pt}(\delta)\}_{j\geq 0}$, which for any δ solves the equation

$$F\left(X_{j-1}^{pt}(\delta), X_j^{pt}(\delta), X_{j+1}^{pt}(\delta) + \mathcal{X}_{j+1,00}\delta, 0\right) = 0,$$

with initial condition $X_{-1}^{pt}(\delta) = X$ and terminal condition $X_j^{pt}(\delta) \to X$ as $j \to \infty$, and the $\mathcal{X}_{j+1,00}$ term calculated using (21). Then appendix A.7 shows that we have:

$$\left(\mathcal{X}_{\sigma\sigma j}\right)^{tvp} = (X_j^{pt})'(0) \tag{25}$$

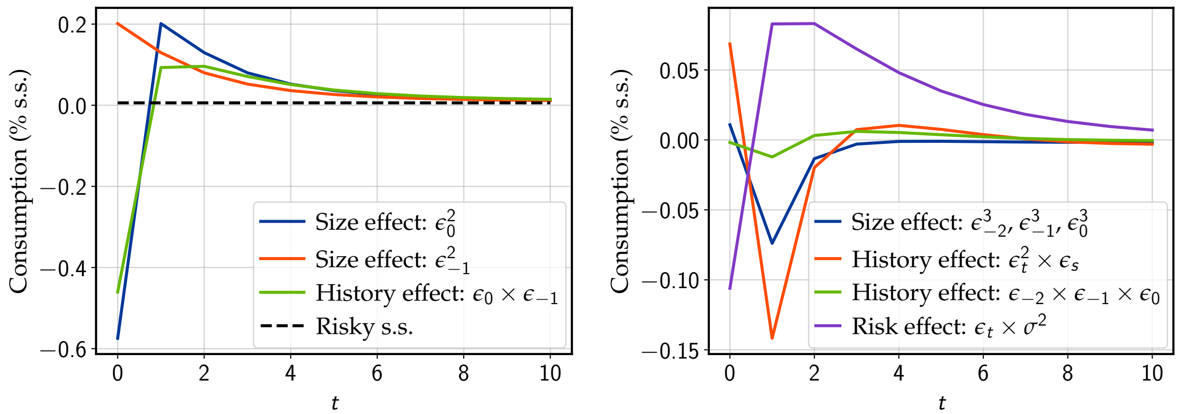
This completes our argument that we can obtain all the terms in the third-order perturbation using perfect-foresight sequences.

Proposition 4 (Third-order certainty correspondence)**.** *The third-order perturbation of X_t in the sequence space is given by the cubic MA(∞) in equation (19). The coefficients on the new terms are given by equations (20), (21), (23), (24), and (25), so they can be obtained from perfect-foresight impulse responses.*

Again, the argument in Lombardo and Uhlig (2018) can be used to show that the third order perturbation obtained here is equivalent to the third-order pruned state space solution. As for the second-order case, in appendix A.9 we use this result to test that our routines are correct by showing that they deliver the same solution as the pruned third-order state space perturbation on the neoclassical growth model.

Application to heterogeneous-agent models. We continue to illustrate these results using our smooth SIM model in PE and GE. Figure 1 shows that the fit is now almost perfect compared to a third order solution: the dash-dot green line (which shows the terms $\mathcal{X}_j + \frac{1}{2}\mathcal{X}_{jj}\nu + \frac{1}{6}\mathcal{X}_{jjj}\nu^2$ for $\nu = 10\%$ of quarterly GDP), is nearly identical to the solid black line in either model. Figure 2 shows that this continues to be the case for shocks of up to 10% of quarterly GDP in the PE model, and Figure A.2 shows the same for the GE model. Hence, compared to the nonlinear perfect foresight solution, a third order perturbation works extremely well in our context: it not only captures the basic concavity/convexity patterns of nonlinear iMPCs, but also corrects for the changes in this curvature in a way that extrapolates well for fiscal shocks of a plausible size.

The right panel of figure 4 visualizes the new terms that come from the third order perturbation. First, the purple line shows the $\mathcal{X}_{\sigma\sigma t}$ term, reflecting the overall interaction of risk with the impulse response. This term is negative on impact and positive later on, notably because of pre-



(a) Second order

(b) Third order

Figure 4: Decomposing second and third order impulse on consumption (PE model)

cautionary savings at the risky steady state, which lowers MPCs on impact and raise them later on. Next, the blue line shows the overall size effects from a combination of three shocks, and the orange line shows the mixed size-history effects from sequences of large and small shocks. These terms have the opposite shape as the risk term. Finally, the green line visualizes the interaction term from a sequence of three shocks, which can be interpreted as the additional effect on the impulse response of consumption from the presence of two shocks at periods $t = -2$ and $t = -1$, relative to the effect of either in isolation. This effect is quite small (compare the magnitudes on the right and the left panel of figure 4), reflecting fast decay in history terms that is suggestive that a fourth order approximation will likely not add much additional precision.

2.6 Generalized impulse response functions and model moments.

Generalized impulse response functions. A common way to summarize the behavior of the model is in terms of its impulse responses. We define the generalized impulse response of X as the revision of the forecast of X, j periods ahead due to an innovation to ϵ at date t, $IRF_j^X \equiv \mathbb{E}_t[X_{t+j}] - \mathbb{E}_{t-1}[X_{t+j}]$. This revision generally depends on the size of perceived shocks σ, on the size of the date-t innovation ν, and on the state entering period t, though it does not further depend on the calendar time t beyond these states. If we consider a state-space solution (3), then the state entering date t is X_{t-1},

$$IRF_j\left(\nu, X_{t-1}, \sigma\right) \equiv \mathbb{E}\left[X_{t+j} | \epsilon_t = \nu, X_{t-1}, \sigma\right] - \mathbb{E}\left[X_{t+j} | X_{t-1}, \sigma\right] \quad j \geq 0 \tag{26}$$

while if we consider a sequence-space solution (2), then the state is $\{\epsilon_{t-k}\}_{k=1}^{\infty}$,

$$IRF_j\left(\nu, \{\epsilon_{t-k}\}_{k=1}^{\infty}, \sigma\right) \equiv \mathbb{E}\left[X_{t+j} | \epsilon_t = \nu, \{\epsilon_{t-k}\}_{k=1}^{\infty}\right] - \mathbb{E}\left[X_{t+j} | \{\epsilon_{t-k}\}_{k=1}^{\infty}\right] \quad j \geq 0 \tag{27}$$

Proposition 5 (GIRFs). *Given history summarized by past shocks $\{\epsilon_{t-l}\}_{l=1}^{\infty}$, the second and third order*

GIRFs are given, respectively, by

$$IRF_j^{(2)}(\nu,\sigma) = \mathcal{X}_j\nu + \frac{1}{2}\mathcal{X}_{jj}\left(\nu^2-\sigma^2\right) + \sum_{k>j}^{\infty}\mathcal{X}_{jk}\nu\epsilon_{t-(k-j)} \tag{28}$$

$$\begin{aligned} IRF_j^{(3)}(\nu,\sigma) &= IRF_j^{(2)}(\nu,\sigma) + \frac{1}{2}\left(\mathcal{X}_{\sigma\sigma j} + \sum_{0\le i<j}\mathcal{X}_{iij}\right)\sigma^2\nu + \frac{1}{6}\mathcal{X}_{jjj}\nu^3 + \frac{1}{2}\left(\sum_{k>j}^{\infty}\mathcal{X}_{jjk}\epsilon_{t-(k-j)}\right)(\nu^2-\sigma^2) \\ &+ \frac{1}{2}\sum_{k>j}^{\infty}\mathcal{X}_{jkk}\epsilon^2_{t-(k-j)}\nu + \sum_{k>j}^{\infty}\sum_{k'>k}^{\infty}\mathcal{X}_{jkk'}\epsilon_{t-(k-j)}\epsilon_{t-(k'-j)}\nu. \end{aligned} \tag{29}$$

In addition to the linear term from the first order impulse response, the generalized impulse response is nonlinear in ν due to the quadratic term in (28), as well as state-dependent. The impulse response at date j is concave in size ν if $\mathcal{X}_{jj} < 0$ and convex otherwise; when past innovations are 0, it coincides with the linear impulse response when $\nu = \pm\sigma$, with this dependence on σ reflecting expected size dependence.[15] Further, a past innovation k periods ago changes the impulse response around 0 to a new shock by an amount proportional to χ_{jk}.

The third-order GIRF similarly contains additional terms from the third order perturbation. For instance, the impulse response around 0 is modified by an amount that scales in risk σ^2 due to both time-variation in precautionary savings and mixed size-history dependence.

Moments The ergodic moments of the model are an object of broad interest. A typical approach in the literature is to use the moments from the perturbation approximation—for instance, second moments calculated from the $MA(\infty)$ implied by the first-order approximation are routinely used to estimate models via method of moments or likelihood methods (e.g. Auclert et al. 2021). Here, we ask how these calculations relate to the perturbation of the primitive ergodic model moments as a function of σ, and show that we can calculate fifth-order accurate moments using the third order approximation derived above.

In particular, for vector-valued X_t, we are interested in the autocovariance matrix

$$\Sigma_h(\sigma) = \mathrm{Cov}[\mathcal{X}(\sigma,\sigma\bar{\epsilon}_t,\sigma\bar{\epsilon}_{t-1},\ldots),\mathcal{X}(\sigma,\sigma\bar{\epsilon}_{t-h},\sigma\bar{\epsilon}_{t-h-1},\ldots)] \tag{30}$$

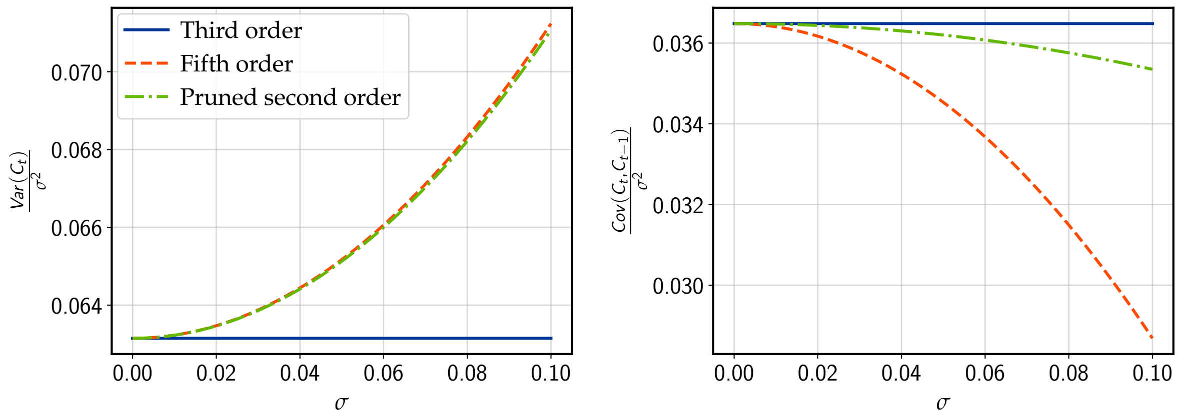
We then have the following proposition, which further assumes that $\bar{\epsilon}$ has the same moments, up to at least fourth order, as a standard normal distribution.[16]

Proposition 6 (Ergodic moments)**.** *A fifth-order accurate solution to the autocovariance matrix in (30) is given by*

$$\Sigma_h(\sigma) = \sigma^2\Sigma_{h,2} + \sigma^4\Sigma_{h,4} + O(\sigma^6)$$

[15] This result was previously noted by Andreasen et al. (2018).

[16] This additional assumption is only required for the fourth-order term $\sigma^4\Sigma_{h,4}$. Our original assumption of symmetric $\bar{\epsilon}$ is sufficient for the $\sigma^2\Sigma_{h,2}$ to be accurate to third order.



(a) Consumption variance

(b) Consumption first autocovariance

Figure 5: Ergodic moments: higher order approximations (PE model)

where $\Sigma_{h,2} = \sum_{i=0}^{\infty} \mathcal{X}_{h+i}\mathcal{X}_i^\top$ *is the autocovariance matrix obtained from the first-order perturbation solution, and*

$$\Sigma_{h,4} = \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \mathcal{X}_{h+i,h+j}\mathcal{X}_{ij}^\top + \frac{1}{2}\sum_{i=0}^{\infty}\left(\mathcal{X}_{h+i}\mathcal{X}_{\sigma\sigma i}^\top + \mathcal{X}_{\sigma\sigma,h+i}\mathcal{X}_i^\top\right) + \frac{1}{2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\left(\mathcal{X}_{h+i}\mathcal{X}_{ijj}^\top + \mathcal{X}_{h+i,jj}\mathcal{X}_i^\top\right).$$

Proposition 6 shows, first, that the second-moments calculated from the first-order solution are accurate up to third order in σ. This provides a justification for the typical procedure in the literature. The proposition further offers a correction to the moments that is accurate up to fifth order. This correction involves the second and third derivatives of $\mathcal{X}$, which we have shown how to obtain in sections 2.4 and 2.5 respectively. Notably, however, it does not require *all terms* from the second and third-order solutions: for instance, the triple interactions $\mathcal{X}_{ijk}$, which are the most numerous and therefore computationally intensive part of the third-order solution, do not appear.[17]

Figure 5 illustrates these results by visualizing the ergodic variance of consumption $\text{Var}(C_t)$ and the first order autocovariance of consumption $\text{Cov}(C_t, C_{t-1})$, normalized by income shock variance σ^2, for different values of σ. The horizontal blue line shows the second-order approximation to variance—the usual term obtained in the literature, including from typical Dynare output—which is constant once normalized by σ^2. The orange line shows how this term is modified in a fifth-order accurate approximation of the moments. For large aggregate shocks ($\sigma \approx 0.10$), the scaled ergodic variance is around 15% larger, and the ergodic autocovariance around 20% lower. The figure also shows, in dash-dot green, the moments that are obtained from the pruned

[17]This is in contrast to the more common approach in, for instance, Andreasen et al. (2018), where the entire second- or third-order pruned perturbation solution is used when approximating autocovariances. Our derivation shows that other components of the solution are not needed for fifth-order accuracy—and while they might be needed to higher order, an even higher-order approximation to moments would involve terms from the fourth- and fifth-order solutions.

second-order solution: in the expression for $\Sigma_{h,4}$ in proposition 6, that solution corresponds to considering the first term only. In our example, the pruned second-order moments do a very good job at approximating the consumption variance to fifth order, but do a poor job at approximating its autocovariance.

In conclusion, using fifth order accurate moments can make a significant difference in certain applications, even compared to a pruned second-order solution, and these moments are obtainable by computing only certain parts of the third-order sequence-space solution, making it especially computationally efficient to do so.

2.7 Comparison to global solution

A key benefit of our partial equilibrium model is that it is possible to solve it globally using a Bellman equation approach. Indeed, the model in (7) can also be written as:

$$\begin{aligned} V(a,e,\beta,Z) \quad &= \quad \mathbb{E}_\xi\left[\max_{c,a'} u(c) + \beta\mathbb{E}_\epsilon[V(a',e',\beta',Z')|e,Z]\right] \\ \text{s.t.} \quad & \quad a' + c = (1+r)a + Z\xi e + Tr; \qquad a' \geq \underline{a} \\ & \quad Z' - Z_{ss} = \rho_Z(Z - Z_{ss}) + \epsilon \end{aligned} \tag{31}$$

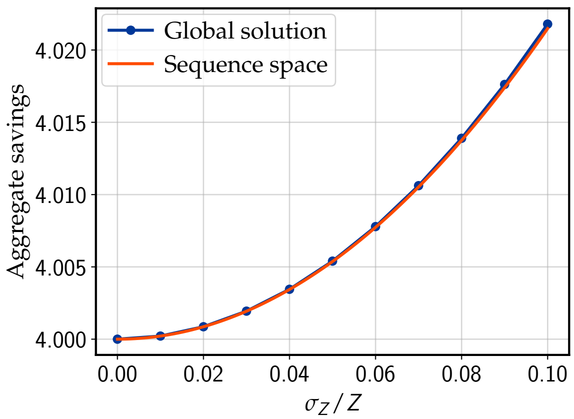
and solving this Bellman equation directly delivers the state space solution $X_t = g(X_{t-1}, \epsilon_t)$. By contrast, our general equilibrium model cannot be cast in this form: as Yang et al. (2025) point out, the general equilibrium market-clearing level of output Y_t (and therefore of post-tax income Z_t) does not have a Markovian structure, so a similar strategy is not available. Our sequence-space perturbation method provides a feasible approach to solve the model up to third order; and section 3 provides additional tools to speed up the general solution specifically.

We implement the global solution as follows. Since Z_t is iid in our main calibration, $\rho_Z = 0$, it does not even need to enter as a state variable in the Bellman equation (31). We assume a normal distribution for ϵ with standard deviation σ and compute expectations with respect to ϵ using Gauss-Hermite quadrature with 9 nodes. We then compare our resulting global solution to those from the third order perturbation.

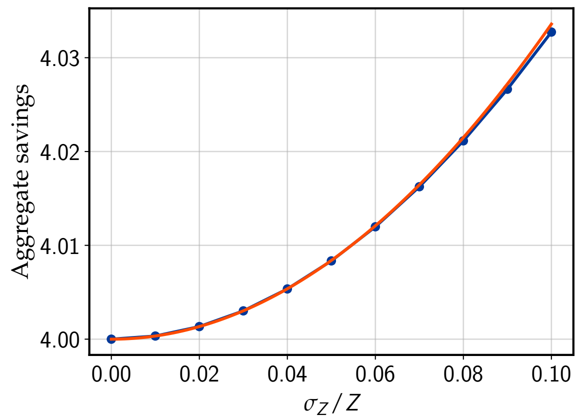
Figure 6 provides this comparison. We first consider how aggregate savings vary with σ in both the risky steady state (panel a) and the model ergodic mean (panel b), with panel c reporting the risky steady state distribution.[18] These terms, which can all be obtained from the second-order perturbation, and figure 6 shows that they are extremely close to the global solution. We also consider a notion of impulse response, where in the global model we start from the risky steady state and then consider the effect of the sequence of shocks $(\epsilon_0, \epsilon_1, \epsilon_2, \ldots) = (1, 0, 0, \ldots)$. Panel d shows this impulse response.[19] Here, we see that the third order solution captures most

[18] We obtain the risky steady state in the global solution by considering the idiosyncratic steady state that results when all agents follow the policy $c(a, e, \beta, Z = 1)$.

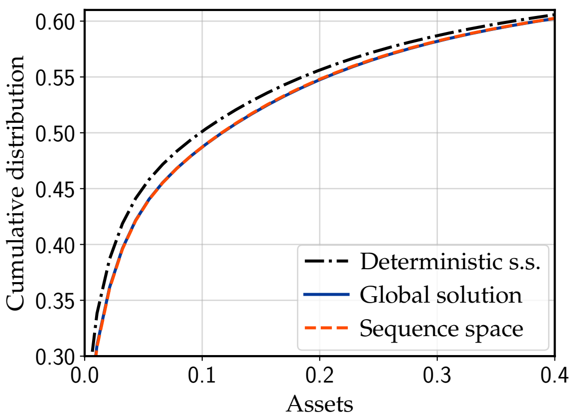
[19] This differs from the usual impulse response concept because it does not trace out the expected outcome on average,



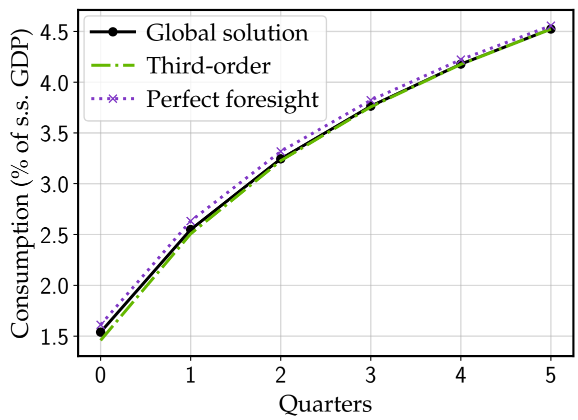
(a) Risky steady state



(b) Ergodic mean



(c) Risky steady state



(d) Cumulative impulse response

Figure 6: Testing global solution vs 3rd order sequence space

of the feature of the global solution, though it undershoots slightly on impact. By contrast, the perfect foresight overshoots slightly the global solution on impact. This is natural, since the perfect foresight does not incorporate the reduced MPCs from higher precautionary savings visible in the risk effect in the purple line of figure 4b.

In summary, the third order perturbation and the global solution get extremely close up to the third order in our PE model, which is comforting. An important question remains to determine how much the global solution adds to the third order perturbation in more complicated, general equilibrium models, for which no standard global method exists.

2.8 How large are nonlinearities?

We have shown that linearization works well: it approximates the perfect-foresight solution accurately to third (and sometime even lower) order, and captures key features of the global solution for a model in which we can compute that solution. This provides a useful path to solving heterogeneous-agent models in which aggregate risk matters more broadly.

In a related context, however, Bianchi and Kaplan (2026) have instead argued that linearization works poorly, including up to third order, in comparison to the perfect foresight solution. In appendix C, we investigate their arguments.

First, as we show in appendix B, it is important to use a smooth model to use second- or higher-order perturbation in the first place. Our particular smooth model, with continuously distributed income, serves this purpose, while the model Bianchi and Kaplan (2026) use is not smooth: its distribution is a set of mass points, rather than a continuous and differentiable density as in ours.[20]

Second, when doing the analysis on this smooth model, we show in appendix C that approximations of increasing order capture increasingly well the nonlinear perfect foresight solution, with third-order perturbation providing an excellent approximation for fiscal shocks of up to 10% of quarterly GDP. Significant errors in the third order approximation relative to the nonlinear perfect foresight solution only appear when have a combination of 1) lump-sum rather than proportional transfers, 2) transitory rather than persistent transfers, 3) looking at the impact effect, rather than cumulative effect over a few quarters, and 4) shocks that outside of a plausible range for business-cycle fluctuations.

3 Implementation in the sequence space

Section 2 took as given a researcher's ability to solve for up to three consecutive nonlinear MIT shocks $X_j^{MIT,l,l'}(\cdot)$ in equations (9), (14), and (22), and to take derivatives of these functions. We showed how to obtain the complete third-order perturbation solution in the sequence space from these derivatives.

but is more easily obtained than a generalized impulse response using the global solution.

[20] Bayer et al. (2026a) make a similar point, and propose an alternative solution based on the method in Bayer et al. (2026b).

In practice, while well-known algorithms do exist to obtain $X_j^{MIT,l,l'}(\cdot)$ in partial equilibrium settings, for general equilibrium models these algorithms require nonlinear fixed point algorithms which can be slow and are not guaranteed to converge. We begin by reviewing the former case, and for the latter case, we establish a result that allows one to bypass the nonlinear fixed point. With this result at hand, we discuss efficiency and speed consideration for our sequence-space algorithm relative to existing state-space perturbation approaches.

3.1 Computing perfect-foresight impulses

The perfect-foresight functions (9), (14), and (22) can be constructed with knowledge of the functions $g_k^*(X_{t-1}, \epsilon_t)$ that, given a value X_{t-1} and ϵ_t for the state and the shock at time t, deliver the path X_{t+k} of the economy at all $t+k$ when $\sigma = 0$ and all future shocks realize at $\epsilon_{t+k} = 0$. Indeed, writing X_{ss} for the deterministic steady state of the model, we have that:

$$X_j^{MIT}(\nu) = g_j^*(X_{ss}, \nu); \quad X_j^{MIT,l}(\epsilon, \nu) = g_j^*(X_l^{MIT}(\epsilon), \nu); \quad X_j^{MIT,l,l'}(\epsilon', \epsilon, \nu) = g_j^*(X_l^{MIT,l'}(\epsilon), \nu) \tag{32}$$

Hence, with knowledge of the g_k^* functions, we can use a differentiation method, numerical or analytical, to obtain the derivatives of (9), (14), and (22) required for the sequence-space perturbation.[21] For instance, if we knew the state-space solution $g(X_{t-1}, \epsilon_t, \sigma)$, then we could construct $g_0^*(X_{t-1}, \epsilon_t) = g(X_{t-1}, \epsilon_t, 0)$ and $g_k^* = g^{k-1}(g(X_{t-1}, \epsilon_t, 0), 0, 0)$, but of course we do not presuppose this knowledge.

Partial equilibrium settings. Sometimes, the F function is such that there is a natural algorithm to directly obtain the g_k^* functions. This is typical of "partial equilibrium" settings in heterogeneous-agent macro. For instance, in our PE model in equation (6), given current ϵ_t and knowing that future $\epsilon_{t+k} = 0$, we can construct the path of future income $Z_{t+k} = 1 + \rho_Z^{k+1}(Z_{t-1} - 1) + \rho_Z^k \epsilon_t$; with this path we can then solve all agent policies c_{t+k} starting from the terminal steady state and all distributions Φ_{t+k} starting from the initial distribution Φ_{t-1}, and finally form consumption by taking a cross-sectional average of the policies with respect to the distribution; so we have all the elements in the deterministic sequence $\{X_t, X_{t+1}, \ldots\}$.

3.2 General equilibrium settings

Often, the F function is more complicated and the functions g_k^* cannot be obtained so simply. This corresponds, typically, to "general equilibrium" situations in which agents need to know endogenous future paths to determine their policies, and these paths in turn are determined to obey market clearing at every date. In these settings, it is often possible to break down the problem as follows.

[21] In our implementation, we use numerical differentiation up to third order, since it is easy to implement and appears to work well in our context, but speed gains could be obtained by exploiting analytical differentiation.

Suppose first that there exists a low dimensional set of aggregates $\mathbf{U} = (\mathbf{U}_t, \mathbf{U}_{t+1}, \dots)$ such that, given $\epsilon_t = \nu$ and all future $\epsilon_{t+k} = 0$, we can solve separately for $\mathbf{U}$ as the solution to some nonlinear system

$$\mathbf{H}(\mathbf{U}(\nu), \nu) = 0 \tag{33}$$

(see, for instance, Auclert et al. (2021), equation 29). Next, suppose that, given knowledge of $\mathbf{U}$, ϵ_t and X_{t-1}, we can form $X_{t+k} = g_k^*(\mathbf{U}, X_{t-1}, \epsilon_t)$. In this case, we show that there is a simple algorithm to solve for the second- and third-order derivatives of the MIT shock without the need to solve a general equilibrium fixed point. Note that it is sufficient to get $\mathbf{U}''(0)$ for the second-order solution and $\mathbf{U}'''(0)$ for the third order solution.

One-shot principle algorithm to get $U''(0)$ and $U'''(0)$. To see how this works, first differentiate (33) with respect to ν to obtain

$$\mathbf{H}_U(\mathbf{U}(\nu), \nu)\mathbf{U}'(\nu) + \mathbf{H}_\nu(\mathbf{U}(\nu), \nu) = 0 \tag{34}$$

Evaluating (34) at $\nu = 0$, letting $\mathbf{H}_U$ and $\mathbf{H}_\nu$ denote derivatives around the steady state gives $\mathbf{H}_U\mathbf{U}'(0) + \mathbf{H}_\nu = 0$. This can be solved to obtain the usual first-order solution

$$\mathbf{U}'(0) = -\mathbf{H}_U^{-1}\mathbf{H}_\nu$$

Calculating this solution only requires the sequence-space Jacobians $\mathbf{H}_U$ and $\mathbf{H}_\nu$, which can be obtained rapidly via the methods in Auclert et al. (2021). This delivers the first-order MIT shock solution in (10).

Next, differentiating (34) again with respect to ν, evaluating at $\nu = 0$, now gives

$$\mathbf{H}_U\mathbf{U}''(0) + \mathbf{H}_{UU}(\mathbf{U}'(0))^2 + 2\mathbf{H}_{U\nu}\mathbf{U}'(0) + \mathbf{H}_{\nu\nu} = 0 \tag{35}$$

where, e.g., $\mathbf{H}_{UU}(\mathbf{U}'(0))^2$ denotes the application of the vector-valued Hessian $\mathbf{H}_{UU}$ to $\mathbf{U}'(0)$ on both input dimensions. (35) can be solved to obtain the second-order term in a solution for $\mathbf{U}$:

$$\mathbf{U}''(0) = -\mathbf{H}_U^{-1}\left(\mathbf{H}_{UU}(\mathbf{U}'(0))^2 + 2\mathbf{H}_{U\nu}\mathbf{U}'(0) + \mathbf{H}_{\nu\nu}\right) \tag{36}$$

This equation for $\mathbf{U}''(0)$ has a simple interpretation. First, calculate the first-order impulse response $\mathbf{U}'(0)$, which satisfies $\mathbf{H}$ to first order since $\mathbf{H}_U\mathbf{U}'(0) + \mathbf{H}_\nu = 0$. Now, evaluate $\mathbf{H}$ to *second order* on these first-order impulse responses, obtaining the term in parentheses in (36). This will generally not be zero, implying a second-order error in $\mathbf{H}$ that must be offset by some $\mathbf{U}''(0)$, which we obtain by applying $-\mathbf{H}_U^{-1}$. Note, importantly, that obtaining $\mathbf{U}''(0)$ requires a single

	Sequence-space		State-space	
Perturbation	Key object	Size	Key object	Size
First order	First-order impulse response	T	First-order LoM	n^2
Second-order	Precautionary effect	1	Second-order LoM	n^3
	Size dependence impulse	T		
	History dependence impulse	TL		
	Total	$\sim TL$	*Total*	n^3
Third-order	Precautionary impulse	T	Third-order LoM	n^4
	Mixed size-history impulses	$2T \times L$		
	History interaction impulses	$T \times L^2$		
	Total for 5th-order moments	$\sim 3TL$		
	Total for full solution	$\sim TL^2$	*Total*	n^4

Table 2: Dimensionality of sequence-space vs state-space perturbation

nonlinear evaluation of the $\mathbf{H}$ function: concretely, we can first obtain the vector

$$\mathbf{h}^{err,2} = \lim_{\nu \to 0} \frac{\mathbf{H}\left(U_{ss} + \mathbf{U}'(0)\nu, \nu\right)}{\nu^2/2}$$

by evaluating $\mathbf{H}$ nonlinearly, and then form $\mathbf{U}''(0) = -\mathbf{H}_U^{-1}\mathbf{h}^{err,2}$ by applying the already-computed inverse steady state sequence-space Jacobian on $-\mathbf{h}^{err,2}$.

Equation (36) therefore delivers the terms necessary for the second-order sequence space solution in (13) and (15). Observe, critically, that this bypasses the need for solving a nonlinear fixed point, since it requires a single nonlinear evaluation of equilibrium conditions, and then applying the inverse Jacobian $\mathbf{H}_U^{-1}$ to re-equilibrate the system.

Continuing this argument, appendix A.12 shows that this "one-shot principle" can be applied to obtain the third order $\mathbf{U}'''(0)$ necessary for equations (20)-(23) as well: evaluate the $\mathbf{H}$ function on the second-order solution to find the third-order error $\mathbf{h}^{err,3}$, and apply $\mathbf{H}_U^{-1}$ to find $\mathbf{U}'''(0)$. In this way, we can dramatically cut down on the time required to solve for the overall third-order perturbation.

3.3 Dimension of the solution

Table 2 shows the dimension of the sequence-space solution, understood as the number of terms to keep track of to obtain one element of $\mathcal{X}$ (such as an aggregate of interest, say output), compared to that of the state-space solution. In the sequence space, impulse responses are truncated at horizon T and history effects have to be kept track up to L past lags. These are choices for an algorithm that will be application specific (T and L must be large enough that the corresponding terms have reverted to 0 at these points), but in our heterogeneous-agent applications with quarterly calibration we have found that $T = 300$ and $L = 60$ works well for second-order effects, and even $L = 30$ for third order effects. In addition, our overall idiosyncratic state-space is $n \simeq 5000$, since we are using 200 grid points for assets, 7 discretized states for productivity and 2 for patience, and that the state space keeps track of both marginal value and distribution at each point on the grid.

Given these values, a second-order perturbation requires calculating approximately 20,000 terms, and a third-order perturbation approximately 300,000 terms. By contrast, a first-order perturbation in the state space already has size 25 million, a second-order perturbation is 125 billion, and a third-order perturbation has size 62 trillion. Concretely, this suggests that calculating a third order perturbation using a state-space approach is infeasible because the resulting system could not even be stored on a computer. Third-order perturbation in the sequence-space is much more economical and practical.

4 Applications

4.1 Application 1: quantitative HANK model

In this section, we relax some of the restrictions from our simple general equilibrium heterogeneous-agent model to study a quantitative HANK model. The model contains the key ingredients required for quantitative evaluation of monetary business cycle: a Taylor rule for the nominal interest rate, and a Phillips curve relating inflation to economic activity. This brings about new sources of nonlinearities.

In this updated model, the central bank follows a Taylor rule for the nominal interest rate of the form

$$1 + i_t = (1 + r)\,(1 + \pi_t)^{\phi} \tag{37}$$

We maintain our assumption that households are only able to invest in real bonds. However, the model also features risk-neutral arbitrageurs, who have the the possibility of investing in either nominal or real bonds. In equilibrium these arbitrageurs make no profits, and their portfolio choice optimality condition implies that the real interest rate promised at t on real bonds paying at $t+1$ is

$$1 + r_t = \mathbb{E}_t\left[\frac{1 + i_t}{1 + \pi_{t+1}}\right] \tag{38}$$

The Taylor rule implies that inflation is now relevant for the determination of aggregate demand. We assume Rotemberg wage setting, as in Auclert et al. (2024a), and further assume that unions evaluate welfare using aggregate consumption, the average discount factor across households $\bar{\beta} = (\beta_L + \beta_H)/2$, and the steady state level of taxes.[22] Motivated by the widespread empirical observation that prices are more flexible than wages, we continue to assume flexible prices, so that the real wage is constant at 1. This combination of assumptions leads to the nonlinear wage Phillips curve:

[22]This assumption avoids large time-varying effects from distortionary taxation.

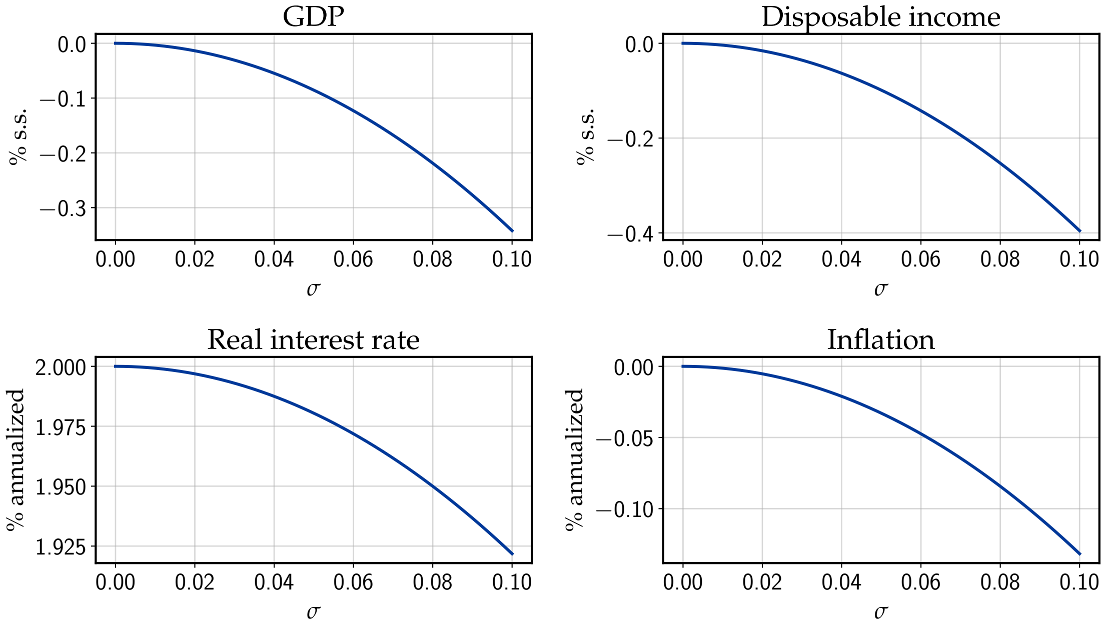


Figure 7: Risky steady state in the quantitative HANK model

$$\pi_t \left(1+\pi_t\right) = \kappa \left(v'\left(N_t\right) - \frac{\epsilon - 1}{\epsilon} \left(1-\tau\right) u'\left(C_t\right) \right) + \bar{\beta} \mathbb{E}_t \left[\pi_{t+1} \left(1+\pi_{t+1}\right)\right] \tag{39}$$

Compared to our simple IKC model, this model contains new sources of nonlinearities brought about by the presence of inflation in the denominator in (38) and entering as a square in (39). We solve the model using the method described in section 3.2, with unknowns $U_t = (Y_t, \pi_t, r_t, Z_t)$. We keep track of $\frac{1}{1+\pi_t}$ and $\pi_t(1+\pi_t)$ as separate variables to that we can calculate the $\mathcal{X}_{\sigma\sigma}$ term. We consider a standard calibration with $\phi = 1.5$, $\kappa = 0.01$, a an isoelastic disutility of labor with constant Frisch elasticity of 0.5 ($v'(N) = N^2$), and a steady state markup $\frac{\epsilon-1}{\epsilon} = 1.1$.

We continue to assume shocks to government bonds with persistence $\rho = 0.95$ and standard deviation σ that finance proportional taxes, as in our IKC model calibration.

Risky steady state. Figure 7 visualizes the risky steady state in our model as a function of the standard deviation of bond shocks σ. Contrary to the PE model where higher σ raised aggregate savings in the risky steady state, here the equilibrium effect is a mix of depressed real interest rate and depressed output and inflation. Contrary to the IKC general equilibrium version of that model in appendix A.11, however, the risky steady state adjustment falls only partly on output, making the effect much less pronounced. For our baseline $\sigma = 5\%$ calibration, output is depressed by less than 0.1%, and interest rates by about 25 basis points, in the risky steady state relative to the deterministic steady state.

Generalized impulse responses. Figure 8 present the impact ($t = 0$) and delayed ($t = 1$ effect) of a deficit-financed bond shock on consumption in our quantitative HANK model as a function of the size of the shock ν, presented as in figure 2 so that it is directly comparable to the perfect foresight solution.[23] The concave/convex pattern present in both the PE and our simple GE model survive the introduction of additional quantitative features, with the third order solution continuing to be an excellent approximation. Adding the anticipated aggregate risk term to the solution at our baseline σ contributes only a modest amount to the third order solution.

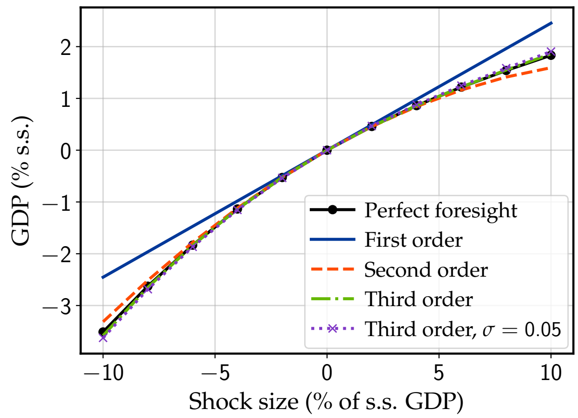
The bottom panel presents the effects on inflation in both periods. Here, we find that nonlinearities are quite modest at second and third order, even though there are multiple sources of such nonlinearities. One reason is that, because inflation is a forward looking average of future aggregate demand, the the concavity in aggregate demand in the first period gets averaged out with the convexity in the latter period (see appendix C for a further discussion of this point). This effect is particularly pronounced at $t = 0$, and goes away at $t = 1$ since the main source of concavity goes away, with convexity being the dominant effect, as expected. The nonlinearities due to the presence of nonlinear terms in inflation in the expectations in both the Fisher equation (38) and the Phillips curve (39) do not appear to play a major role.

Moments. Next, we reevaluate how the variance and the autocovariance of output affected by the computation of a fifth-order accurate solution. Just like in the partial equilibrium case and the simple general equilibrium model, figure 9 finds that the fifth-order accurate correction increases output (and consumption) variance and reduces output (and consumption) autocovariance. However, contrary to the partial equilibrium case in figure 5, the moments implied by the pruned second-order state-space solution now are a poor approximation even for the variance, and this effect is quite large even in our benchmark $\sigma = 0.05$ calibration. This suggests that, for quantitative HANK models to accurately capture moments like output variance, obtaining the relevant parts of a third order solution may be necessary.

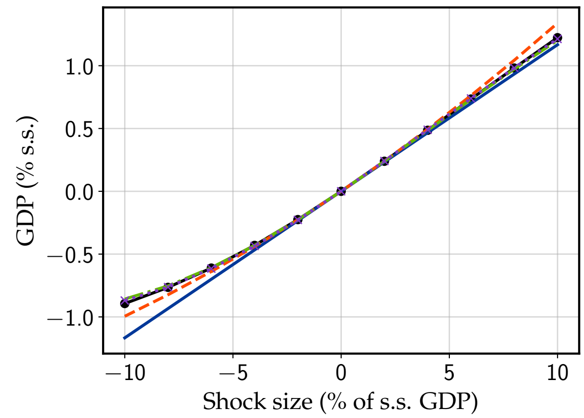
Nonlinear multipliers. To conclude, we study how much history dependence there is in the transfer multiplier dY_t/dB_t. We simulate the model twice, once using the second-order Volterra expansion and once using the third-order Volterra expansion, and calculate at any point in the simulation the first-order responsiveness of output to an innovation to B_t using our formulas for (11) and (19).

Figure 10 finds that the transfer multiplier is about four times as large when a sequence of negative shocks has pushed output 3% below the deterministic steady state, compared to when a sequence of positive shocks has pushed output 2% above that steady state. The model therefore features a wide ergodic distribution of fiscal multipliers. This is consistent with a typical result in the empirical literature on state-dependent fiscal multipliers (eg Auerbach and Gorodnichenko 2012), though here it is unrelated to the zero lower bound (Ramey and Zubairy 2018), and is instead driven by the association between recessions and weak balance sheets, and therefore higher

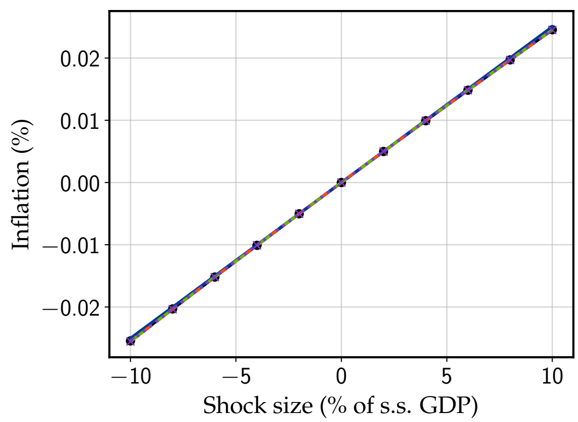
[23] As proposition 5 shows, the generalized impulse responses would look slightly different due to the expected size dependence.



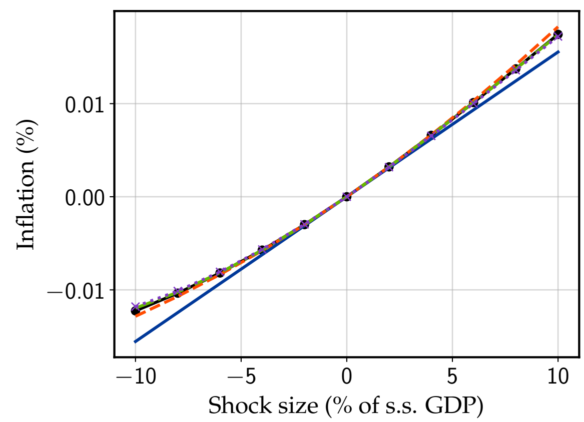
(a) Output at $t = 0$



(b) Output at $t = 1$

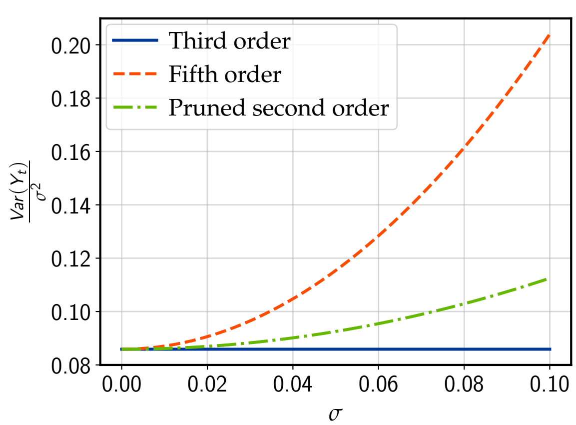


(c) Inflation at $t = 0$

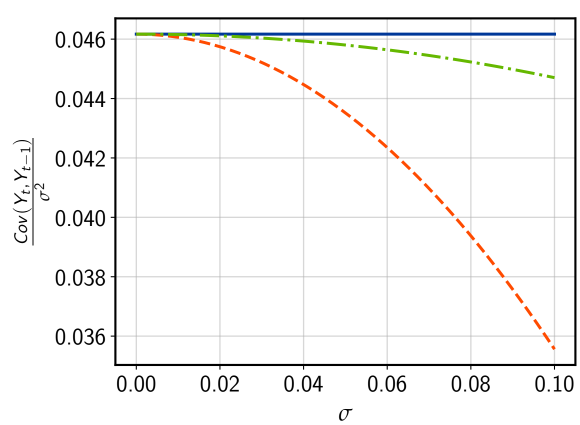


(d) Inflation at $t = 1$

Figure 8: Response to a fiscal shock in the quantitative HANK model



(a) Output variance



(b) Output first autocovariance

Figure 9: Higher order approximation of moments in the quantitative HANK model

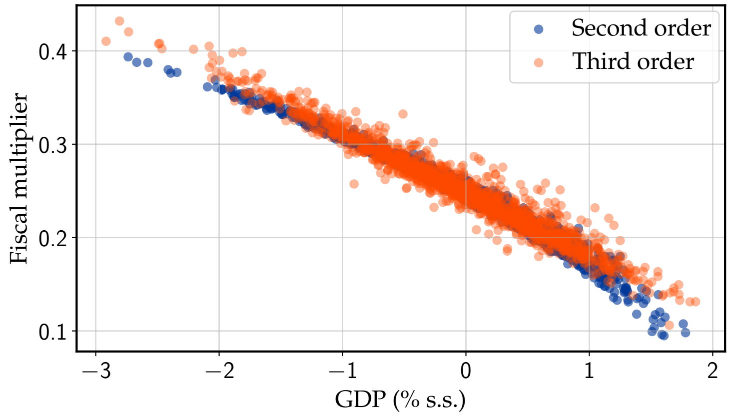


Figure 10: Time-varying fiscal multipliers for $\sigma = 0.025$ in the quantitative HANK model

MPCs. More broadly, just like models solved to first order make contact with the literature estimating impulse responses, models solved to the second and higher order make contact with the literature on state and size dependence in these impulse responses. We view this as a useful prospect for the literature going forward.

4.2 Application 2: nonlinear Phillips curves

Our second general application is a menu cost model with idiosyncratic shocks in the spirit of Golosov and Lucas (2007) and Alvarez, Lippi and Souganidis (2023). The model features a random menu cost, trend inflation, and strategic complementarity, as in Auclert, Rigato, Rognlie and Straub (2024c) and Alvarez et al. (2023). Firms have a quadratic objective, they try to keep their price gap x as close as possible to aggregate log nominal marginal cost MC_t. They can reset their price either by paying a fixed menu cost ζ, or by waiting for a free opportunity to adjust their cost, which arrives with Calvo probability λ. The only state is the price gap x, and their Bellman equation at date t is given by:

$$V_t(x) = \min \frac{1}{2}(x - MC_t)^2 + \beta(1-\lambda)\mathbb{E}_\epsilon[\min\{V_{t+1}(x+\epsilon), \zeta + \min_{x^*} V_{t+1}(x^*)\}] + \beta\lambda \min_{x^*} V_{t+1}(x^*)$$

We assume linear aggregation, so that the log price index is $P_t = \int x dD_t(x)$, where D_t is the distribution of gaps. Finally, we model strategic complementarity as roundabout production, with firms producing from an aggregate of labor and an input bundle made up of production from other firms; this implies that log nominal marginal cost is

$$MC_t = \chi W_t + (1-\chi)P_t$$

The log nominal wage trends at rate μ, $W_t = \mu t + \tilde{W}_t$, where $\tilde{W}_t = \rho\tilde{W}_{t-1} + \sigma\bar{\epsilon}_t$ follows an AR(1) process and is the source of aggregate risk in the model. We calibrate the model at a quarterly frequency so that, in the deterministic steady state, the price adjustment frequency is freq $= 0.25$,

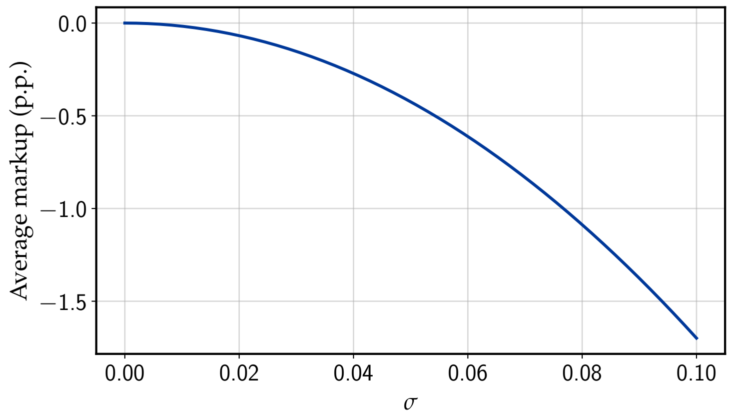


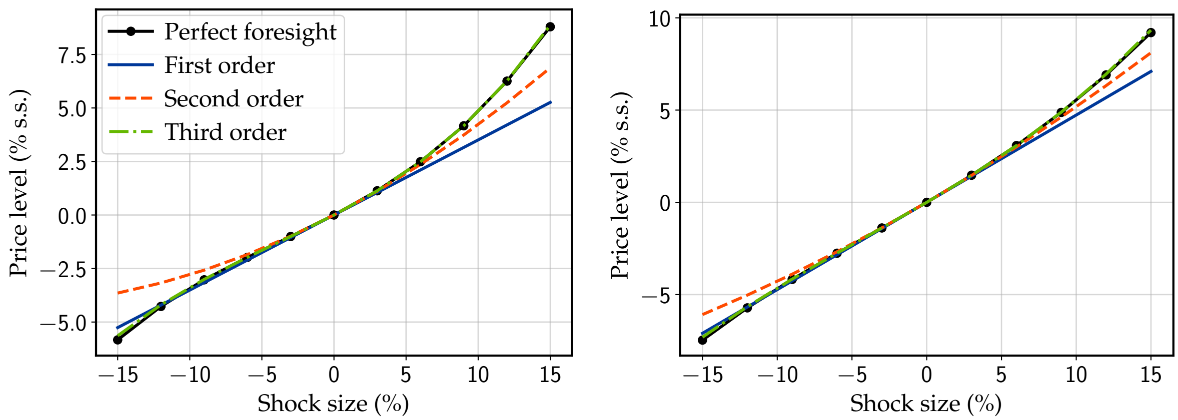
Figure 11: Average markup in the risky steady state

the average size of price changes is $\mathbb{E}[\Delta p] = 0.1$, the share of free adjustments in total adjustments if 0.75, strategic complementarity is $1 - \chi = 0.3$, trend inflation is $\mu = 5\%$ annually, and the shock persistence is $\rho = 0.8$. General equilibrium requires solving for the nonlinear fixed point $MC_t = \chi W_t + (1 - \chi) P_t(\{MC_s\})$. We therefore employ the one-shot principle method discussed in section 3.2 to solve for the unknown MC_t. Further details on the methods used to solve the steady state of the model are provided in Auclert et al. (2024c)

Risky steady state. We first consider effect on the risky steady state average markup. Figure 11 shows that cost uncertainty reduces risky steady-state markups. This effect is due to a widening of the inaction region because of the presence of cost uncertainty, which in turn erodes average real prices. This is related to the precautionary pricing mechanism highlighted by Baley and Blanco (2019), but here the uncertainty that of general equilibrium aggregate costs.

Impulse responses. Next, we turn to the impulse response of the price level to cost shocks of different sizes. As is well known of menu cost models, cost shocks tend to have nonlinear effects on prices, as larger shocks in either direction trigger more repricing. Figure 12 shows this effect: there is convexity in the immediate response of the price level to shocks. Contrary to the case of income shocks on consumption, here the effects of shocks persistent, so the delayed response is convex in the size of the shock as well.

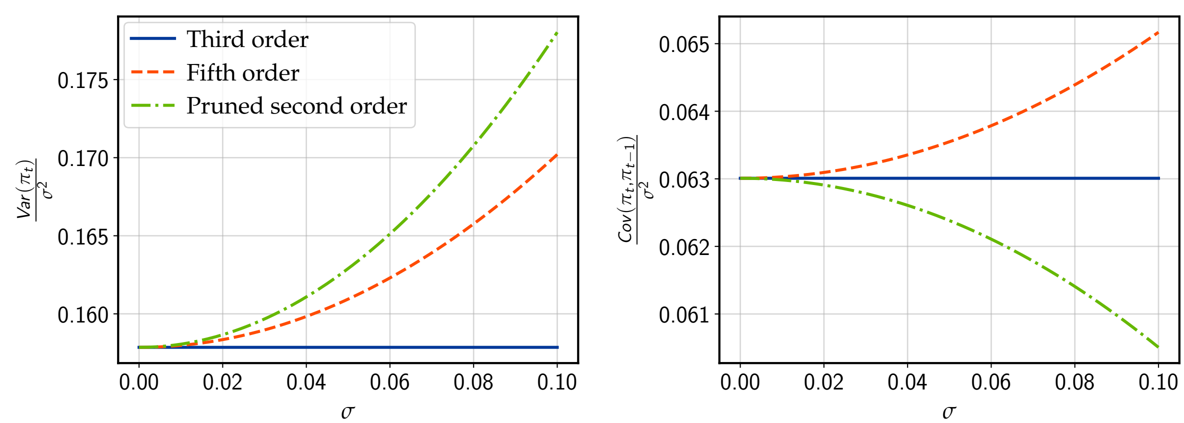
Moments. We then turn to two ergodic moments of the model: the variance of inflation and the autocovariance of inflation. Relative to the typical second-order accurate moment, with large cost shocks a fifth-order accurate solution significantly raises the variance of inflation and lowers its autocovariance. Further, the pruned second-order autocovariance of inflation here corrects in the wrong direction, relative to the fifth-order accurate moment, demonstrating again the benefit of third order perturbation to accurately solve for these moments.



(a) Response at $t = 0$

(b) Response at $t = 1$

Figure 12: Response of the price level to marginal cost shocks



(a) Inflation variance

(b) Inflation first autocovariance

Figure 13: Higher order approximation of moments in the menu cost model

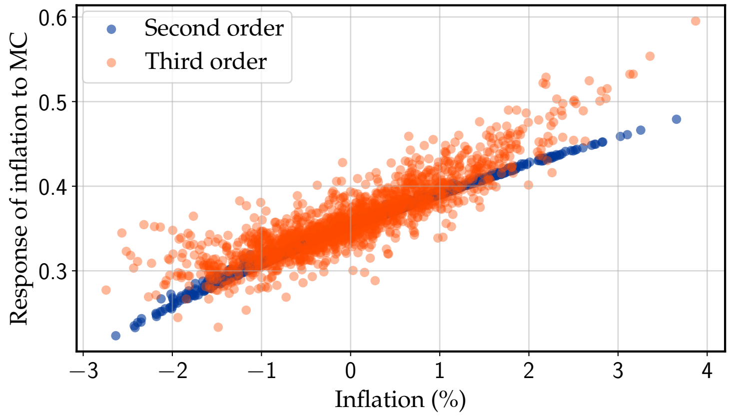


Figure 14: Time-varying response of inflation to marginal costs for $\sigma = 0.025$

State-dependent responsiveness of inflation. Finally, we evaluate the state dependence in the responsiveness of inflation to marginal cost shocks, $d\pi_t/d\epsilon_t$, conducting the same types of second and third order simulations as in our fiscal multiplier application. As pointed out by other studies on nonlinear Phillips curves (eg Blanco et al. 2024), at higher rates of current inflation, the responsiveness of inflation steepens. This effect is particularly pronounced in the third order solution, since then the relationship between average inflation and inflation responsiveness becomes convex.

5 Conclusion

This paper shows how to obtain higher-order perturbation solutions to heterogeneous-agent models using information derived from well-known routines to solve the model under perfect foresight, which we augment to solve rapidly for higher-order general equilibrium. This allows us to solve a HANK model and a menu cost model to third order at a much lower computational burden than a state-space solution would require. Our checks for accuracy against both nonlinear perfect foresight and a global solution perform well. An open question, however, is the extent to which our third order perturbation solutions differ from those obtained by global solution methods. This is an active area of research.

References

Alvarez, Fernando, Francesco Lippi, and Panagiotis Souganidis, "Price Setting with Strategic Complementarities as a Mean Field Game," *Econometrica*, 2023, *91* (6), 2005–2039.

Andreasen, Martin M, Jesús Fernández-Villaverde, and Juan F Rubio-Ramírez, "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications," *Review of Economic Studies*, 2018, *85* (1), 1–49.

Auclert, Adrien and Matthew Rognlie, "Inequality and Aggregate Demand," *Working Paper*, 2018.

__ , **Bence Bardóczy, Matthew Rognlie, and Ludwig Straub**, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

__ , **Evan Majic, Matthew Rognlie, and Ludwig Straub**, "Determinacy and Large-Scale Solutions in the Sequence Space," *Journal of Political Economy: Macroeconomics (forthcoming)*, 2025.

__ , **Matthew Rognlie, and Ludwig Straub**, "The Trickling up of Excess Savings," *AEA Papers and Proceedings*, 2023, *113*, 70–75.

__ , __ , **and** __ , "The Intertemporal Keynesian Cross," *Journal of Political Economy*, 2024, *132* (12), 4068–4121.

__ , __ , __ , **and Tomas Tapak**, "When do Endogenous Portfolios Matter for HANK?," *Working Paper*, 2024.

__ , **Rodolfo Rigato, Matthew Rognlie, and Ludwig Straub**, "New Pricing Models, Same Old Phillips Curves?," *Quarterly Journal of Economics*, 2024, *139* (1), 121–186.

Auerbach, Alan J and Yuriy Gorodnichenko, "Measuring the Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 2012, *4* (2), 1–27.

Azinovic-Yang, Marlon and Jan Žemlička, "Deep Learning in the Sequence Space," *arXiv preprint arXiv:2509.13623*, 2025.

Baley, Isaac and Andrés Blanco, "Firm uncertainty cycles and the propagation of nominal shocks," *American Economic Journal: Macroeconomics*, 2019, *11* (1), 276–337.

Bayer, Christian, Luigi-Maria Briglia, Ralph Luetticke, Christian Weiss, and Yannik Winkelmann, "Local Methods for Large Transfers," *Working Paper*, 2026.

__ , **Ralph Luetticke, Maximilian Weiss, and Yannik Winkelmann**, "An Endogenous Gridpoint Method for Distributional Dynamics," *Journal of Monetary Economics*, 2026, p. 103895.

Bhandari, Anmol, Thomas Bourany, David Evans, and Mikhail Golosov, "A Perturbational Approach for Approximating Heterogeneous-Agent Models," *Working Paper*, 2023.

Bianchi, Javier and Greg Kaplan, "How Small is Small? Non-linearities in Heterogeneous Agent Models," *Working Paper*, 2026.

Bilal, Adrien, "Solving Heterogeneous Agent Models with the Master Equation," *Working Paper*, 2023.

Blanchard, Olivier Jean and Charles M. Kahn, "The Solution of Linear Difference Models Under Rational Expectations," *Econometrica*, 1980, *48* (5), 1305–1311.

Blanco, Andres, Corina Boar, Callum Jones, and Virgiliu Midrigan, "Non-linear Inflation Dynamics in Menu Cost Economies," *Working Paper*, 2024.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT Shocks in Heterogeneous-Agent Economies: The Impulse Response as a Numerical Derivative," *Journal of Economic Dynamics and Control*, 2018, *89*, 68–92.

Carroll, Christopher D., "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics Letters*, 2006, *91* (3), 312–320.

Coeurdacier, Nicolas, Hélène Rey, and Pablo Winant, "The Risky Steady State," *American Economic Review Papers and Proceedings*, 2011, *101* (3), 398–401.

Druedahl, Jeppe, Raphael Huleux, and Inge Røpke, "Fiscal Multipliers in a Globally Solved HANK Model with Aggregate Risk," *Working Paper*, 2026.

Fernández-Villaverde, Jesús, Juan Rubio-Ramírez, and Frank Schorfheide, "Chapter 9 - Solution and Estimation Methods for DSGE Models," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, 2016, pp. 527–724.

Ganong, Peter, Pascal Noel, Christina Patterson, Joseph Vavra, and Alexander Weinberg, "Earnings Instability," *Working Paper*, 2025.

Golosov, Mikhail and Robert E. Lucas, "Menu Costs and Phillips Curves," *Journal of Political Economy*, 2007, *115* (2), 171–199.

Grand, François Le and Xavier Ragot, "Refining the Truncation Method to Solve Heterogeneous-Agent Models," *Annals of Economics and Statistics*, 2022, *146*, 65–92.

Hamilton, James D., *Time Series Analysis*, Princeton University Press, 1994.

Han, Jiequn, Yucheng Yang et al., "DeepHAM: A Global Solution Method for Heterogeneous Agent Models with Aggregate Shocks," *arXiv preprint arXiv:2112.14377*, 2021.

Iskhakov, Fedor, Thomas H. Jørgensen, John Rust, and Bertel Schjerning, "The Endogenous Grid Method for Discrete-Continuous Dynamic Choice Models with (or without) Taste Shocks," *Quantitative Economics*, 2017, *8* (2), 317–365.

Kase, Hanno, Leonardo Melosi, and Matthias Rottner, "Estimating Nonlinear Heterogeneous Agents Models with Neural Networks," *Working Paper*, 2022.

Kim, Jinill, Sunghyun Kim, Ernst Schaumburg, and Christopher A. Sims, "Calculating and Using Second-Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models," *Journal of Economic Dynamics and Control*, 2008, *32* (11), 3397–3414.

Klein, Paul, "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 2000, *24* (10), 1405–1423.

Kopecky, Karen and Richard Suen, "Finite State Markov-chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 2010, *13* (3), 701–714.

Lan, Hong and Alexander Meyer-Gohde, "Pruning in Perturbation DSGE Models: Guidance from Nonlinear Moving Average Approximations," *Working Paper*, 2013.

_ and _ , "Solving DSGE Models with a Nonlinear Moving Average," *Journal of Economic Dynamics and Control*, 2013, *37* (12), 2643–2667.

Lee, Hanbaek, "Global Nonlinear Solutions in Sequence Space and the Generalized Transition Function," 2025.

Lombardo, Giovanni and Harald Uhlig, "A Theory of Pruning," *International Economic Review*, 2018, *59* (4), 1825–1836.

Ramey, Valerie A and Sarah Zubairy, "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of political economy*, 2018, *126* (2), 850–901.

Reiter, Michael, "State Reduction and Second-order Perturbations of Heterogeneous Agent Models," *Working Paper*, 2023.

Schmitt-Grohé, Stephanie and Martín Uribe, "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function," *Journal of Economic Dynamics and Control*, 2004, *28* (4), 755–775.

_ and Martín Uribe, "Heterogeneous Downward Nominal Wage Rigidity: Foundations of a Nonlinear Phillips Curve," *Working Paper*, 2022.

Yang, Yucheng, Chiyuan Wang, Andreas Schaab, and Benjamin Moll, "Structural Reinforcement Learning for Heterogeneous Agent Macroeconomics," *arXiv preprint arXiv:2512.18892*, 2025.

Young, Eric R., "Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell–Smith Algorithm and Non-Stochastic Simulations," *Journal of Economic Dynamics and Control*, 2010, *34* (1), 36–41.

Appendix for "Higher-Order Perturbation in the Sequence Space: the Certainty Correspondence"

A Appendix to sections 2 and 3

A.1 Rank condition for assumption 1

Define

$$M \equiv \begin{pmatrix} -F_2 & -F_1 \\ I & 0 \end{pmatrix}, \qquad N \equiv \begin{pmatrix} F_3 & 0 \\ 0 & I \end{pmatrix},$$

where, as in the main text, F_1, F_2, and F_3 are partial derivatives of F with respect to its first three arguments at the deterministic steady state. Note that these satisfy the identity $\det(M - \lambda N) = \det(F_1 + \lambda F_2 + \lambda^2 F_3)$, where we use the latter in the main text for our root condition.

Let $Z_s \in \mathbb{R}^{2n\times n}$ consist of the columns of Z associated with the n roots inside the unit circle in an ordered real QZ decomposition of (M, N), and partition

$$Z_s = \begin{pmatrix} Z_0 \\ Z_{-1} \end{pmatrix}, \qquad Z_0, Z_{-1} \in \mathbb{R}^{n\times n},$$

where the two blocks correspond to x_t and x_{t-1}, respectively. Then we assume

$$\operatorname{rank}(Z_{-1}) = n, \qquad \text{or equivalently} \qquad \det Z_{-1} \neq 0. \tag{A.1}$$

Interpretation. For any solution of the deterministic linearized system $F_1 x_{t-1} + F_2 x_t + F_3 x_{t+1} = 0$ that converges back to the steady state, the pair (x_t, x_{t-1}) can be written as

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = Z_s a = \begin{pmatrix} Z_0 a \\ Z_{-1} a \end{pmatrix}$$

for some $a \in \mathbb{R}^n$. Condition (A.1) implies that, for each lagged state x_{t-1}, there is a unique coefficient vector

$$a = Z_{-1}^{-1} x_{t-1},$$

and therefore a unique current state consistent with convergence back to the steady state:

$$x_t = Z_0 Z_{-1}^{-1} x_{t-1}.$$

If, on the other hand, (A.1) is violated and Z_{-1} is singular, there is some nonzero a such that $Z_{-1}a = 0$. The columns of Z_s are independent by construction, so $a \neq 0$ implies $Z_s a \neq 0$; and since $Z_{-1}a = 0$, this means $Z_0 a \neq 0$. Hence in this case, the same initial state $x_{t-1} = 0$ is consistent both with $x_t = 0$ and with the nonzero convergent path beginning at $x_t = Z_0 a$, violating uniqueness.

(Similarly, for some x_{t-1} there is no a such that $Z_{-1}a = x_{t-1}$, meaning that there is no convergent continuation path, violating existence.)

Condition (A.1) is the same, up to notation and ordering, as the invertibility part of Assumption 4.5 in Klein (2000). See also section 5.3.1 in that paper for discussion of why root count alone is not sufficient.

A.2 Sequence-space form of assumption 1

We now show that assuming continuous differentiability of F, the rest of assumption 1 is equivalent to the invertibility of a certain sequence-space Jacobian. This makes it possible to directly test the assumption using sequence-space methods, as in Auclert, Majic, Rognlie and Straub (2025).

Concretely, let ℓ^2 be the space of square-summable paths $x = (x_0, x_1, \dots)$ in $\mathbb{R}^n$. Define the bounded linear operator $J : \ell^2 \to \ell^2$ as

$$(Jx)_0 = F_2 x_0 + F_3 x_1, \qquad (Jx)_t = F_1 x_{t-1} + F_2 x_t + F_3 x_{t+1} \quad (t \geq 1). \tag{A.2}$$

Lemma 2. The root and rank conditions in assumption 1 hold if and only if J is invertible on ℓ^2.

Proof. Suppose first that the root and rank conditions hold. Write the full ordered real QZ decomposition from appendix A.1:

$$Q^\top N Z = \begin{pmatrix} S_s & S_{su} \\ 0 & S_u \end{pmatrix}, \qquad Q^\top M Z = \begin{pmatrix} T_s & T_{su} \\ 0 & T_u \end{pmatrix}. \tag{A.3}$$

Write $Z = (Z_s \; Z_u)$, and let Z^u_{-1} denote the lower n rows of Z_u. For $Jx = r$, write

$$\begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix} = Z \begin{pmatrix} y^s_t \\ y^u_t \end{pmatrix}, \qquad Q^\top \begin{pmatrix} r_t \\ 0 \end{pmatrix} = \begin{pmatrix} q^s_t \\ q^u_t \end{pmatrix}. \tag{A.4}$$

The lower block of the transformed system is $S_u y^u_{t+1} = T_u y^u_t + q^u_t$. Since this is unstable, we can solve backward, with convergence guaranteed by square summability:

$$y^u_t = -\sum_{j=0}^{\infty} (T_u^{-1} S_u)^j T_u^{-1} q^u_{t+j}. \tag{A.5}$$

The upper, stable block can then be solved forward:

$$y^s_{t+1} = S_s^{-1} T_s y^s_t + S_s^{-1} \left(T_{su} y^u_t + q^s_t - S_{su} y^u_{t+1} \right), \tag{A.6}$$

$$y^s_0 = -Z_{-1}^{-1} Z^u_{-1} y^u_0, \tag{A.7}$$

where the second equation imposes $x_{-1} = 0$. Both $T_u^{-1} S_u$ and $S_s^{-1} T_s$ have geometrically decaying powers. Equations (A.5)–(A.7) therefore give a unique $x \in \ell^2$ for every $r \in \ell^2$, with $\|x\|_2 \leq C \|r\|_2$. Hence J is invertible.

Conversely, suppose that J is invertible. First, there can be no root on the unit circle. Otherwise, there are $|\lambda| = 1$ and $v \neq 0$ such that $(F_1 + \lambda F_2 + \lambda^2 F_3)v = 0$. We observe that J has a bounded inverse on complex-valued paths, obtained by applying J^{-1} separately to their real and imaginary parts. For $L \geq 1$, set $x_t^{(L)} = L^{-1/2}\lambda^{t+1}v$ for $0 \leq t < L$ and zero thereafter. Then $\|x^{(L)}\|_2 = |v|$ while $\|Jx^{(L)}\|_2 = O(L^{-1/2})$, contradicting boundedness of J^{-1}.

Now let m be the number of roots inside the unit circle, and let $Z_s \in \mathbb{R}^{2n\times m}$ consist of the corresponding columns of Z. With no forcing, any component outside this stable subspace fails to converge back to zero. A convergent path must therefore begin in this subspace. Writing the lower $n \times m$ block of Z_s as Z_{-1}, its lagged state is $x_{-1} = Z_{-1}c$ for some $c \in \mathbb{R}^m$.

For any $a \in \mathbb{R}^n$, define $\tilde{r} \in \ell^2$ by $\tilde{r}_0 = -F_1 a$ and $\tilde{r}_t = 0$ for $t \geq 1$. By the definition of J, the unique solution to $Jx = \tilde{r}$ is the unique square-summable solution of the homogeneous deterministic system with $x_{-1} = a$. Thus every a corresponds to exactly one c satisfying $a = Z_{-1}c$. This map is therefore a bijection, which requires $m = n$ and $\operatorname{rank}(Z_{-1}) = n$. These are the root and rank conditions in assumption 1. □

A.3 Proof of proposition 1 and lemma 1

The main challenge for proposition 1 is that the date-0 equilibrium equation contains $\mathbb{E}_0 X_1$, so X_0 must be solved in a way that is consistent with its continuation after every possible sequence of future innovations. Our proof proceeds in three parts. First, we show that the linearized continuation problem has a unique bounded solution. Second, the implicit-function theorem gives a nonlinear continuation whose date-0 value defines the locally stable state-space solution g, and we establish its local uniqueness. Third, we use g to construct the unique sequence-space solution $\mathcal{X}$ and establish the equivalence in (4). Lemma 1 then follows as a simple corollary of uniqueness.

1. The linear continuation problem. The argument follows the standard QZ approach to linear rational expectations models, as in Klein (2000). Here it is written for bounded continuations indexed by innovation histories.

Let $\mathcal{E}$ denote the bounded support of $\bar{\epsilon}_t$. Write $\bar{\epsilon}^t = (\bar{\epsilon}_1, \ldots, \bar{\epsilon}_t)$. By a *continuation* we mean a bounded sequence $x = \{x_t\}_{t\geq 0}$, where $x_0 \in \mathbb{R}^n$ and each $x_t : \mathcal{E}^t \to \mathbb{R}^n$ is measurable for $t \geq 1$, with norm

$$\|x\|_\infty \equiv \sup_{t\geq 0} \sup_{\bar{\epsilon}^t \in \mathcal{E}^t} |x_t(\bar{\epsilon}^t)|. \tag{A.8}$$

These continuations form a Banach space, and conditional expectation is a bounded linear operator on this space.

Lemma 3 (Linear continuation)**.** Under assumption 1, let $x_{-1} \in \mathbb{R}^n$ and let r be any uniformly bounded continuation. There is a unique uniformly bounded continuation x satisfying

$$F_1 x_{t-1} + F_2 x_t + F_3 \mathbb{E}_t x_{t+1} = r_t, \qquad t \geq 0, \tag{A.9}$$

after every history. Moreover,

$$\|x\|_\infty \leq C\left(|x_{-1}| + \|r\|_\infty\right) \tag{A.10}$$

for a constant C independent of x_{-1} and r.

Proof. Use the QZ decomposition in (A.3) and the transformed coordinates in (A.4). Conditional expectations change the unstable block to $S_u \mathbb{E}_t y^u_{t+1} = T_u y^u_t + q^u_t$. The same backward calculation as in (A.5) gives

$$y^u_t = -\sum_{j=0}^{\infty} (T_u^{-1} S_u)^j T_u^{-1} \mathbb{E}_t q^u_{t+j}. \tag{A.11}$$

The stable equation (A.6) continues to hold with y^s_{t+1} and y^u_{t+1} replaced by their conditional expectations, and therefore determines $\mathbb{E}_t y^s_{t+1}$. With the same partition of Z, the lower half of (x_{t+1}, x_t) must equal x_t after every realization of $\bar{\epsilon}_{t+1}$. Its innovation must therefore vanish:

$$y^s_{t+1} - \mathbb{E}_t y^s_{t+1} = -Z_{-1}^{-1} Z^u_{-1}\left(y^u_{t+1} - \mathbb{E}_t y^u_{t+1}\right). \tag{A.12}$$

Together with the conditional mean, (A.12) determines y^s_{t+1} after each realization and makes the reconstructed lower block equal x_t. The initial condition is $y^s_0 = Z_{-1}^{-1}(x_{-1} - Z^u_{-1} y^u_0)$.

The matrix powers in both recursions decay geometrically, as established in the proof of lemma 2, and conditional expectation does not increase the sup norm. The same geometric-series bounds therefore give a unique bounded continuation satisfying (A.10). □

2. The state-space solution. We now solve the nonlinear continuation problem. For a continuation $\mathbf{Y} = (Y_0, Y_1, \ldots)$, define the residual mapping $\mathcal{R}$, which maps $\mathbf{Y}$ to its equilibrium residual after each possible history:

$$\mathcal{R}_0(\mathbf{Y}, X_{-1}, \epsilon_0, \sigma) = F\left(X_{-1}, Y_0, \mathbb{E}_0 Y_1, \epsilon_0\right), \tag{A.13}$$

$$\mathcal{R}_t(\mathbf{Y}, X_{-1}, \epsilon_0, \sigma)(\bar{\epsilon}^t) = F\left(Y_{t-1}(\bar{\epsilon}^{t-1}), Y_t(\bar{\epsilon}^t), \mathbb{E}_t Y_{t+1}(\bar{\epsilon}^t), \sigma\bar{\epsilon}_t\right), \qquad t \geq 1. \tag{A.14}$$

Thus $\mathcal{R} = 0$ requires $\mathbf{Y}$ to satisfy the equilibrium equation at date 0 and after every possible future innovation history. At $(X_{-1}, \epsilon_0, \sigma) = (X, 0, 0)$, the constant continuation $\mathbf{Y} = (X, X, \ldots)$ sets every residual to zero. Restrict $\mathbf{Y}$ and $(X_{-1}, \epsilon_0, \sigma)$ to a sufficiently small neighborhood, so that every argument passed to F lies in a compact subset of the neighborhood on which F is continuously differentiable. The derivatives of F are then bounded and uniformly continuous, so pointwise application of F, together with the bounded linear expectation operator, makes $\mathcal{R}$ continuously differentiable. If F is $\mathcal{C}^r$, then so is $\mathcal{R}$.

The derivative of $\mathcal{R}$ with respect to $\mathbf{Y}$ at the steady state is

$$[D_{\mathbf{Y}}\mathcal{R}\, x]_t = F_1 x_{t-1} + F_2 x_t + F_3 \mathbb{E}_t x_{t+1}, \qquad x_{-1} = 0.$$

Lemma 3 gives its bounded inverse. The implicit-function theorem therefore gives a unique nearby continuation $\mathbf{Y}(X_{-1}, \epsilon_0, \sigma)$ satisfying $\mathcal{R} = 0$, with the same differentiability as F. Define

$$g(X_{-1}, \epsilon_0, \sigma) \equiv Y_0(X_{-1}, \epsilon_0, \sigma). \tag{A.15}$$

In particular, $g(X, 0, 0) = X$.

After any realization $\bar{\epsilon}_1$, the continuation from date 1 onward, $(Y_1, Y_2, \ldots)$, satisfies $\mathcal{R} = 0$ with lagged state Y_0 and current shock $\sigma\bar{\epsilon}_1$. Uniqueness gives

$$Y_1(\bar{\epsilon}_1) = g(Y_0, \sigma\bar{\epsilon}_1, \sigma),$$

and repeating this at each later date shows that g generates the continuation and satisfies the state-space equilibrium equation.

Set $\epsilon_0 = \sigma = 0$ and differentiate the continuation with respect to X_{-1}. The resulting perturbation is a bounded deterministic solution of (A.9) with $r = 0$. Equation (A.11) sets $y_t^u = 0$, while $y_{t+1}^s = S_s^{-1} T_s y_t^s$ converges to zero. Appendix A.1 identifies this solution as $x_t = Z_0 Z_{-1}^{-1} x_{t-1}$. Because the continuation is generated by g, this gives $g_X(X, 0, 0) = Z_0 Z_{-1}^{-1}$. The resulting path converges to zero for every initial perturbation, so g_X has spectral radius less than one. Thus g is locally stable at X.

It remains to establish uniqueness among continuous locally stable solutions. Let $\widetilde{g}$ be another such solution, and let $\{X_t\}_{t \geq 0}$ denote the continuation it generates. Choose m with $\|\widetilde{g}_X^m\| < 1$. The deterministic m-period map is differentiable at X with derivative $\widetilde{g}_X^m$, so it maps a sufficiently small closed ball strictly inside itself. After shrinking the ball and shocks if necessary, continuity implies that whenever X_t lies in the ball, X_{t+m} also lies in the ball, while $X_{t+1}, \ldots, X_{t+m-1}$ remain in the neighborhood where $\mathcal{R} = 0$ has a unique solution. Thus every X_t remains in that neighborhood. Because $\widetilde{g}$ solves the state-space equilibrium equation, this continuation satisfies $\mathcal{R} = 0$ and must equal $\mathbf{Y}$. Hence $\widetilde{g} = g$.

3. The sequence-space solution. Let ℓ_-^∞ denote the space of bounded sequences indexed by the nonpositive integers. Write $\boldsymbol{\epsilon} = (\epsilon_0, \epsilon_{-1}, \ldots)$ and $\mathbf{X} = (X_0, X_{-1}, \ldots)$. For $\mathbf{X}$ near the constant sequence $(X, X, \ldots)$ and $(\boldsymbol{\epsilon}, \sigma)$ near $(0, 0)$, define

$$[\Phi(\mathbf{X}, \boldsymbol{\epsilon}, \sigma)]_j = X_j - g(X_{j-1}, \epsilon_j, \sigma), \qquad j \leq 0. \tag{A.16}$$

At the deterministic steady state, $\mathbf{X} = (X, X, \ldots)$ and $(\boldsymbol{\epsilon}, \sigma) = (0, 0)$, so $\Phi = 0$. Since g is continuously differentiable, so is Φ on a sufficiently small sup-norm neighborhood. Its derivative with respect to $\mathbf{X}$ is

$$[D_{\mathbf{X}} \Phi\, h]_j = h_j - g_X h_{j-1},$$

where g_X is evaluated at $(X, 0, 0)$. For $r \in \ell^\infty_-$, the solution to $D_{\mathbf{X}}\Phi\, h = r$ is

$$h_j = \sum_{k=0}^{\infty} g_X^k r_{j-k}. \tag{A.17}$$

The series defines a bounded h because $\rho(g_X) < 1$ implies $\sum_{k\geq 0} \|g_X^k\| < \infty$. A bounded solution to the homogeneous equation satisfies $h_j = g_X^k h_{j-k}$ for every $k \geq 1$ and must therefore vanish. Thus $D_{\mathbf{X}}\Phi$ has a bounded inverse.

The implicit-function theorem gives a unique $\mathbf{X}(\boldsymbol{\epsilon}, \sigma)$ near $(X, X, \ldots)$ satisfying (A.16) for every sufficiently small bounded $\boldsymbol{\epsilon}$ and sufficiently small σ. If F is $\mathcal{C}^r$, then g, Φ, and $\mathbf{X}$ are also $\mathcal{C}^r$. Define

$$\mathcal{X}(\sigma, \epsilon_0, \epsilon_{-1}, \ldots) \equiv X_0(\boldsymbol{\epsilon}, \sigma). \tag{A.18}$$

Thus $\mathcal{X}$ is $\mathcal{C}^r$ whenever F is $\mathcal{C}^r$.

The sequence $(X_{-1}, X_{-2}, \ldots)$ satisfies (A.16) for the shock history $(\epsilon_{-1}, \epsilon_{-2}, \ldots)$. Uniqueness therefore gives

$$X_{-1}(\boldsymbol{\epsilon}, \sigma) = \mathcal{X}(\sigma, \epsilon_{-1}, \epsilon_{-2}, \ldots).$$

The date-0 equation in (A.16) is therefore (4). After a future innovation $\bar{\epsilon}_1$, equation (4) gives

$$\mathcal{X}(\sigma, \sigma\bar{\epsilon}_1, \epsilon_0, \epsilon_{-1}, \ldots) = g\left(X_0, \sigma\bar{\epsilon}_1, \sigma\right).$$

Taking $\mathbb{E}_0$ of this equality, the state-space equilibrium equation for g becomes (1) for $\mathcal{X}$.

It remains to prove uniqueness. Let $\widetilde{\mathcal{X}}$ be another local sequence-space solution and fix $\boldsymbol{\epsilon} = (\epsilon_0, \epsilon_{-1}, \ldots)$. Set $Y_0 = \widetilde{\mathcal{X}}(\sigma, \epsilon_0, \epsilon_{-1}, \ldots)$ and, after future innovations $\bar{\epsilon}^t$, set

$$Y_t(\bar{\epsilon}^t) = \widetilde{\mathcal{X}}\left(\sigma, \sigma\bar{\epsilon}_t, \ldots, \sigma\bar{\epsilon}_1, \epsilon_0, \epsilon_{-1}, \ldots\right), \qquad t \geq 1.$$

Continuity of $\widetilde{\mathcal{X}}$ and boundedness of $\bar{\epsilon}_t$ imply that, for sufficiently small $\boldsymbol{\epsilon}$ and σ, every Y_t remains in the neighborhood where $\mathcal{R} = 0$ has a unique solution. Since $\widetilde{\mathcal{X}}$ solves (1), the continuation $\mathbf{Y} = (Y_0, Y_1, \ldots)$ satisfies $\mathcal{R} = 0$ with lagged value $\widetilde{\mathcal{X}}(\sigma, \epsilon_{-1}, \epsilon_{-2}, \ldots)$ and current shock ϵ_0. The uniqueness proved in Part 2 therefore gives (4) with $\mathcal{X}$ replaced by $\widetilde{\mathcal{X}}$.

For the fixed shock history, define $\widetilde{X}_j = \widetilde{\mathcal{X}}(\sigma, \epsilon_j, \epsilon_{j-1}, \ldots)$ for $j \leq 0$. Substitution into (4) gives $\Phi(\widetilde{\mathbf{X}}, \boldsymbol{\epsilon}, \sigma) = 0$. Continuity of $\widetilde{\mathcal{X}}$ places $\widetilde{\mathbf{X}}$ in the neighborhood where (A.16) has a unique solution. Therefore $\widetilde{\mathbf{X}} = \mathbf{X}$. Since $\boldsymbol{\epsilon}$ was arbitrary, $\widetilde{\mathcal{X}} = \mathcal{X}$ throughout the neighborhood. This completes the proof of proposition 1.

Symmetry in σ and proof of lemma 1. Because $\bar{\epsilon}_t$ is symmetric around zero, $\sigma\bar{\epsilon}_t$ and $-\sigma\bar{\epsilon}_t$ have the same distribution. Replacing σ by $-\sigma$ therefore leaves the conditional distribution of future shocks, and hence the equilibrium equations after every history, unchanged. The local uniqueness established above implies $g(X_{t-1}, \epsilon_t, \sigma) = g(X_{t-1}, \epsilon_t, -\sigma)$ and $\mathcal{X}(\sigma, \epsilon_t, \epsilon_{t-1}, \ldots) = \mathcal{X}(-\sigma, \epsilon_t, \epsilon_{t-1}, \ldots)$. Thus both functions are even in σ. Every derivative involving an odd number of differentiations with respect to σ is therefore zero at $\sigma = 0$, whenever that derivative exists.

A.4 Calibration of heterogeneous-agent models

Here we provide more details about the computation and calibration of our two heterogeneous-agent models introduced in section 2.2.

Partial equilibrium (smooth SIM) model. Optimal consumption and asset policy conditional on (a, e, β, ξ) are characterized by a first-order condition $u'(c) \geq \beta \mathbb{E}_t[V_{a,t+1}(a', e', \beta')|e, \beta]$, which holds with equality unless the borrowing constraint binds. The marginal value function $V_{a,t}$ is then given by the envelope condition $V_{a,t}(a, e, \beta) = (1 + r)\mathbb{E}_\xi[u'(c(a, e, \beta, \xi))]$.

For each e, β, let $\Phi_t(a', e, \beta)$ denote the end-of-period CDF of assets a' saved by households, conditional on current income state e and patience state β. Given the Φ_{t-1} from the end of the previous period, the asset policy at t, together with the fixed Π and distribution of ξ, determines Φ_t. Finally, aggregate consumption at t is given by

$$C_t = \sum_{(e,\beta),(e_-,\beta_-)} \Pi_{(e_-,\beta_-),(e,\beta)} \int\int c_t(a, e, \beta, \xi) f(\xi) d\xi d\Phi_{t-1}(a, e_-, \beta_-).$$

In practice, we represent $V_{a,t}(a, e, \beta)$ using its value at n_a asset gridpoints a, for each of the n_e income states e and n_β patience states β, and let $\boldsymbol{v}_t$ be the vector of these $n_a \times n_e \times n_\beta$ values. We represent $\Phi_t(a', e, \beta)$ with a cubic spline interpolating over the same gridpoints for each e, β, and let $\boldsymbol{\phi}_t$ be the vector of the $n_a \times n_e \times n_\beta$ spline coefficients.

To put the system in the notation of (1), we stack these vectors, along with aggregate C_t and Z_t, in the vector $X_t \equiv (\boldsymbol{v}_t, \boldsymbol{\phi}_t, C_t, Z_t)$. We then stack the $2(n_a n_e n_\beta + 1)$ equations into the system displayed in 7 in the main text, where $\mathcal{V}$, Λ, and $\mathcal{C}$ are constructed as described above, acting on the discretized representation.

Our calibration, which is at a quarterly frequency, mostly follows Auclert et al. (2024a), except for the addition of the smooth i.i.d. income shocks. We set the annual interest rate to $r = 2\%$, and pick a two-point process for β with support (β_L, β_H), with half of the population in each state and probability of drawing a new β state of 1% quarterly. We set $Tr = 0.14$ and $Z = 0.67$ to be consistent with U.S. tax progressivity, as discussed in the next section.

We assume that the primitive log income process is the sum of two independent components, with the persistent component following an AR(1) with persistence $\rho_e = 0.91$ at the annual level. The innovations to both processes are normally distributed with standard deviation σ_e for the persistent component and σ_ξ for the transitory component. We solve for σ_e and σ_ξ so that the cross-sectional standard deviation of income is 0.92 overall and 0.24 in quarterly changes, as reported in Ganong, Noel, Patterson, Vavra and Weinberg (2025). This yields $\sigma_e = 0.1964$ and $\sigma_\xi = 0.0964$.

We set $\underline{a} = 0$, and set the asset gridpoints on which to store the value function (which are also the knots for the cubic spline representing the distribution) to be $n_a = 200$ double-exponentially spaced points. In the deterministic steady state, we pick β_L and β_H so that the first-order aggregate marginal propensity to consume $\frac{\partial C_0}{\partial Z_0}$ is equal to 0.2 and aggregate assets are $A = 4$. This procedure yields $(\beta_L, \beta_H) = (0.96, 0.99)$: as is common for these models, some patience heterogeneity is required to simultaneously hit our high MPC and high asset targets. For our aggregate income process, we assume that aggregate income follows an i.i.d. income process $\rho_Z = 0$.

General equilibrium (IKC) model. Our calibration of the GE model exactly matches that of our PE model, as follows. We normalize income to $Y = 1$, set government spending at $G = 0.17$ and let the government transfer be $Tr = 0.14$. This implies that the tax rate is $\tau = rB + G + Tr = 0.02 + 0.17 + 0.14 = 0.33$, so net-of-tax aggregate income is $Z = 0.67$. In turn, $Tr = 0.14$ and $\tau = 0.33$ are consistent with the U.S. relationship between pre- and post-tax income estimated in Auclert and Rognlie (2018). For our aggregate bond process, we assume a quarterly persitence of $\rho_B = 0.95$.

In this model, positive shocks to B_t enable the government to temporarily lower tax revenue $T_t = \tau_t Y_t$ and therefore increasing post-tax household income $Z_t = Y_t - T_t$. Since households are not Ricardian, this decline in T_t in turn affects consumption, which in equilibrium affects aggregate income Y_t and feeds back into consumption C_t. Formally, the model has the same equations as the partial equilibrium model, with the addition of the goods market clearing condition and government budget constraint,

$$F\left(X_{t-1}, X_t, \mathbb{E}_t\left[X_{t+1}\right], \epsilon_t\right) \equiv \begin{bmatrix} \boldsymbol{v}_t - \mathcal{V}(Y_t - T_t, \mathbb{E}_t \boldsymbol{v}_{t+1}) \\ \boldsymbol{\phi}_t - \Lambda(Y_t - T_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1}) \\ C_t - \mathcal{C}(Y_t - T_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1}) \\ T_t + (1+r)B_{t-1} - B_t - Tr \\ C_t + G - Y_t \\ B_t - 1 - \rho_B(B_{t-1} - 1) - \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{A.19}$$

The key difficulty is the endogeneity of aggregate income Y_t, so that the model can no longer be solved globally in the state space.

A.5 Proof of proposition 2

By Lemma 1, $\mathcal{X}_\sigma = 0$ at the deterministic steady state. The first-order Taylor expansion of the sequence-space solution therefore has no term in σ, leaving

$$X_t \approx X + \sum_{j=0}^{\infty} \mathcal{X}_j \epsilon_{t-j}.$$

It remains only to identify coefficients $\mathcal{X}_j$ with perfect-foresight impulse responses. By definition,

$$X_j^{MIT}(\nu) = \mathcal{X}(0, 0, \ldots, 0, \nu, 0, \ldots),$$

where ν is in the coordinate corresponding to ϵ_{t-j}. Differentiating with respect to ν at $\nu = 0$ gives

$$(X_j^{MIT})'(0) = \mathcal{X}_j.$$

This proves the first-order certainty equivalence result.

A.6 Proof of proposition 3

MIT shock terms. The size and history terms follow directly from the definitions of the perfect-foresight impulse responses. By definition,

$$X_j^{MIT}(\nu) = \mathcal{X}(0, 0, \ldots, 0, \nu, 0, \ldots),$$

where ν is in the coordinate corresponding to ϵ_{-j}. Differentiating twice with respect to ν gives

$$(X_j^{MIT})''(0) = \mathcal{X}_{jj}.$$

Similarly, for $l > 0$,

$$X_j^{MIT,l}(\epsilon, \nu) = \mathcal{X}(0, 0, \ldots, 0, \nu, 0, \ldots, 0, \epsilon, 0, \ldots),$$

where ν is in the coordinate corresponding to ϵ_{-j} and ϵ is in the coordinate corresponding to $\epsilon_{-(j+l)}$. Differentiating with respect to ν and ϵ gives

$$\frac{\partial^2 X_j^{MIT,l}}{\partial \nu \, \partial \epsilon}(0, 0) = \mathcal{X}_{j,j+l}.$$

This proves the size and history parts of proposition 3. It remains to identify the risk term $\mathcal{X}_{\sigma\sigma}$.

Risk terms. Now set all realized shocks through date t to zero, but let agents anticipate future shocks with standard deviation σ. Then

$$X_t = \mathcal{X}(\sigma, 0, 0, \ldots) = X + \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2 + o(\sigma^2). \tag{A.20}$$

At date $t+1$, the new shock is $\epsilon_{t+1} = \sigma\bar{\epsilon}_{t+1}$. Using $\mathbb{E}\bar{\epsilon}_{t+1} = 0$ and $\mathbb{E}\bar{\epsilon}_{t+1}^2 = 1$,

$$\mathbb{E}_t X_{t+1} = \mathbb{E}_t \mathcal{X}(\sigma, \sigma\bar{\epsilon}_{t+1}, 0, 0, \ldots) = X + \frac{1}{2}\left(\mathcal{X}_{\sigma\sigma} + \mathcal{X}_{00}\right)\sigma^2 + o(\sigma^2) \tag{A.21}$$

where the mixed $\sigma\epsilon$ term is zero by the symmetry in σ from lemma 1. The first-order shock term also drops out after taking the expectation.

Substitute (A.20) and (A.21) into

$$F(X_{t-1}, X_t, \mathbb{E}_t X_{t+1}, \epsilon_t) = 0.$$

Since the realized shock is zero and the path is constant through date t,

$$F\left(X+\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2,\ X+\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2,\ X+\frac{1}{2}\left(\mathcal{X}_{\sigma\sigma}+\mathcal{X}_{00}\right)\sigma^2,\ 0\right)=o(\sigma^2). \tag{A.22}$$

Differentiating (A.22) with respect to σ^2 around the steady state $\sigma^2=0$, we get $(F_1+F_2+F_3)\mathcal{X}_{\sigma\sigma}/2+F_3\mathcal{X}_{00}/2=0$, or

$$\mathcal{X}_{\sigma\sigma}=-(F_1+F_2+F_3)^{-1}F_3\mathcal{X}_{00} \tag{A.23}$$

Now define $X^{pss}(\delta)$ by

$$F\left(X^{pss}(\delta),\ X^{pss}(\delta),\ X^{pss}(\delta)+\mathcal{X}_{00}\delta,\ 0\right)=0 \tag{A.24}$$

where we note that $\delta=0$ is the deterministic steady state. Differentiating with respect to δ, we have $(X^{pss})'(0)=-(F_1+F_2+F_3)^{-1}F_3\mathcal{X}_{00}$, and hence

$$\mathcal{X}_{\sigma\sigma}=(X^{pss})'(0). \tag{A.25}$$

Alternative F formulation. Suppose instead that the model is written, as in Lan and Meyer-Gohde (2013b), as

$$\mathbb{E}_t F(X_{t-1},X_t,X_{t+1},\epsilon_t)=0. \tag{A.26}$$

Set realized shocks through date t to zero. Then at date t, we have the second-order approximation $X_t\approx X+\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2$. But the future variable inside F is random:

$$X_{t+1}\approx X+\mathcal{X}_0\epsilon_{t+1}+\frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2+\frac{1}{2}\mathcal{X}_{00}\epsilon_{t+1}^2+o(\sigma^2).$$

Substituting into (A.26), expanding, and taking expectations gives

$$(F_1+F_2+F_3)\mathcal{X}_{\sigma\sigma}+F_3\mathcal{X}_{00}+F_{33}(\mathcal{X}_0,\mathcal{X}_0)=0.$$

The first two terms are the same as in (A.22), but the last term (where $F_{33}(\mathcal{X}_0,\mathcal{X}_0)$ denotes applying the Hessian F_{33} to $\mathcal{X}_0$ on both input dimensions) is new. It appears because we take the expectation of applying F to the random X_{t+1}, so that the curvature of F matters.

Thus, in this formulation,

$$\mathcal{X}_{\sigma\sigma}=-(F_1+F_2+F_3)^{-1}\left[F_3\mathcal{X}_{00}+F_{33}(\mathcal{X}_0,\mathcal{X}_0)\right].$$

By contrast, in our formulation $F(X_{t-1},X_t,\mathbb{E}_tX_{t+1},\epsilon_t)=0$, the random first-order term $\mathcal{X}_0\epsilon_{t+1}$ is averaged before F is applied, so the extra term does not appear.

Indeed, Lan and Meyer-Gohde (2013b) have precisely this additional term in their equation (36); in their notation, it appears as $f_{y^{+2}}(y_0\otimes y_0)\mathbb{E}_t(\varepsilon_{t+1}\otimes\varepsilon_{t+1})$. Without this term, their equation (36) would parallel our $\mathcal{X}_{\sigma\sigma}$.

Although either formulation is valid and can be used to solve a model when the equations are written in a compatible way, our approach makes possible the simple certainty-correspondence statement of proposition 3, where the risk term $\mathcal{X}_{\sigma\sigma}$ can be obtained by perturbing expectations at the steady state with the size dependence of MIT shocks. With the (A.26) formulation, the extra term $F_{33}(\mathcal{X}_0, \mathcal{X}_0)$ cannot be obtained from size dependence alone.

A.7 Proof of proposition 4

MIT shock terms. As in the first- and second-order cases, the coefficients involving only realized shocks follow by differentiating the perfect-foresight impulse responses. By definition,

$$X_j^{MIT}(\nu) = \mathcal{X}(0, 0, \dots, 0, \nu, 0, \dots),$$

where ν is in the coordinate corresponding to ϵ_{-j}. Differentiating three times gives

$$(X_j^{MIT})'''(0) = \mathcal{X}_{jjj}.$$

For $k > j$, the two-shock impulse response

$$X_j^{MIT,k-j}(\epsilon, \nu) = \mathcal{X}(0, 0, \dots, 0, \nu, 0, \dots, 0, \epsilon, 0, \dots)$$

places ν in the coordinate corresponding to ϵ_{-j} and ϵ in the coordinate corresponding to ϵ_{-k}. Therefore

$$\frac{\partial^3 X_j^{MIT,k-j}}{\partial \nu^2\, \partial \epsilon}(0,0) = \mathcal{X}_{jjk}, \qquad \frac{\partial^3 X_j^{MIT,k-j}}{\partial \nu\, \partial \epsilon^2}(0,0) = \mathcal{X}_{jkk}.$$

Finally, for $k' > k > j$,

$$X_j^{MIT,k-j,k'-k}(\epsilon', \epsilon, \nu) = \mathcal{X}(0, 0, \dots, 0, \nu, 0, \dots, 0, \epsilon, 0, \dots, 0, \epsilon', 0, \dots),$$

with ν, ϵ, and ϵ' in the coordinates corresponding to ϵ_{-j}, ϵ_{-k}, and $\epsilon_{-k'}$. Differentiating once with respect to each shock gives

$$\frac{\partial^3 X_j^{MIT,k-j,k'-k}}{\partial \nu\, \partial \epsilon\, \partial \epsilon'}(0,0,0) = \mathcal{X}_{jkk'}.$$

This proves the pure MIT-shock parts of proposition 4. It remains to identify the mixed risk terms $\mathcal{X}_{\sigma\sigma j}$.

Mixed risk terms. Fixing $j \geq 0$, we will differentiate the date-t equilibrium condition

$$F(X_{t-1}, X_t, \mathbb{E}_t X_{t+1}, \epsilon_t) = 0$$

with respect to $\epsilon_{t-j}, \sigma, \sigma$, evaluating at the deterministic steady state, writing

$$u = [X_{t-1}, X_t, \mathbb{E}_t X_{t+1}, \epsilon_t]$$

for the stacked argument of F. Derivatives such as F_u, F_{uu}, and F_{uuu} are taken with respect to this stacked argument and evaluated at $[X, X, X, 0]$, and we will use subscripts j and σ on u for derivatives with respect to ϵ_{t-j} and σ. The resulting mixed derivative is:

$$F_u u_{j\sigma\sigma} + F_{uu}(u_{\sigma\sigma}, u_j) + 2F_{uu}(u_{j\sigma}, u_\sigma) + F_{uuu}(u_j, u_\sigma, u_\sigma) = 0 \tag{A.27}$$

where we use e.g. the notation $F_{uu}(v, w)$ to refer to the Hessian F_{uu} evaluated on vectors v and w, and we note that the last two terms vanish because $u_\sigma = 0$.

We now evaluate u_j, $u_{\sigma\sigma}$, and $u_{j\sigma\sigma}$ as needed for (A.27). First, we have

$$u_j = \left[\mathcal{X}_{j-1},\ \mathcal{X}_j,\ \mathcal{X}_{j+1},\ \mathbf{1}_{j=0}\right],$$

with $\mathcal{X}_{-1} = 0$. Then

$$u_{\sigma\sigma} = [\mathcal{X}_{\sigma\sigma},\ \mathcal{X}_{\sigma\sigma},\ \mathcal{X}_{\sigma\sigma} + \mathcal{X}_{00},\ 0],$$

where we note that here the third entry includes $\mathcal{X}_{00}$ because the new shock at $t+1$ is $\sigma\bar{\epsilon}_{t+1}$, and we take the expectation of its square in $\mathbb{E}_t X_{t+1}$. Finally,

$$u_{j\sigma\sigma} = \left[\mathcal{X}_{j-1,\sigma\sigma},\ \mathcal{X}_{j\sigma\sigma},\ \mathcal{X}_{j+1,\sigma\sigma} + \mathcal{X}_{j+1,00},\ 0\right],$$

where we similarly get a $\mathcal{X}_{j+1,00}$ term from expectations, and we write $\mathcal{X}_{-1,\sigma\sigma} = 0$.

Substituting these into the two surviving terms from (A.27), we get

$$\begin{aligned} F_1\mathcal{X}_{j-1,\sigma\sigma} + F_2\mathcal{X}_{j\sigma\sigma} + F_3\mathcal{X}_{j+1,\sigma\sigma} + F_{uu}\left([\mathcal{X}_{\sigma\sigma}, \mathcal{X}_{\sigma\sigma}, \mathcal{X}_{\sigma\sigma} + \mathcal{X}_{00}, 0],\right. \\ \left.[\mathcal{X}_{j-1}, \mathcal{X}_j, \mathcal{X}_{j+1}, \mathbf{1}_{j=0}]\right) + F_3\mathcal{X}_{j+1,00} = 0. \end{aligned} \tag{A.28}$$

We can interpret (A.28) as a linear system that can be solved for the unknown sequence $\{\mathcal{X}_{j,\sigma\sigma}\}$, given the forcing terms $F_{uu}(\cdot, \cdot)$ and $F_3\mathcal{X}_{j+1,00}$. We will characterize the contributions from each of these terms in turn.

RSS term. First reproduce the F_{uu} term in (A.28). Let δ shift the expected-future term by $\mathcal{X}_{00}\delta$, as in (A.24), and let ν be the size of the one-time shock:

$$F\left(Y^{rss}_{k-1}(\nu, \delta),\ Y^{rss}_k(\nu, \delta),\ Y^{rss}_{k+1}(\nu, \delta) + \mathcal{X}_{00}\delta,\ \nu\mathbf{1}_{k=0}\right) = 0, \qquad k \geq 0, \tag{A.29}$$

with $Y^{rss}_{-1}(\nu, \delta) = X^{pss}(\delta)$ and $Y^{rss}_k(\nu, \delta) \to X^{pss}(\delta)$ as $k \to \infty$.

For the date-j equation, the F_{uu} term in the mixed ν, δ derivative is F_{uu} evaluated on the δ and ν derivatives of the four arguments of F. The ν derivative at zero is

$$[\mathcal{X}_{j-1},\ \mathcal{X}_j,\ \mathcal{X}_{j+1},\ \mathbf{1}_{j=0}].$$

At $\nu = 0$, the path is constant: $Y^{rss}_k(0, \delta) = X^{pss}(\delta)$. Using the second-order result above, the δ

derivative at zero is

$$[\mathcal{X}_{\sigma\sigma},\ \mathcal{X}_{\sigma\sigma},\ \mathcal{X}_{\sigma\sigma}+\mathcal{X}_{00},\ 0].$$

Define $(\mathcal{X}_{j\sigma\sigma})^{rss} \equiv \partial_{\nu\delta} Y_j^{rss}(0,0)$. Differentiating (A.29) with respect to ν and δ at zero gives

$$F_1(\mathcal{X}_{j-1,\sigma\sigma})^{rss} + F_2(\mathcal{X}_{j\sigma\sigma})^{rss} + F_3(\mathcal{X}_{j+1,\sigma\sigma})^{rss} \\ + F_{uu}\left([\mathcal{X}_{\sigma\sigma},\mathcal{X}_{\sigma\sigma},\mathcal{X}_{\sigma\sigma}+\mathcal{X}_{00},0],[\mathcal{X}_{j-1},\mathcal{X}_j,\mathcal{X}_{j+1},\mathbf{1}_{j=0}]\right) = 0. \quad \text{(A.30)}$$

Equation (A.30) is (A.28) with the same linear operator and only the F_{uu} forcing term. Thus $(\mathcal{X}_{j\sigma\sigma})^{rss}$ is the part of $\mathcal{X}_{j\sigma\sigma}$ that comes from changing the steady state at which the first-order shock response is computed, in the direction of the risky steady state.

TVP term. The remaining term in (A.28) is the forcing $F_3\mathcal{X}_{j+1,00}$. To isolate it, define $Y_k^{tvp}(\delta)$ by

$$F\left(Y_{k-1}^{tvp}(\delta),\ Y_k^{tvp}(\delta),\ Y_{k+1}^{tvp}(\delta)+\delta\mathcal{X}_{k+1,00},\ 0\right) = 0, \qquad k \geq 0,$$

with $Y_{-1}^{tvp}(\delta) = X$ and $Y_k^{tvp}(\delta) \to X$ as $k \to \infty$. Define

$$(\mathcal{X}_{j\sigma\sigma})^{tvp} = \frac{d}{d\delta} Y_j^{tvp}(0).$$

Differentiating at $\delta = 0$ gives

$$F_1(\mathcal{X}_{j-1,\sigma\sigma})^{tvp} + F_2(\mathcal{X}_{j\sigma\sigma})^{tvp} + F_3(\mathcal{X}_{j+1,\sigma\sigma})^{tvp} + F_3\mathcal{X}_{j+1,00} = 0, \quad \text{(A.31)}$$

which is (A.28) with the same linear operator and only the $F_3\mathcal{X}_{j+1,00}$ forcing term.

Putting it together. Adding (A.30) and (A.31) gives (A.28), so

$$\mathcal{X}_{j\sigma\sigma} = (\mathcal{X}_{j\sigma\sigma})^{rss} + (\mathcal{X}_{j\sigma\sigma})^{tvp}. \quad \text{(A.32)}$$

This is the third-order certainty correspondence for the mixed $\epsilon_{t-j}\sigma^2$ term. The *rss* part comes from the fact that background risk σ^2 means that we are computing the impulse around the risky steady state rather than the deterministic steady state. The *tvp* part comes from the expectation of the mixed size and state-dependence terms $\mathcal{X}_{j+1,00}$ along a transition. Both use the same linearized perfect-foresight equations as the first-order MIT response.

A.8 Proof of propositions 5 and 6

Second-order GIRF. Isolating the terms that depend on ϵ_t in the second-order expansion, we have

$$\mathcal{X}_j\epsilon_t + \frac{1}{2}\mathcal{X}_{jj}\epsilon_t^2 + \sum_{k>j}\mathcal{X}_{jk}\epsilon_t\epsilon_{t-(k-j)}. \quad \text{(A.33)}$$

plus terms involving future shocks dated after t, whose expectations will be unchanged before and after ϵ_t is realized.

In the forecast after observing $\epsilon_t = \nu$, the terms in (A.33) become

$$\mathcal{X}_j \nu + \frac{1}{2}\mathcal{X}_{jj}\nu^2 + \sum_{k>j} \mathcal{X}_{jk}\nu\epsilon_{t-(k-j)}.$$

In the forecast before observing ϵ_t, ϵ_t has zero mean and variance σ^2, so (A.33) is $\mathcal{X}_{jj}\sigma^2/2$ in expectation. Subtracting this from (A.33) gives

$$IRF_j^{(2)}(\nu, \sigma) = \mathcal{X}_j \nu + \frac{1}{2}\mathcal{X}_{jj}(\nu^2 - \sigma^2) + \sum_{k>j} \mathcal{X}_{jk}\nu\epsilon_{t-(k-j)},$$

which is (28).

Third-order GIRF. The third-order terms add a complication relative to the second-order proof: future shocks still have mean zero, but a future shock can now enter squared and be multiplied by ϵ_t. Thus, after taking expectations over shocks dated after t, those terms can depend on the realization of ϵ_t.

Beyond the second-order terms already handled above, the terms whose expectations can change when ϵ_t is observed are

$$\begin{aligned}\frac{1}{2}\mathcal{X}_{\sigma\sigma j}\sigma^2\epsilon_t + \frac{1}{6}\mathcal{X}_{jjj}\epsilon_t^3 + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jjk}\epsilon_t^2\epsilon_{t-(k-j)} + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jkk}\epsilon_t\epsilon_{t-(k-j)}^2 \\ + \sum_{k>j}\sum_{k'>k}\mathcal{X}_{jkk'}\epsilon_t\epsilon_{t-(k-j)}\epsilon_{t-(k'-j)} + \frac{1}{2}\sum_{i<j}\mathcal{X}_{iij}\epsilon_{t+j-i}^2\epsilon_t \quad \text{(A.34)}\end{aligned}$$

where we note that the last sum is over shocks dated after t.

After observing $\epsilon_t = \nu$, the future shocks in the last term have variance σ^2, so the conditional expectation of these third-order terms is

$$\begin{aligned}\frac{1}{2}\mathcal{X}_{\sigma\sigma j}\sigma^2\nu + \frac{1}{6}\mathcal{X}_{jjj}\nu^3 + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jjk}\nu^2\epsilon_{t-(k-j)} + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jkk}\nu\epsilon_{t-(k-j)}^2 \\ + \sum_{k>j}\sum_{k'>k}\mathcal{X}_{jkk'}\nu\epsilon_{t-(k-j)}\epsilon_{t-(k'-j)} + \frac{1}{2}\sigma^2\nu\sum_{i<j}\mathcal{X}_{iij}. \quad \text{(A.35)}\end{aligned}$$

Before observing ϵ_t, symmetry gives $E_{t-1}\epsilon_t = E_{t-1}\epsilon_t^3 = 0$, while $E_{t-1}\epsilon_t^2 = \sigma^2$. Hence the only nonzero contribution from the displayed third-order terms is

$$\frac{1}{2}\sum_{k>j}\mathcal{X}_{jjk}\sigma^2\epsilon_{t-(k-j)}.$$

Subtracting this forecast from the forecast after observing $\epsilon_t = \nu$, and adding the second-order part already derived, gives

$$IRF_j^{(3)}(\nu,\sigma) = IRF_j^{(2)}(\nu,\sigma) + \frac{1}{2}\mathcal{X}_{\sigma\sigma j}\sigma^2\nu + \frac{1}{2}\sigma^2\nu\sum_{i<j}\mathcal{X}_{iij} + \frac{1}{6}\mathcal{X}_{jjj}\nu^3 + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jjk}(\nu^2-\sigma^2)\epsilon_{t-(k-j)} + \frac{1}{2}\sum_{k>j}\mathcal{X}_{jkk}\nu\epsilon^2_{t-(k-j)} + \sum_{k>j}\sum_{k'>k}\mathcal{X}_{jkk'}\nu\epsilon_{t-(k-j)}\epsilon_{t-(k'-j)}. \quad \text{(A.36)}$$

Ergodic moments. Write

$$X_t(\sigma) \equiv \mathcal{X}(\sigma, \sigma\bar{\epsilon}_t, \sigma\bar{\epsilon}_{t-1}, \ldots)$$

and evaluate all derivatives at $\sigma = 0$, holding fixed the sequence $\{\bar{\epsilon}_t, \bar{\epsilon}_{t-1}, \ldots\}$. Then

$$X_t' = \sum_i \mathcal{X}_i\bar{\epsilon}_{t-i},$$

$$X_t'' = \mathcal{X}_{\sigma\sigma} + \sum_{i,j}\mathcal{X}_{ij}\bar{\epsilon}_{t-i}\bar{\epsilon}_{t-j},$$

and

$$X_t''' = 3\sum_i \mathcal{X}_{\sigma\sigma i}\bar{\epsilon}_{t-i} + \sum_{i,j,k}\mathcal{X}_{ijk}\bar{\epsilon}_{t-i}\bar{\epsilon}_{t-j}\bar{\epsilon}_{t-k}.$$

To expand $\Sigma_h(\sigma)$, we first center $X_t(\sigma)$,

$$\tilde{X}_t(\sigma) \equiv X_t(\sigma) - \mu(\sigma), \qquad \mu(\sigma) \equiv \mathbb{E}[X_t(\sigma)],$$

so that $\Sigma_h(\sigma) = \mathbb{E}[\tilde{X}_t(\sigma)\tilde{X}_{t-h}(\sigma)^\top]$. Since $\tilde{X}_t(0) = 0$, we have

$$\tilde{X}_t(\sigma) = \sigma X_t' + \frac{\sigma^2}{2}(X_t'' - \mu'') + \frac{\sigma^3}{6}(X_t''' - \mu''') + \cdots,$$

where all derivatives are evaluated at $\sigma = 0$. Multiplying the expansions for t and $t-h$,

$$\begin{aligned}\Sigma_h(\sigma) = {} & \sigma^2\mathbb{E}[X_t'(X_{t-h}')^\top] \\ & + \frac{\sigma^3}{2}\mathbb{E}\left[X_t'(X_{t-h}''-\mu'')^\top + (X_t''-\mu'')(X_{t-h}')^\top\right] \\ & + \sigma^4\left\{\frac{1}{4}\mathbb{E}[(X_t''-\mu'')(X_{t-h}''-\mu'')^\top] + \frac{1}{6}\mathbb{E}[X_t'(X_{t-h}'''-\mu''')^\top] + \frac{1}{6}\mathbb{E}[(X_t'''-\mu''')(X_{t-h}')^\top]\right\} + \cdots.\end{aligned}$$

The coefficient on σ^3 is zero by symmetry, as is the coefficient on σ^5, so

$$\Sigma_h(\sigma) = \sigma^2\Sigma_{h,2} + \sigma^4\Sigma_{h,4} + O(\sigma^6).$$

Identifying these coefficients with terms from the expansion above, we get

$$\Sigma_{h,2} = \mathbb{E}[X_t'(X_{t-h}')^\top] = \sum_i \mathcal{X}_{h+i}\mathcal{X}_i^\top,$$

and

$$\Sigma_{h,4} = \frac{1}{4}\operatorname{Cov}(X_t'', X_{t-h}'') + \frac{1}{6}\mathbb{E}[X_t'(X_{t-h}''')^\top] + \frac{1}{6}\mathbb{E}[X_t'''(X_{t-h}')^\top].$$

The centering of X''' drops out in the last two terms because $\mathbb{E}[X_t'] = \mathbb{E}[X_{t-h}'] = 0$.

For independent normal $\bar{\epsilon}$s, Isserlis's theorem implies the identity

$$\mathbb{E}[\bar{\epsilon}_a\bar{\epsilon}_b\bar{\epsilon}_c\bar{\epsilon}_d] = \mathbf{1}_{a=b}\mathbf{1}_{c=d} + \mathbf{1}_{a=c}\mathbf{1}_{b=d} + \mathbf{1}_{a=d}\mathbf{1}_{b=c},$$

which continues to hold if the $\bar{\epsilon}$ are not normal but have the same moments as a normal up to the fourth. This, in turn, implies

$$\frac{1}{4}\operatorname{Cov}(X_t'', X_{t-h}'') = \frac{1}{2}\sum_{i,j}\mathcal{X}_{h+i,h+j}\mathcal{X}_{ij}^\top.$$

Also, since

$$X_t''' = 3\sum_i \mathcal{X}_{\sigma\sigma i}\bar{\epsilon}_{t-i} + \sum_{i,j,k}\mathcal{X}_{ijk}\bar{\epsilon}_{t-i}\bar{\epsilon}_{t-j}\bar{\epsilon}_{t-k},$$

we have

$$\frac{1}{6}\mathbb{E}[X_t'(X_{t-h}''')^\top] = \frac{1}{2}\sum_i \mathcal{X}_{h+i}\mathcal{X}_{\sigma\sigma i}^\top + \frac{1}{2}\sum_{i,j}\mathcal{X}_{h+i}\mathcal{X}_{ijj}^\top,$$

and

$$\frac{1}{6}\mathbb{E}[X_t'''(X_{t-h}')^\top] = \frac{1}{2}\sum_i \mathcal{X}_{\sigma\sigma,h+i}\mathcal{X}_i^\top + \frac{1}{2}\sum_{i,j}\mathcal{X}_{h+i,jj}\mathcal{X}_i^\top.$$

Combining these terms gives

$$\Sigma_{h,4} = \frac{1}{2}\sum_{i,j}\mathcal{X}_{h+i,h+j}\mathcal{X}_{ij}^\top + \frac{1}{2}\sum_i\left(\mathcal{X}_{h+i}\mathcal{X}_{\sigma\sigma i}^\top + \mathcal{X}_{\sigma\sigma,h+i}\mathcal{X}_i^\top\right) + \frac{1}{2}\sum_{i,j}\left(\mathcal{X}_{h+i}\mathcal{X}_{ijj}^\top + \mathcal{X}_{h+i,jj}\mathcal{X}_i^\top\right).$$

A.9 Equivalence to pruned state space and testing.

Coefficient recursions. We recall equation (4) from proposition 1:

$$\mathcal{X}(\sigma, \epsilon_t, \epsilon_{t-1}, \ldots) = g\left(\mathcal{X}(\sigma, \epsilon_{t-1}, \epsilon_{t-2}, \ldots), \epsilon_t, \sigma\right).$$

Differentiating this identity around the deterministic steady state gives us recursions that relate derivatives of the sequence-space solution $\mathcal{X}$ and the state-space solution g. For instance, differentiating (4) once with respect to ϵ_{t-j} gives $\mathcal{X}_j = g_X\mathcal{X}_{j-1} + g_\epsilon\mathbf{1}_{j=0}$, which provides a simple way to build the first-order sequence-space solution from the state-space solution. (Here, and also below, we adopt the convention that any coefficients with negative indices are zero.)

Differentiating twice with respect to σ, recalling that single σ derivatives are zero, gives an equation that can be solved for the second-order risk term:

$$\mathcal{X}_{\sigma\sigma} = g_X \mathcal{X}_{\sigma\sigma} + g_{\sigma\sigma} \qquad \Longleftrightarrow \qquad \mathcal{X}_{\sigma\sigma} = (I - g_X)^{-1} g_{\sigma\sigma}. \tag{A.37}$$

Differentiating with respect to ϵ_{t-j} and ϵ_{t-k} gives a recursion for the other second-order terms:

$$\mathcal{X}_{jk} = g_X \mathcal{X}_{j-1,k-1} + g_{XX}(\mathcal{X}_{j-1}, \mathcal{X}_{k-1}) + g_{X\epsilon}(\mathcal{X}_{j-1})\mathbf{1}_{k=0} + g_{X\epsilon}(\mathcal{X}_{k-1})\mathbf{1}_{j=0} + g_{\epsilon\epsilon}\mathbf{1}_{j=0}\mathbf{1}_{k=0}. \tag{A.38}$$

As usual, we write e.g. $g_{XX}(\mathcal{X}_{j-1}, \mathcal{X}_{k-1})$ to denote applying the vector-valued Hessian g_{XX} to $\mathcal{X}_{j-1}$ and $\mathcal{X}_{k-1}$. Here, we also write $g_{X\epsilon}(\mathcal{X}_{j-1})\mathbf{1}_{k=0}$ to denote applying $g_{X\epsilon}$ to $\mathcal{X}_{j-1}$ (for X) and the scalar $\mathbf{1}_{k=0}$ (for ϵ).

Next, for third order, taking a mixed partial with respect to ϵ_{t-j}, σ, and σ gives us a recursion for the mixed risk-shock term:

$$\mathcal{X}_{j\sigma\sigma} = g_X \mathcal{X}_{j-1,\sigma\sigma} + g_{XX}(\mathcal{X}_{j-1}, \mathcal{X}_{\sigma\sigma}) + g_{X\epsilon}(\mathcal{X}_{\sigma\sigma})\mathbf{1}_{j=0} + g_{X\sigma\sigma}(\mathcal{X}_{j-1}) + g_{\epsilon\sigma\sigma}\mathbf{1}_{j=0}. \tag{A.39}$$

Finally, to get a recursion for the general triple-shock third-order terms $\mathcal{X}_{jkl}$, we can differentiate with respect to ϵ_{t-j}, ϵ_{t-k}, and ϵ_{t-l}. This gives the following somewhat unwieldy recursion:

$$\begin{aligned}
\mathcal{X}_{jkl} = {} & g_X \mathcal{X}_{j-1,k-1,l-1} \\
& + g_{XX}(\mathcal{X}_{j-1,k-1}, \mathcal{X}_{l-1}) + g_{XX}(\mathcal{X}_{j-1,l-1}, \mathcal{X}_{k-1}) + g_{XX}(\mathcal{X}_{k-1,l-1}, \mathcal{X}_{j-1}) \\
& + g_{X\epsilon}(\mathcal{X}_{j-1,k-1})\mathbf{1}_{l=0} + g_{X\epsilon}(\mathcal{X}_{j-1,l-1})\mathbf{1}_{k=0} + g_{X\epsilon}(\mathcal{X}_{k-1,l-1})\mathbf{1}_{j=0} \\
& + g_{XXX}(\mathcal{X}_{j-1}, \mathcal{X}_{k-1}, \mathcal{X}_{l-1}) \\
& + g_{XX\epsilon}(\mathcal{X}_{j-1}, \mathcal{X}_{k-1})\mathbf{1}_{l=0} + g_{XX\epsilon}(\mathcal{X}_{j-1}, \mathcal{X}_{l-1})\mathbf{1}_{k=0} + g_{XX\epsilon}(\mathcal{X}_{k-1}, \mathcal{X}_{l-1})\mathbf{1}_{j=0} \\
& + g_{X\epsilon\epsilon}(\mathcal{X}_{j-1})\mathbf{1}_{k=0}\mathbf{1}_{l=0} + g_{X\epsilon\epsilon}(\mathcal{X}_{k-1})\mathbf{1}_{j=0}\mathbf{1}_{l=0} + g_{X\epsilon\epsilon}(\mathcal{X}_{l-1})\mathbf{1}_{j=0}\mathbf{1}_{k=0} \\
& + g_{\epsilon\epsilon\epsilon}\mathbf{1}_{j=0}\mathbf{1}_{k=0}\mathbf{1}_{l=0}.
\end{aligned} \tag{A.40}$$

Note, however, that (A.40) is long mostly to handle all the base cases where one or more of the indices is zero.

Equivalence with pruning. The coefficient recursions above become the usual pruned state-space recursions from Andreasen et al. (2018) once we evaluate the sequence-space expansions along a realized shock history. For instance, at second order, write the first- and second-order contributions to (11) as

$$X_t^{[1]} \equiv \sum_{j \geq 0} \mathcal{X}_j \epsilon_{t-j}, \qquad X_t^{[2]} \equiv \frac{1}{2}\mathcal{X}_{\sigma\sigma}\sigma^2 + \frac{1}{2}\sum_{j,k\geq 0} \mathcal{X}_{jk}\epsilon_{t-j}\epsilon_{t-k},$$

so that the full second-order approximation is $X_t^{(2)} = X + X_t^{[1]} + X_t^{[2]}$.

Summing the first-order coefficient recursion over the realized shocks gives

$$X_t^{[1]} = g_X X_{t-1}^{[1]} + g_\epsilon \epsilon_t.$$

To get the second-order equation, multiply (A.37) by $\sigma^2/2$, multiply (A.38) by $\epsilon_{t-j}\epsilon_{t-k}/2$, and sum over $j, k \geq 0$. This gives

$$X_t^{[2]} = g_X X_{t-1}^{[2]} + \frac{1}{2} g_{XX}(X_{t-1}^{[1]}, X_{t-1}^{[1]}) + g_{X\epsilon}(X_{t-1}^{[1]})\epsilon_t + \frac{1}{2} g_{\epsilon\epsilon}\epsilon_t^2 + \frac{1}{2} g_{\sigma\sigma}\sigma^2.$$

This is the standard second-order pruned state-space recursion: we calculate new second-order interactions from the first-order path, and the result is then propagated forward linearly by g_X. The same calculation at third order similarly gives the usual third-order pruning equations from the recursions for $\mathcal{X}_{j\sigma\sigma}$ and $\mathcal{X}_{jkl}$.

Hence, the nth-order sequence-space solution is exactly the same as the pruned nth-order state-space solution. This is not a new observation on our part: indeed, it is the central idea of Lombardo and Uhlig (2018), who use a related sequence-space concept as motivation for pruning in the state space.[A-1] The upshot is that in the sequence space, there is no need to add an auxiliary procedure like "pruning" on top of the perturbation solution; instead, the stability given by pruning is already built into the perturbation solution itself.

Testing certainty correspondence. Given the equivalence between the pruned state-space and sequence-space solutions, we can use a standard state-space perturbation solution of the neoclassical model as an alternative means to construct the sequence-space solution, and validate that this gives the same results as using the certainty correspondence. We have done so, finding that the two agree up to a small numerical error from differentiation.

A.10 Multiple exogenous shocks

The scalar-shock notation in section 2 extends directly to multiple independent exogenous shocks. Let $\epsilon_t = (\epsilon_{t,1}, \ldots, \epsilon_{t,m})$, with $\epsilon_{t,a} = \sigma\bar{\epsilon}_{t,a}$. The assumptions on F, local existence, and differentiability are unchanged. For the shocks, assume that each $\bar{\epsilon}_{t,a}$ is symmetric around zero, has unit variance, and is independent across dates and components:

$$\mathbb{E}\bar{\epsilon}_{t,a} = 0, \qquad \mathbb{E}\bar{\epsilon}_{t,a}^2 = 1, \qquad \{\bar{\epsilon}_{t,a}\}_{t,a} \text{ are mutually independent.}$$

Each shock derivative now carries a component index. Thus $\mathcal{X}_j^a$ is the derivative with respect to shock a dated $t-j$, $\mathcal{X}_{jk}^{ab}$ is the derivative with respect to shocks a and b dated $t-j$ and $t-k$, and similarly at higher order. With this convention, (11) and (19) keep the same structure, with sums

[A-1] Their concept differs from ours by assuming only a finite history of shocks, starting at date 0, so that the sequence-space solution is not stationary, but the logic of why sequence-space and pruned state-space perturbation are equivalent is essentially unchanged. Similar observations are made by Bhandari et al. (2023).

over shock components added to the sums over dates.

The risk terms are obtained by replacing each expected squared scalar shock by the sum over componentwise variances. In (16), the expectation perturbation $\mathcal{X}_{00}\delta$ becomes $\delta \sum_a \mathcal{X}_{00}^{aa}$, so (17) holds with this replacement. At third order, (24) holds for each component a:

$$\mathcal{X}_{\sigma\sigma j}^{a} = (\mathcal{X}_{\sigma\sigma j}^{a})^{rss} + (\mathcal{X}_{\sigma\sigma j}^{a})^{tvp}.$$

The rss term uses the same multi-shock risky steady state across all a. For the tvp term, the perturbed-expectations transition for component a uses the forcing $\delta \sum_b \mathcal{X}_{j+1,00}^{abb}$, since the future squared shock can be in any component b.

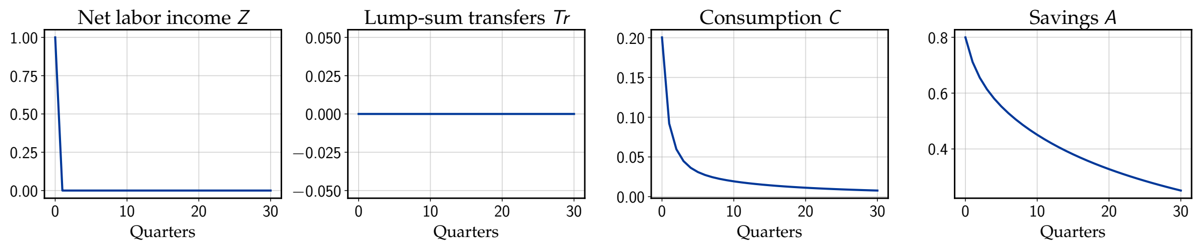
The GIRF formulas change in a similar way. If the date-t innovation is $\nu = (\nu_1, \ldots, \nu_m)$, then (28) becomes

$$IRF_j^{(2)}(\nu, \sigma) = \sum_a \mathcal{X}_j^a \nu_a + \frac{1}{2} \sum_{a,b} \mathcal{X}_{jj}^{ab} (\nu_a \nu_b - \sigma^2 \mathbf{1}_{a=b}) + \sum_{k>j} \sum_{a,b} \mathcal{X}_{jk}^{ab} \nu_a \epsilon_{t-(k-j),b}.$$

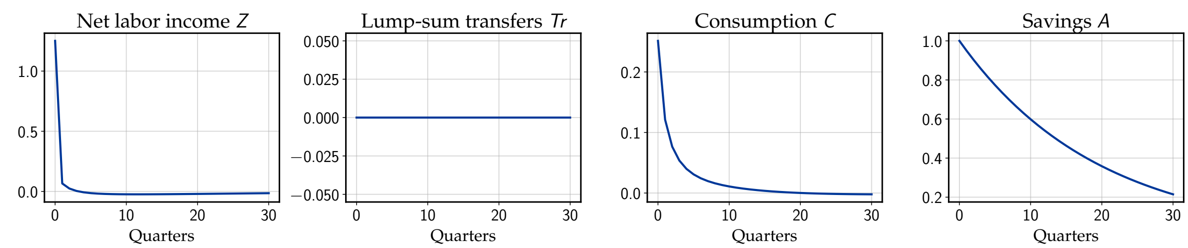
The third-order formula (29) is obtained by adding the same component indices to every shock argument. For instance, $\mathcal{X}_{\sigma\sigma j}\sigma^2\nu$ becomes $\sum_a \mathcal{X}_{\sigma\sigma j}^{a}\sigma^2\nu_a$, while $\sum_{i<j} \mathcal{X}_{iij}\sigma^2\nu$ becomes $\sum_{i<j}\sum_{a,b} \mathcal{X}_{iij}^{bba}\sigma^2\nu_a$.

The moment formulas in proposition 6 also extend by matching both dates and components. The leading covariance term becomes $\Sigma_{h,2} = \sum_{i,a} \mathcal{X}_{h+i}^{a}(\mathcal{X}_i^a)^\top$. In the fourth-order term, each repeated shock index is replaced by a repeated date-component pair: for example, $\sum_{i,j} \mathcal{X}_{h+i,h+j}\mathcal{X}_{ij}^\top$ becomes $\sum_{i,j,a,b} \mathcal{X}_{h+i,h+j}^{ab}(\mathcal{X}_{ij}^{ab})^\top$, and $\sum_{i,j} \mathcal{X}_{h+i}\mathcal{X}_{ijj}^\top$ becomes $\sum_{i,j,a,b} \mathcal{X}_{h+i}^{a}(\mathcal{X}_{ijj}^{abb})^\top$, with the transposed terms adjusted analogously. For the fourth-order term in proposition 6, we also assume the corresponding fourth-moment condition $\mathbb{E}\bar{\epsilon}_{t,a}^4 = 3$; together with independence, this gives the same fourth-moment identities used in the scalar case.

A.11 Additional quantitative results: general equilibrium IKC model



(a) Disposable income shock (PE model)



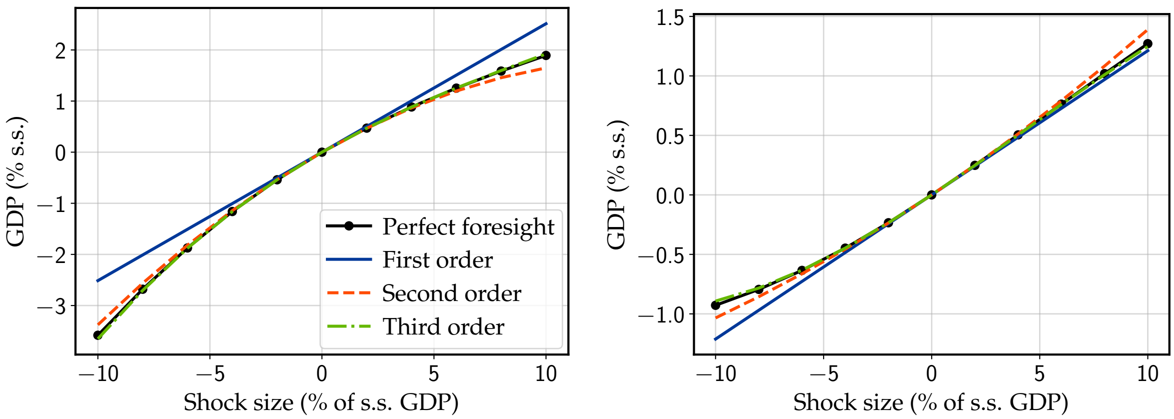
(b) Government bonds shock (GE model)

Figure A.1: First-order impulse response functions

Figure A.1 displays the shocks and resulting impulses in our heterogeneous-agent models. The top panel shows the PE model, with the shock to aggregate income Z on the left. The right panel shows GE model, with the shock to aggregate B with persistence $\rho_B = 0.95$ (which must also be aggregate assets in general equilibrium) on the right. As can be seen by comparing the top and bottom panel, the underlying shocks to net labor income Z are very similar in the two models—in particular, the impact effect on aggregate taxes T is the same. However, the income shock has present discounted value of one in the PE model, while the tax shock has zero present discounted value in the GE model, as it must to satisfy the intertemporal budget constraint.

Figures A.2–A.5 provide the counterpart of figures 2–5 in the general equilibrium IKC model (the exception is figure 6, which requires a global solution which we do not have for that model). The general patterns are very similar as those from the PE model, because the same underlying effects are operative in both.

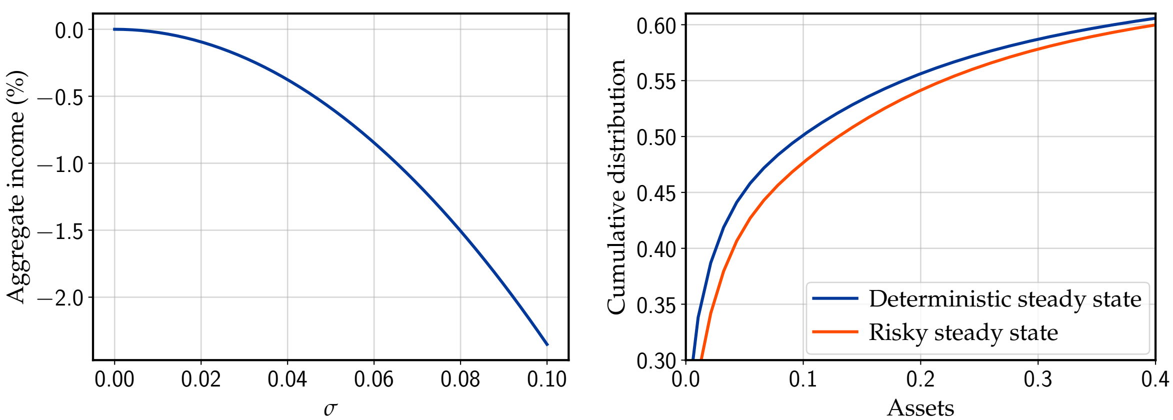
One exception is that the risky steady state in figure A.3 is now expressed in terms of an effect on aggregate income. Here, as risk increases, agents attempt to save more and consume less, but in general equilibrium monetary policy maintains the real interest rate constant at the deterministic steady state r, requiring a decline in equilibrium income to clear the goods and asset market. This effect is nontrivial (a little under 2.5% for a shock standard deviation of 0.1); this effect is mitigated in our quantitative HANK model of section 4.1, since there we have a more responsive monetary policy rule.



(a) Response at $t = 0$, GE

(b) Response at $t = 1$, GE

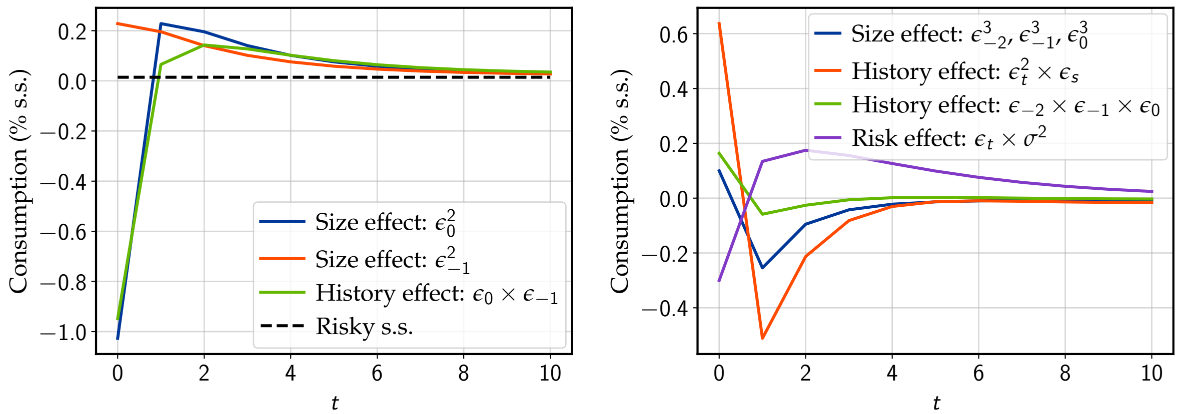
Figure A.2: Size dependence in the response of output to bond shocks (GE model)



(a) Disposable income

(b) Wealth distribution

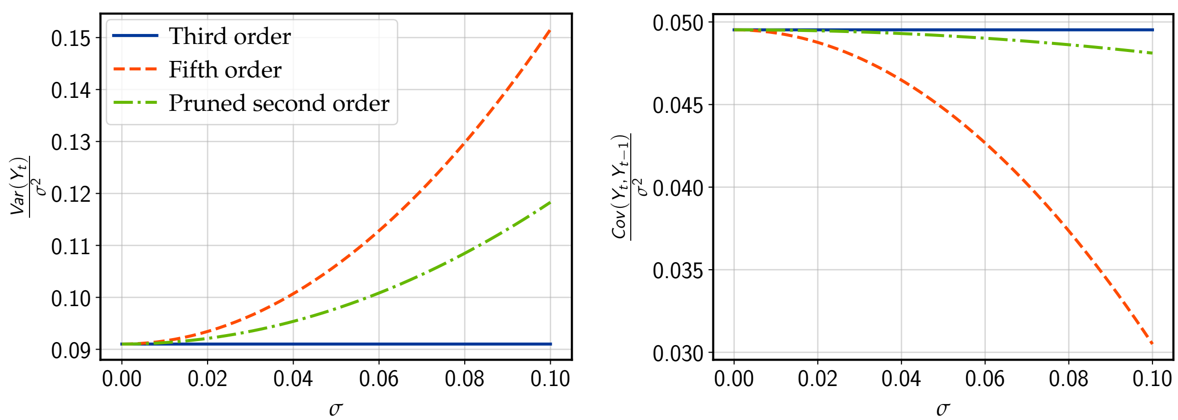
Figure A.3: The risky steady state: aggregate income and aggregate uncertainty (GE model)



(a) Second order

(b) Third order

Figure A.4: Decomposing second and third order impulse on consumption (GE model)



(a) Output variance

(b) Output first autocovariance

Figure A.5: Higher order approximation of moments in the IKC model

A.12 Proof of one-shot principle

Assume that $\mathbf{H}$ is sufficiently smooth and that $\mathbf{H}_U$ is invertible at $(\mathbf{U}_{ss}, 0)$. All derivatives of $\mathbf{H}$ below are evaluated at this point. Differentiating

$$\mathbf{H}(\mathbf{U}(\nu), \nu) = 0$$

once with respect to ν gives

$$\mathbf{H}_U \mathbf{U}'(0) + \mathbf{H}_\nu = 0,$$

and hence

$$\mathbf{U}'(0) = -\mathbf{H}_U^{-1} \mathbf{H}_\nu. \tag{A.41}$$

Differentiating a second time gives

$$\mathbf{H}_U \mathbf{U}''(0) + \mathbf{H}_{UU}\left[\mathbf{U}'(0), \mathbf{U}'(0)\right] + 2\mathbf{H}_{U\nu}\left[\mathbf{U}'(0)\right] + \mathbf{H}_{\nu\nu} = 0. \tag{A.42}$$

Therefore,

$$\mathbf{U}''(0) = -\mathbf{H}_U^{-1}\left(\mathbf{H}_{UU}\left[\mathbf{U}'(0), \mathbf{U}'(0)\right] + 2\mathbf{H}_{U\nu}\left[\mathbf{U}'(0)\right] + \mathbf{H}_{\nu\nu}\right). \tag{A.43}$$

As argued in section 3.2, this expression has a simple "one-shot principle" interpretation. Indeed, consider the the first-order approximation

$$\mathbf{U}^{[1]}(\nu) = \mathbf{U}_{ss} + \mathbf{U}'(0)\nu.$$

By definition, this satisfies the equilibrium conditions through first order. We therefore have:

$$\mathbf{H}\big(\mathbf{U}^{[1]}(\nu), \nu\big) = \frac{\nu^2}{2}\mathbf{h}^{\text{err},2} + O(\nu^3),$$

On the one hand, this implies that

$$\mathbf{h}^{\text{err},2} = \lim_{\nu \to 0} \frac{\mathbf{H}\left(\mathbf{U}_{ss} + \mathbf{U}'(0)\nu, \nu\right)}{\nu^2/2}. \tag{A.44}$$

on the other, a Taylor expansion implies that

$$\mathbf{h}^{\text{err},2} = \mathbf{H}_{UU}\left[\mathbf{U}'(0), \mathbf{U}'(0)\right] + 2\mathbf{H}_{U\nu}\left[\mathbf{U}'(0)\right] + \mathbf{H}_{\nu\nu}.$$

It follows that

$$\mathbf{U}''(0) = -\mathbf{H}_U^{-1}\mathbf{h}^{\text{err},2}.$$

Thus, after evaluating the nonlinear equilibrium residual along the first-order path, a single application of the inverse steady-state Jacobian re-equilibrates the system through second order.

The same principle applies at third order. Differentiating the equilibrium conditions a third time gives

$$\begin{aligned} 0 = {} & \mathbf{H}_U \mathbf{U}'''(0) + \mathbf{H}_{UUU}\left[\mathbf{U}'(0), \mathbf{U}'(0), \mathbf{U}'(0)\right] \\ & + 3\mathbf{H}_{UU}\left[\mathbf{U}'(0), \mathbf{U}''(0)\right] + 3\mathbf{H}_{UU\nu}\left[\mathbf{U}'(0), \mathbf{U}'(0)\right] \\ & + 3\mathbf{H}_{U\nu}\left[\mathbf{U}''(0)\right] + 3\mathbf{H}_{U\nu\nu}\left[\mathbf{U}'(0)\right] + \mathbf{H}_{\nu\nu\nu}. \end{aligned} \tag{A.45}$$

Consequently,

$$\begin{aligned} \mathbf{U}'''(0) = {} & -\mathbf{H}_U^{-1}\Big(\mathbf{H}_{UUU}\left[\mathbf{U}'(0), \mathbf{U}'(0), \mathbf{U}'(0)\right] \\ & + 3\mathbf{H}_{UU}\left[\mathbf{U}'(0), \mathbf{U}''(0)\right] + 3\mathbf{H}_{UU\nu}\left[\mathbf{U}'(0), \mathbf{U}'(0)\right] \\ & + 3\mathbf{H}_{U\nu}\left[\mathbf{U}''(0)\right] + 3\mathbf{H}_{U\nu\nu}\left[\mathbf{U}'(0)\right] + \mathbf{H}_{\nu\nu\nu}\Big). \end{aligned} \tag{A.46}$$

To obtain the one-shot representation, consider the second-order path

$$\mathbf{U}^{[2]}(\nu) = \mathbf{U}_{ss} + \mathbf{U}'(0)\nu + \frac{1}{2}\mathbf{U}''(0)\nu^2. \tag{A.47}$$

Since this path satisfies the equilibrium conditions through second order,

$$\mathbf{H}\big(\mathbf{U}^{[2]}(\nu), \nu\big) = \frac{\nu^3}{6}\mathbf{h}^{\text{err},3} + O(\nu^4),$$

where

$$\mathbf{h}^{\text{err},3} = \lim_{\nu \to 0} \frac{\mathbf{H}\left(\mathbf{U}_{ss} + \mathbf{U}'(0)\nu + \frac{1}{2}\mathbf{U}''(0)\nu^2, \nu\right)}{\nu^3/6}. \tag{A.48}$$

In terms of the derivatives of $\mathbf{H}$,

$$\begin{aligned}\mathbf{h}^{\text{err},3} = {} & \mathbf{H}_{UUU}\big[\mathbf{U}'(0), \mathbf{U}'(0), \mathbf{U}'(0)\big] + 3\mathbf{H}_{UU}\big[\mathbf{U}'(0), \mathbf{U}''(0)\big] \\ & + 3\mathbf{H}_{UU\nu}\big[\mathbf{U}'(0), \mathbf{U}'(0)\big] + 3\mathbf{H}_{U\nu}\big[\mathbf{U}''(0)\big] \\ & + 3\mathbf{H}_{U\nu\nu}\big[\mathbf{U}'(0)\big] + \mathbf{H}_{\nu\nu\nu}.\end{aligned} \tag{A.49}$$

The third-order correction is therefore

$$\mathbf{U}'''(0) = -\mathbf{H}_U^{-1}\mathbf{h}^{\text{err},3}. \tag{A.50}$$

Hence the third-order solution also requires only an evaluation of the nonlinear residual along the previously computed path, followed by one application of the same inverse steady-state Jacobian.

B Smooth Standard Incomplete Markets Model

B.1 Kinks and higher-order differentiation: the problem with non-smooth models

B.1.1 Discussion

To motivate our smooth version of the standard incomplete markets model, we first discuss why the standard incomplete markets model can be problematic for higher-order differentiation without smoothing. The central issue is that there is a kink in the policy function at the asset level where the borrowing constraint becomes binding. At this kink, the second derivative of the consumption function is infinite. Furthermore, when income risk is discrete—including when a continuous income process is discretized via Rouwenhorst or a similar procedure—this kink leads to *secondary kinks* at other asset levels. These occur at the asset levels where, if a certain discrete income path is realized, the borrowing constraint will become exactly binding in $n > 1$ periods.[A-2]

[A-2] Even in discretizations with a decent number of points, these secondary kinks are often quite pronounced, because there is still a high probability of remaining at the same income state for the next several periods. One benefit of continuous-time solutions is that they avoid these secondary kinks—although there is still an important kink at the constraint. Also see Iskhakov, Jørgensen, Rust and Schjerning (2017) for more discussion of secondary kinks, which in

Much of the overall nonlinearity of the consumption function is concentrated at these kinks, which makes differentiation potentially problematic: the second derivative at these kinks is infinite, while the second derivative away from these kinks will understate the overall curvature of the function. If we integrate over a smooth, continuous distribution of households, however, the resulting aggregate consumption (as a function, say, of a lump-sum transfer) is nevertheless well-behaved, and its higher derivatives will reflect the nonlinearity from kinks.

Unfortunately, typical calibrations of the standard incomplete markets model, with a discrete Markov chain representing idiosyncratic income, do not have such a distribution over assets. Indeed, their stationary distributions consist of countably many mass points: with probability 1, every household has hit the borrowing constraint at some previous date, and households' current asset levels are then determined by the finite history of income states received since the most recent visit to the constraint.[A-3] In practice, a large fraction of the distribution often concentrates at a limited number of mass points—including the large mass point at the constraint itself.[A-4] In implementation, this is often made worse by the standard "lottery" method of discretizing the distribution (Young, 2010), which constrains the distribution to consist entirely of mass points.[A-5] This makes higher derivatives of aggregate consumption unreliable. In the unlikely event that a mass of households is at a kink point of the policy function, then these derivatives will be infinite; otherwise, they will understate the curvature of aggregate consumption.

In appendix B.2, we will show how to fix these pathologies in practice with a "smooth model", which includes an smoothly distributed iid component of income, and where we carefully perform numerical integration with respect to this component separately on both sides of the primary kink. This approach eliminates secondary kinks, and ensures a smooth distribution of cash on hand, resulting in a smooth aggregate consumption function that is amenable to perturbation.[A-6]

For the remainder of this section, we will illustrate these concepts with a stark, analytically tractable model.

B.1.2 Analytical illustration

Consider a household with log utility, discount factor $\beta < 1$, no income risk, income normalized to one, no interest, and a borrowing constraint. Given incoming assets a_{t-1}, it chooses consumption c_t and assets a_t carried into the next period:

$$\max_{\{c_t,a_t\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \log c_t \qquad \text{subject to} \qquad c_t + a_t = 1 + a_{t-1}, \quad a_t \geq 0, \quad a_{-1} = a.$$

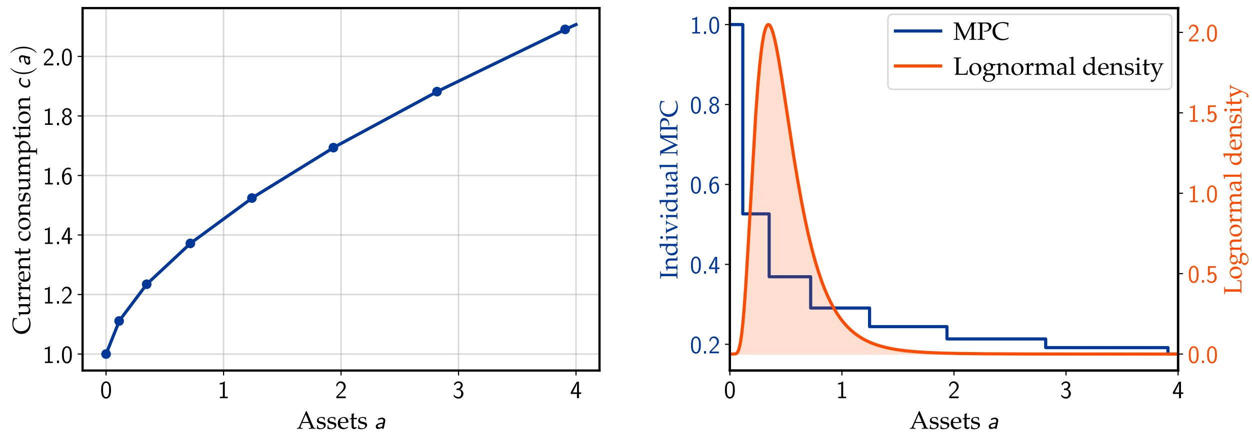
their context are kinks in the value function rather than the policy function.

[A-3] Indeed, Le Grand and Ragot (2022) have used this insight as part of a computational strategy.

[A-4] See some discussion and visualization of mass points away from the constraint in the Lecture 1 notebook at `https://github.com/shade-econ/nber-workshop-2023/`.

[A-5] See Bhandari et al. (2023) for a critique of this method in the context of higher-order perturbation, as well as Bilal (2023). Bayer et al. (2026b) offer an alternative approach to iterating on the distribution that is better-suited to higher-order perturbation.

[A-6] For an example of using heterogeneity to smooth out a kink and facilitate perturbation in a very different context, see Schmitt-Grohé and Uribe (2022)'s model of downward nominal wage rigidity.



(a) Individual consumption

(b) Individual MPC and asset density

Figure B.1: Individual consumption function and marginal propensity to consume

With a gross interest rate of one, the interior Euler equation is $c_{t+1} = \beta c_t$. The household is impatient relative to the interest rate, so it eventually runs down its assets and reaches the constraint.[A-7]

Suppose a household with initial assets a first reaches the constraint after n periods. Consumption follows a path $c_t = \beta^t c_0$ for $t = 0, \ldots, n-1$, and all initial assets have been spent by the end of these n periods. Defining

$$S_n = 1 + \beta + \cdots + \beta^{n-1},$$

the corresponding budget constraint is $S_n c_0 = a + n$. Current consumption is therefore

$$c(a) = \frac{a+n}{S_n}, \qquad a \in [a^{n-1}, a^n].$$

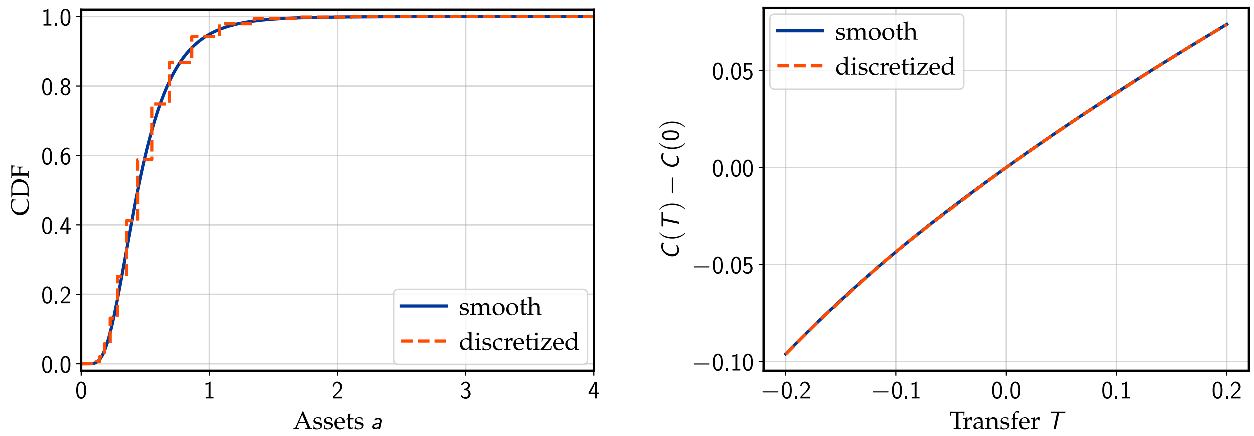
The kink point a^n is the boundary at which the household arrives at zero assets after n periods and its consumption path implies $c_n = 1$, exactly equal to next period's income. Thus $c_0 = \beta^{-n}$, which together with the budget constraint gives

$$a^n = \frac{\beta^{-n} - 1}{1 - \beta} - n.$$

Consumption is continuous and piecewise linear, with MPC $1/S_n$ on the nth segment. Note in this example, the *entire* nonlinearity of the consumption function is at the kink points.

Figure B.1(a) plots this consumption function for our assumed $\beta = 0.9$, with solid dots denoting the kink points. For the aggregation below, we will further suppose that initial assets are lognormally distributed with mean 0.5 and log standard deviation 0.5; panel (b) overlays this density on the individual MPC schedule.

[A-7]Note this model is not stationary, so its derivatives are not derivatives around a steady state as in most of the paper.



(a) Asset CDFs

(b) Aggregate consumption $C(T)$

Figure B.2: Asset distributions and implied aggregate consumption

Aggregation and discretization. An aggregate transfer T shifts each household from assets a to $a+T$. If f is the smooth lognormal asset density, aggregate current consumption is

$$C(T) = \int_0^\infty c(a+T) f(a)\, da.$$

We compare this distribution in figure B.2 with a 21-point Rouwenhorst discretization that matches mean assets and the standard deviation of log assets. Note that the nonlinear consumption functions $C(T)$ implied by the smooth density and discretization are visually indistinguishable in panel (b).

Higher-order numerical differentiation. Let F and f denote the CDF and density of assets. At kink a^q, the individual MPC changes by

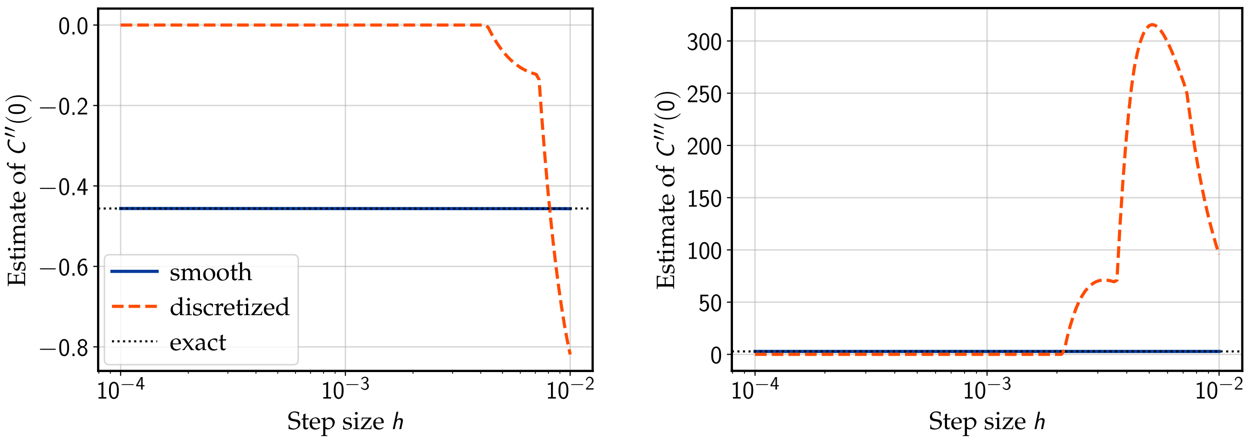
$$\Delta m_q = \frac{1}{S_{q+1}} - \frac{1}{S_q}.$$

The aggregate derivatives at $T = 0$ are therefore

$$C''(0) = \sum_{q \geq 1} \Delta m_q f(a^q), \qquad C'''(0) = -\sum_{q \geq 1} \Delta m_q f'(a^q).$$

which we can calculate directly using the smooth density f.

Figure B.3 compares these exact derivatives with derivatives obtained using numerical differentiation of $C(T)$, implemented using centered finite differences over various step sizes $h \in [10^{-4}, 10^{-2}]$, when $C(T)$ is calculated using either the smooth density or its discretization. Although these two versions of $C(T)$ were visually indistinguishable in figure B.2(b), their higher-order numerical derivatives are not: the derivatives for the discretized version of $C(T)$ are unreli-



(a) Second derivative $C''(0)$

(b) Third derivative $C'''(0)$

Figure B.3: Numerical estimates of higher derivatives of aggregate consumption

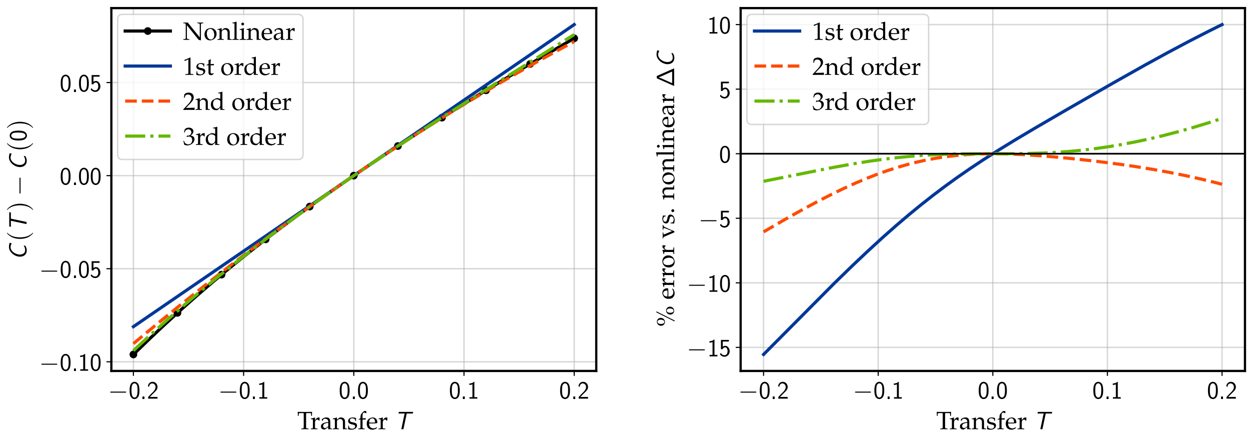
able and wildly erratic, while the derivatives for the smooth $C(T)$ agree with the exact computation.

The reason is closely related to our earlier discussion: when $C(T)$ is calculated using a discretized distribution, its higher-order numerical derivatives at 0 depend on whether any of the mass points a are close to a kink. If they are not at a kink, so that the finite differences (i.e. $a + h$ and $a - h$ for second order) do not cross any kinks for small h, the estimated derivatives will be exactly zero. This is indeed the case for small h in both panels of figure B.3. But once h becomes sufficiently large, then the finite differences start to cross kinks, at which higher derivatives are locally infinite. This results in erratic estimates for $C''(0)$ and $C'''(0)$, with the latter being especially extreme.

Taylor approximation. Figure B.4(a) compares first-, second-, and third-order Taylor approximations around $T = 0$ with the nonlinear change in consumption over transfers from -0.2 to 0.2. Panel (b) reports the corresponding errors relative to the nonlinear consumption change $\Delta C = C(T) - C(0)$. The first-order approximation is visibly inaccurate for larger transfers, with errors reaching roughly 10–15% at the endpoints. The second-order approximation captures most of the curvature, while the third-order approximation keeps the absolute error below about 3% throughout the range, although errors begin to grow near the boundaries.[A-8]

Average consumption over several periods. The preceding results concern current consumption. We can also ask how well the Taylor approximations perform for *cumulative* consumption

[A-8]Interestingly, the Taylor series around $T = 0$ has radius of convergence of only $a^1 = \beta^{-1} - 1 \simeq 0.11$. This is because at $T = a^1$, the first consumption kink reaches the lower endpoint of the asset distribution (0), where the lognormal density is smooth but not analytic. An infinite series therefore diverges for $|T| > a^1$, but the accuracy shown here illustrates that finite Taylor approximations can still work well beyond this radius when the relevant lower-order derivatives are sufficiently well-behaved.



(a) Nonlinear response and approximations

(b) Approximation errors

Figure B.4: Taylor approximations to aggregate consumption

over several periods. For a household starting with assets a, define cumulative and average consumption over the next H periods as

$$G_H(a) = \sum_{t=0}^{H-1} c_t(a), \qquad \bar{G}_H(a) = \frac{G_H(a)}{H}.$$

If the household reaches the constraint within the horizon, it spends all of its initial assets during these H periods. Cumulative consumption is then initial assets plus total income, so $G_H(a) = a + H$, regardless of exactly when the constraint is reached. If $n > H$, the household remains unconstrained throughout the horizon, so consumption continues to satisfy $c_t = \beta^t c_0$ and

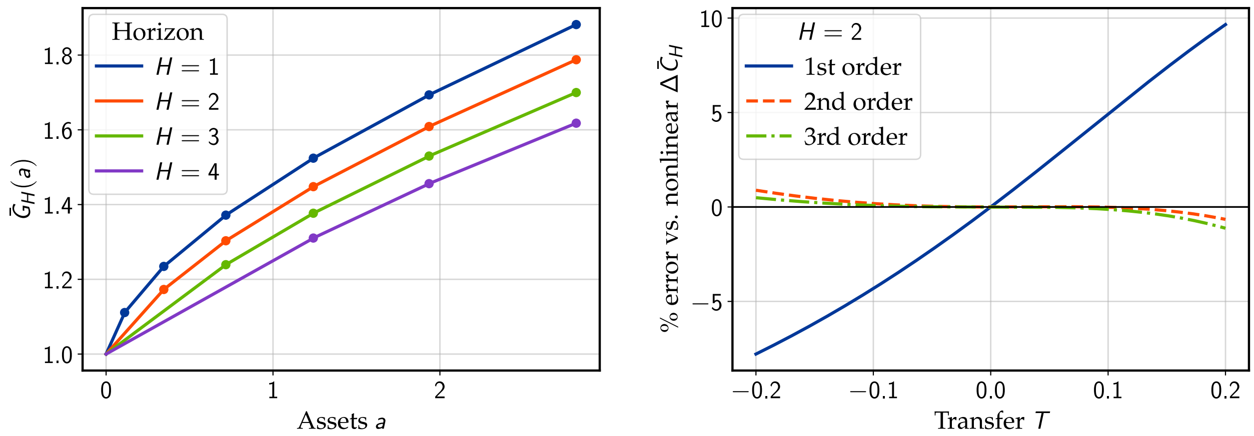
$$G_H(a) = S_H \frac{a+n}{S_n}, \qquad a \in [a^{n-1}, a^n].$$

Hence, the segments on which the constraint is reached within the horizon all collapse to the same line, with slope $1/H$ for average consumption. The kinks at $a^1, \ldots, a^{H-1}$ disappear from $\bar{G}_H$, and its first remaining kink is at a^H. Figure B.5(a) shows how increasing the horizon successively removes these low-asset kinks.

Finally, define aggregate average consumption as

$$\bar{C}_H(T) = \int_0^\infty \bar{G}_H(a+T) f(a)\, da.$$

Figure B.5(b) repeats the Taylor-approximation exercise for $H = 2$, which is average consumption over the first two periods rather than one. The linear approximation is better but remains noticeably inaccurate; the second- and third-order errors, however, are now around 1% or less over nearly the entire range of transfers. Intuitively, as the kink points near the lognormal distribution's extreme curvature just above zero (figure B.1(b)) are removed, aggregate consumption becomes



(a) Individual averages by horizon

(b) Approximation errors for $H = 2$

Figure B.5: Consumption averaged over multiple periods

much better-behaved: although it is still nonlinear, it is regular enough to be well-captured by a second- or third-order Taylor expansion.[A-9] This helps explain why, in appendix C, we find that perturbation does even better for cumulative outcomes over a few periods.

Taking stock. Kinks in the individual consumption function, which emerge naturally in models with a borrowing constraint, do not by themselves rule out higher-order approximation: when households are smoothly distributed, aggregation produces a smooth consumption function whose derivatives capture households crossing the kinks. The problem is the combination of kinked policies and a distribution with mass points. As figures B.2 and B.3 show, a discretization can match aggregate consumption almost exactly in levels, while producing unreliable higher derivatives. By contrast, with a smooth distribution, second- and third-order approximations can capture substantial nonlinearities, especially for cumulative outcomes over which the most important individual kinks disappear.

B.2 Smooth income shocks: solution and implementation

We now describe the smooth standard incomplete markets model used in the paper. The key addition to the usual model is a continuously distributed i.i.d. component of income. This component is realized and integrated out within the period, so it smooths the distribution of cash on hand without adding a state variable. It thereby eliminates the mass points in cash on hand that caused the difficulty in Appendix B.1, as well as the secondary kinks in the policy function: consumption is smooth everywhere except where the borrowing constraint starts to bind. Code and slides for the implementation of the model, although the notation differs, is available at lecture 9 of `https://github.com/shade-econ/nber-workshop-2025`.

[A-9]This improved approximation can be understood in part via the observation in footnote A-8: averaging over H periods expands the radius of convergence from a^1 to a^H.

Household problem and timing. Let $s = (e, \beta)$ denote the household's discrete state and let $\Pi_{ss'}$ be its transition matrix. At the beginning of period t, the household enters with assets a, observes s and the transitory shock ξ, and then chooses consumption c and assets a'. In the notation of equation (6), the Bellman equation is

$$V_t(a,s) = \mathbb{E}_\xi \left[\max_{a' \geq \underline{a}} \left\{ u(c) + \beta \sum_{s'} \Pi_{ss'} \mathbb{E}_t V_{t+1}(a', s') \right\} \right], \tag{A.51}$$
$$c + a' = (1 + r_t)a + Z_t e \xi + Tr_t.$$

We retain time subscripts on r_t and Tr_t because both vary in some of the exercises in the paper. To be consistent with equation (7) in the main text, however, we suppress them as arguments of the household maps below. In the baseline calibration, $u(c) = \log c$, $\underline{a} = 0$, and

$$\log \xi \sim N(-\sigma_\xi^2/2, \sigma_\xi^2), \qquad \mathbb{E}\xi = 1.$$

We solve the household problem in three backward stages. This decomposition also makes explicit the construction of the map $\mathcal{V}(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1})$ in the stacked household block in (7).

1. Continuation values. We iterate on the marginal value of incoming assets, $v_t(a,s) \equiv V_{a,t}(a,s)$. Let $w_t(a',s)$ denote the discounted expected continuation value, whose derivative is $w_{a,t}(a',s) = \beta \sum_{s'} \Pi_{ss'} \mathbb{E}_t v_{t+1}(a', s')$.

2. Endogenous-grid solution conditional on cash on hand. Write cash on hand as

$$m = (1 + r_t)a + Z_t e \xi + Tr_t.$$

Conditional on s and m, the household solves

$$\max_{a' \geq \underline{a}} \{ u(m - a') + w_t(a', s) \}.$$

At each asset gridpoint a_i,[A-10] the interior first-order condition gives

$$c_{s,i} = (u')^{-1}(w_{a,t}(a_i, s)), \qquad m_{s,i} = a_i + c_{s,i}. \tag{A.52}$$

For CRRA utility with coefficient γ, this is $c_{s,i} = w_{a,t}(a_i, s)^{-1/\gamma}$. Thus the endogenous gridpoint method of Carroll (2006) directly supplies observations $(m_{s,i}, c_{s,i})$. We interpolate these observations with a cubic spline. If $m_s^* \equiv m_{s,0}$ is the first endogenous gridpoint, the complete policy

[A-10] The current implementation uses 200 double-exponentially spaced asset gridpoints on $[0, 1000]$. If u_i is uniformly spaced from zero to $\log[1 + \log(1 + 1000)]$, then $a_i = \exp(\exp(u_i) - 1) - 1$.

is

$$c_t(m,s) = \begin{cases} m - \underline{a}, & m \le m_s^*, \\ \widetilde{c}_t(m,s), & m > m_s^*, \end{cases} \tag{A.53}$$

where $\widetilde{c}_t$ is the spline. Hence, for each discrete state s, the consumption policy has a single kink, where the borrowing constraint ceases to bind.

3. Integration over the transitory shock. The envelope condition returns the marginal value at the original asset gridpoints:

$$v_t(a,s) = (1+r_t)\mathbb{E}_\xi\left[u'(c_t((1+r_t)a + Z_t e\xi + Tr_t, s))\right]. \tag{A.54}$$

This expectation must be evaluated carefully because there is a kink at m_s^*. Let $d_t(a) = (1+r_t)a + Tr_t$, $x = \log(Z_t e\xi)$, and $\mu_s = \log(Z_t e) - \sigma_\xi^2/2$. When $d_t(a) < m_s^*$, set $x_s^*(a) = \log(m_s^* - d_t(a))$. Then

$$\begin{aligned} \frac{v_t(a,s)}{1+r_t} = & \int_{-\infty}^{x_s^*(a)} u'(d_t(a) + e^x - \underline{a})\, p_s(x)\, dx \\ & + \int_{x_s^*(a)}^{\infty} u'(\widetilde{c}_t(d_t(a) + e^x, s))\, p_s(x)\, dx, \end{aligned} \tag{A.55}$$

where p_s is the density of a normal random variable with mean μ_s and standard deviation σ_ξ. If $d_t(a) \ge m_s^*$, only the second integral is needed. Numerically, we apply a fixed, high-order Gauss-Legendre rule separately to each nonempty interval. Splitting at the kink is important: a single quadrature rule whose nodes move across the kink is much less accurate and can lead to erratic numerical derivatives, similar to those we will see in the non-smooth model in appendix B.3.[A-11]

Forward iteration of the distribution. The smooth shock also permits a continuous-distribution implementation of the forward step. For each discrete state, we represent the conditional CDF of assets on the same asset grid used for the backward iteration. The CDF may have a mass point at $\underline{a}$, but is smooth above the constraint. We interpolate the stored CDF values with a cubic spline whenever the CDF is evaluated between gridpoints. Thus the vector $\boldsymbol{\phi}_t$ in the main text may equivalently be viewed as CDF values at the knots or as coefficients of the cubic spline representation; the two are related by a fixed linear transformation.

To see the forward step, hold the current state s fixed and let $\Phi_{t-1,s}(a) \equiv \Pr(a_{t-1} \le a \mid s_t = s)$ denote the CDF of incoming assets. The CDF of cash on hand is the convolution

$$\Phi_{t,s}^{\text{coh}}(m) = \int \Phi_{t-1,s}\left(\frac{m - Tr_t - Z_t e\xi}{1+r_t}\right) f_\xi(\xi)\, d\xi, \tag{A.56}$$

[A-11] Concretely, we truncate at $\mu_s \pm 8\sigma_\xi$ and use 40 Gauss–Legendre nodes on each nonempty side. In cases where we are far from the kink, we could also use Gauss-Hermite with a smaller number of nodes, but we stick to Gauss-Legendre for simplicity.

where $\Phi_{t-1,s}$ is understood to be zero below the borrowing limit. Since the saving policy is monotone and equation (A.52) maps $m_{s,i}$ into a_i, no inversion or lottery is required:[A-12]

$$\widehat{\Phi}_{t,s}(a_i) = \Phi^{\text{coh}}_{t,s}(m_{s,i}). \tag{A.57}$$

In particular, $\widehat{\Phi}_{t,s}(\underline{a}) = \Phi^{\text{coh}}_{t,s}(m^*_s)$ records the mass at the constraint. Equation (A.56) is evaluated with the same quadrature rule in log transitory income.

The last forward stage applies the transition of the discrete state. If ϖ is the stationary distribution of s, the reverse conditional transition matrix is

$$\Pi^{\text{rev}}_{s's} = \Pr(s_t = s \mid s_{t+1} = s') = \frac{\varpi_s \Pi_{ss'}}{\varpi_{s'}}. \tag{A.58}$$

The CDF carried into the next period, conditional on s', is therefore $\Phi_{t,s'} = \sum_s \Pi^{\text{rev}}_{s's}\widehat{\Phi}_{t,s}$. The combination of equations (A.56)–(A.58) is the map $\Lambda(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1})$ in (7).[A-13]

Aggregate consumption can be computed by integrating the policy over incoming assets, current discrete states, and ξ, as described in appendix A.4. In the implementation we use the equivalent aggregate budget identity

$$C_t = Z_t + (1 + r_t)A_{t-1} + Tr_t - A_t, \qquad A_t = \sum_s \varpi_s \int_{\underline{a}}^{\bar{a}} (1 - \Phi_{t,s}(a))\, da, \tag{A.59}$$

where $\bar{a}$ is the upper endpoint of the numerical asset grid and the baseline borrowing limit is $\underline{a} = 0$. The survival functions are integrated as cubic splines. This defines $\mathcal{C}(Z_t, \mathbb{E}_t \boldsymbol{v}_{t+1}, \boldsymbol{\phi}_{t-1})$. Although the individual policy retains a kink and the asset distribution an atom at the borrowing constraint, integration over the continuous ξ shock makes $\mathcal{V}$, Λ, and $\mathcal{C}$ smooth functions of their aggregate arguments.

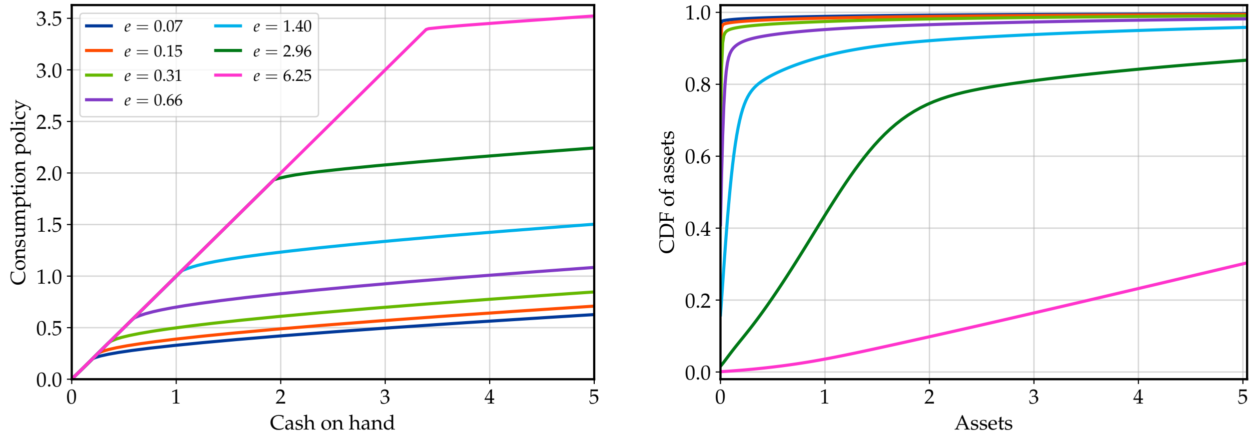
All parameter values follow the calibration reported in appendix A.4. We discretize persistent income using the Rouwenhorst method (Kopecky and Suen, 2010) and compute the steady state by standard backward and forward iteration until convergence.[A-14]

Figure B.6 shows the two objects that result from the backward and forward iterations. To make the kink most visible, both panels condition on the lower discount-factor type. The consumption function has one kink for each persistent-income state and is smooth thereafter. The conditional asset CDFs have an atom at zero but no other jumps.

[A-12] This has previously been used as the key insight behind the distributional EGM method of Bayer et al. (2026b).

[A-13] To avoid instability in the spline near the boundary, our current implementation also smoothly forces the CDF to one near the upper end of the asset grid, which we choose so that it is far from much probability mass.

[A-14] The convergence criteria are a maximum change below 10^{-9} in the endogenous cash-on-hand grid and below 10^{-11} in the conditional CDFs.



(a) Consumption policy

(b) Conditional asset CDFs

Figure B.6: Policies and conditional distributions in the smooth SIM model

B.3 Smooth vs. non-smooth model

We next compare the smooth model with a conventional finite-state approximation and non-smooth implementation of an otherwise identical model.

Discrete non-smooth benchmark. We now approximate the i.i.d. shock ξ with a five-state, zero-persistence Rouwenhorst process, following the same construction used for the persistent component of income (Kopecky and Suen, 2010). The process is normalized to match the mean and level variance of the continuous lognormal.[A-15]

We then solve the non-smooth model using the endogenous-grid method of Carroll (2006) with linear interpolation and the lottery method of Young (2010), following our typical approach (e.g., Auclert et al., 2021). Holding the smooth model's discount factors fixed gives $A = 4.01314$ and an impact income MPC of 0.19889, compared with targets of 4 and 0.2. These small gaps suggest that the two models are very close in levels and first-order behavior.[A-16]

Numerical derivatives. The distinction becomes sharp at higher order. Let $C_0(z)$ denote impact consumption when $Z_0 = Z_{ss} + z$, with subsequent Z_t and transfers held at their baseline values. For each step size h, we estimate its first three derivatives at $z = 0$ using centered finite differences. Figure B.7 reports these estimates over $h \in [10^{-4}, 10^{-2}]$, holding the smooth discount factors fixed in both models. The first derivatives are stable and nearly identical. In the smooth model, both the second and third derivatives are stable around $h = 10^{-3}$, where their estimated values are approximately -0.963 and 4.11. In the non-smooth model, the second derivative varies substantially

[A-15] A five-point Gauss–Hermite approximation, treating its quadrature nodes as discrete i.i.d. shock states, produces the same qualitative results as in figure B.7: the first numerical derivative is stable, while the second and especially the third vary sharply with the finite-difference step size.

[A-16] Recalibrating the non-smooth model to the same targets instead gives $(\beta_L^{\text{disc}}, \beta_H^{\text{disc}}) = (0.9597140, 0.9904922)$, changes of -0.0002504 and -0.0000026 relative to the smooth calibration.

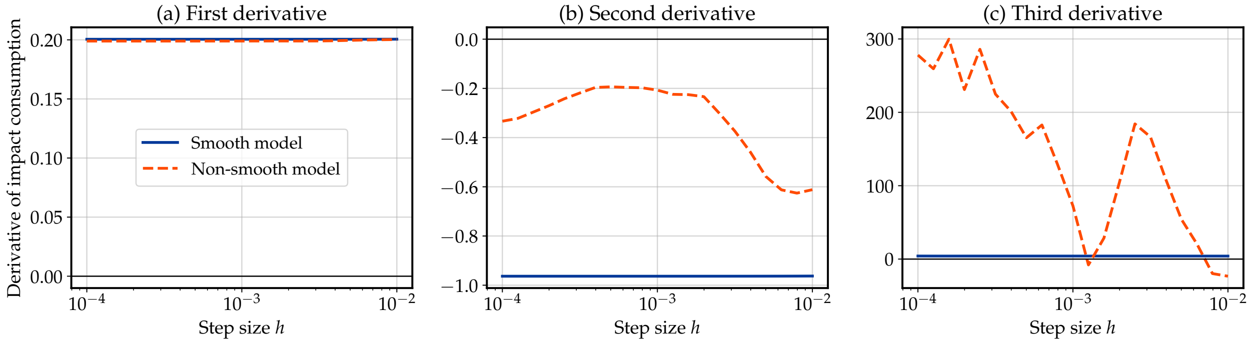


Figure B.7: Numerical derivatives of impact consumption with respect to a one-time income Z shock

with h, while the third derivative is erratic and becomes very large at small step sizes.

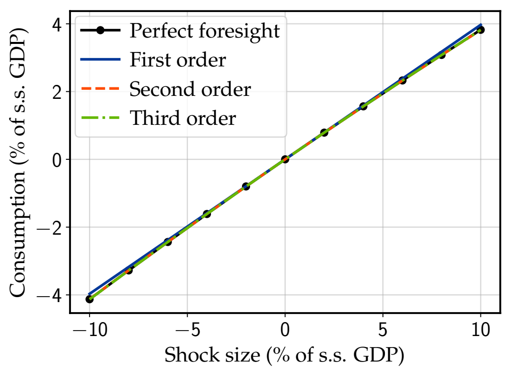
The mechanism is the one described in Appendix B.1: mass points in the non-smooth distribution interact with kinks in the individual policy. The problem is more severe here because linear interpolation creates many closely spaced kinks throughout the policy function. As a result, the estimated third derivative is unstable and remains far from the smooth benchmark throughout the range of h shown in Figure B.7.

C Quantifying nonlinearities

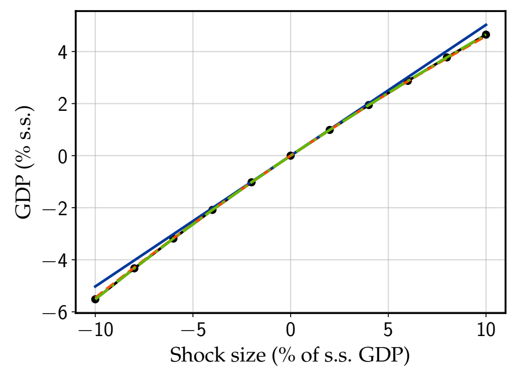
Figures 1, 2 and A.2 show that third order very accurately captures the nonlinear solution in our model, for aggregate income shocks of up to $\pm10\%$ of GDP, in both partial and general equilibrium. Bianchi and Kaplan (2026) instead argue that the third order approximation works poorly: in their words, "local methods of any order are unlikely to be reliable in settings where a failure of Ricardian equivalence from high MPCs is important." Here, we show that a third order solution continues to work well for many variants of our model—all featuring lack of Ricardian equivalence from high MPCs—and also explore the types of situations in which it works less well.

Smooth vs unsmooth model. As discussed in section B, doing a third order approximation correctly requires a smooth underlying model, which Bianchi and Kaplan (2026) do not have. We therefore focus on our smooth model. Appendix B discusses at length the types of problems one encounters with perturbation when using a non-smooth model so we do not discuss this issue further here.

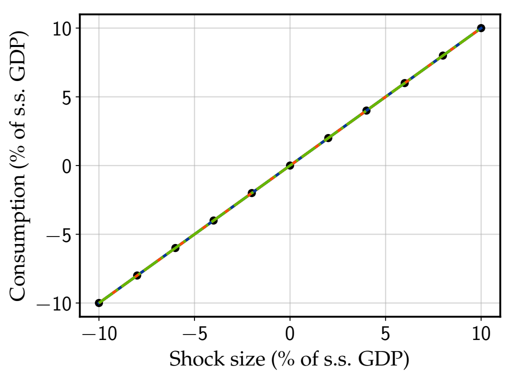
Impact vs cumulative multipliers. As figures 2 and A.2 clearly illustrate, the size effect of income on consumption is concave on impact but convex later on, irrespective of the nature of the shock. As discussed in section 2.4, this is due to the intertemporal budget constraint, which holds nonlinearly. Hence, concavity is largest when we look at the impact effect: when we consider the cumulative effect on consumption from an income shock over longer horizons, we should expect the nonlinearity to progressively disappear.



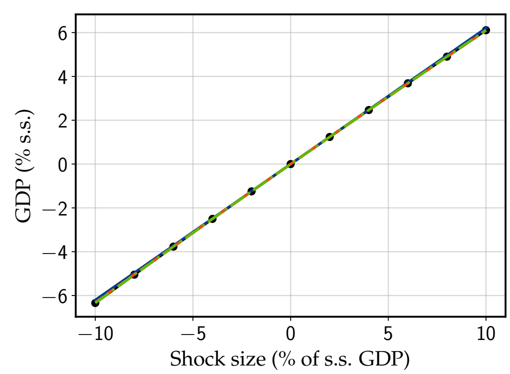
(a) 4-quarter cumulative, PE



(b) 4-quarter cumulative, GE



(c) Infinite horizon cumulative, PE



(d) Infinite horizon cumulative, GE

Figure C.1: Cumulative impulse response of consumption

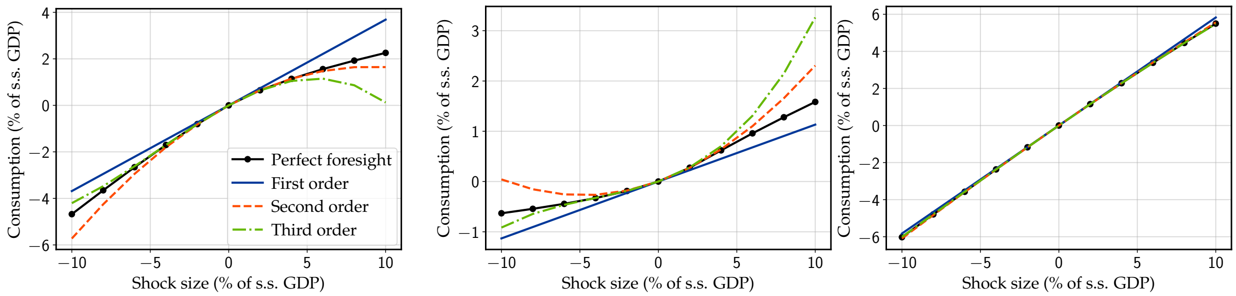
Figure C.1 shows that this is exactly the case. The left panel considers the discounted cumulative impulse response of consumption, $\sum_{t=0}^{k}(1+r)^{-k}(C_t - C_{ss})$, from our baseline shock to income in the PE model, over the first four quarters ($k = 3$). The concavity and convexity in size from figure 2 average out, so that cumulative results over just the first year the predictions from first, second, and third order already line up very closely with those from the nonlinear perfect foresight version of the model. Panel (b) shows that this is also true in the GE model. Panel (c) shows what happens as we look at the infinite horizon cumulative impulse: in the PE model, we get all responses on a line with slope 1, as predicted by the intertemporal budget constraint. In the GE model, there is not such a clean theoretical benchmark, but the nonlinear responses are nevertheless all essentially quantitatively linear with slope equal to the cumulative transfer multiplier (which is around 0.6 in our model).

We consider the partial equilibrium model first, and then turn to the general equilibrium model, in which the same broad conclusions hold.

Lump-sum vs proportional transfers. Our baseline model has a shock to transfers that is distributed proportionately to household income (via Z in the PE model or τ in the GE model), as would be the case for changes in tax rates. In practice, some transfers are distributed more progressively, the classic case being that of lump-sum transfers. Figure C.2 considers the effect of a one-off lump-sum transfer, where the shock is to Tr_t rather than Z_t in equation 6. As the perfect-foresight lines, nonlinearities are bigger for these transfers: there is both more concavity in the first quarter (left panel) and more convexity in the second quarter (right panel), implying that large shocks have a lower impact and more persistent effect on consumption relative to their size. This type of nonlinearity is intuitive: lump-sum shocks of a very large size have a very disproportionate effect on households with low income and low assets. In fact, for agents with approximately zero productivity and zero incoming assets, a negative income shock that exceeds the steady state Tr is not even compatible with positive consumption; which a positive shock of that size would lead to large savings.

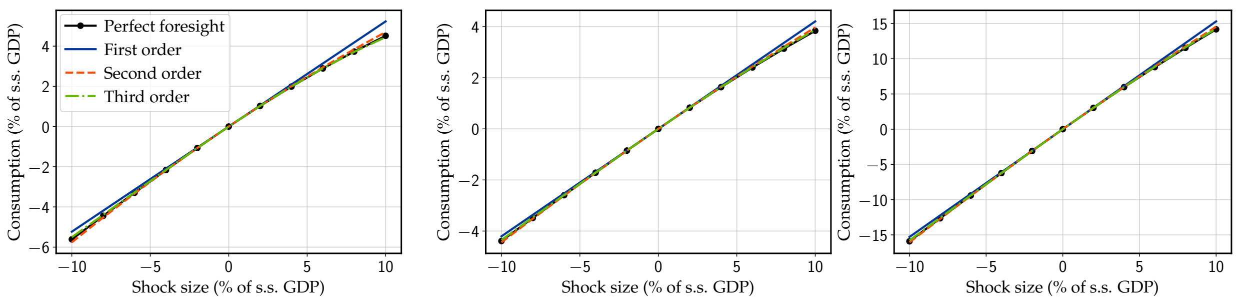
In conclusion, lump-sum transfers do create larger nonlinearities than proportional transfers, and these imply that the second and even the third order perturbation can fail to capture the nonlinearities correctly at very large shock sizes—for instance, the third order does poorly with shocks of +10% of quarterly income, even though it does fine with shocks of -10% of quarterly income (recall that steady state transfers are 14% of quarterly income, so -10% is close to the upper bound of size.) Even then, the right panel shows that these nonlinear errors continue to average out when looking over the four quarter horizon (right panel of Figure C.2).

Transitory vs persistent transfers. So far, we have only considered fully transitory transfers, both proportional and lump-sum transfer ($\rho_Z = \rho_{Tr} = 0$). In practice, however, fiscal shocks tend to be persistent. Persistence, in turns, matters for the nonlinearity of the impact response of a transfer on consumption, because individuals at the borrowing constraint are unable to borrow



(a) Response at $t = 0$ (b) Response at $t = 1$ (c) Cumulative over first 4 quarters

Figure C.2: Response to a one-time lump-sum transfer (PE model)



(a) Response at $t = 0$ (b) Response at $t = 1$ (c) Cumulative over first 4 quarters

Figure C.3: Response to a lump-sum transfer with persistence 0.7 (PE model)

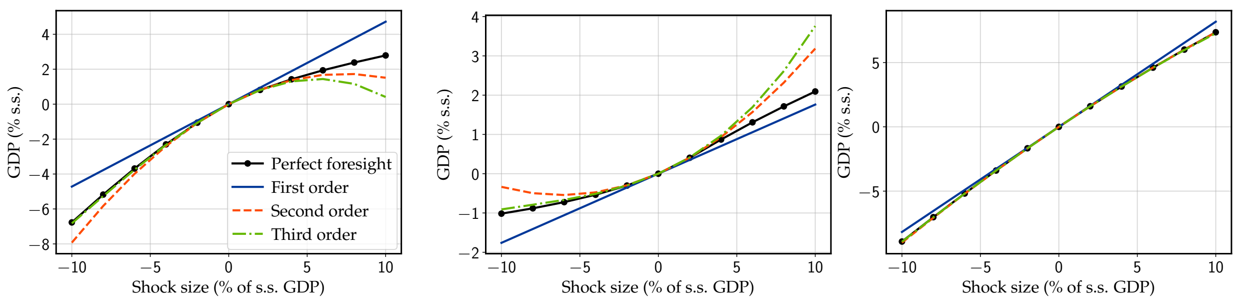
against anticipated future transfers while those further away from the constraint can to some extent, making effective marginal propensities to consume out of transfer receipts less unequal.

Indeed, figure C.3 shows that for aggregate, lump-sum income shocks of persistence $\rho_{Tr} = 0.7$, much of the nonlinearities in both the impact and the delayed effect on consumption goes away—in fact, with this amount of persistence, the $t = 1$ response is still mildly concave in size, with convexity appearing at longer horizons; the cumulative response over 4 quarters is already quite linear.

General equilibrium vs partial equilibrium. We have seen that, in our partial equilibrium model, nonlinearities are greatest when shocks are lump-sum, have no peristence, and their effect is considered over a single quarter. These results carry over to the general equilibrium version of the model, which features amplification from consumption to income across periods but does not fundamentally alter the nature of the nonlinearities in the effect of the fiscal shock itself.

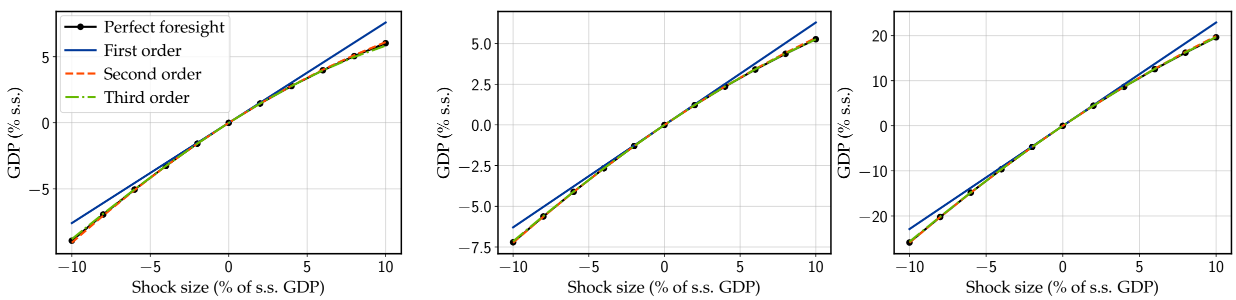
Indeed, figures C.4 and C.5 reproduce C.2 and C.3 in the general equilibrium version of our model, with our background fiscal rule in place to bring bonds back to the steady state at rate $\rho_B = 0.95$. While the amplification effect from the intertemporal keynesian cross alters the slope of the first-order lines, similar nonlinearities as in the partial equilibrium model exist in the perfect foresight solution, and the second- and third-order solutions do a similar job at capturing them:

the third order solution performs poorly for very large one-off lump-sum transfers, but does better once we cumulate the effects over a quarter, consider a more persistent shock, or both.



(a) Response at $t = 0$ (b) Response at $t = 1$ (c) Cumulative over first 4 quarters

Figure C.4: Response to a one-time lump-sum transfer (GE model)



(a) Response at $t = 0$ (b) Response at $t = 1$ (c) Cumulative over first 4 quarters

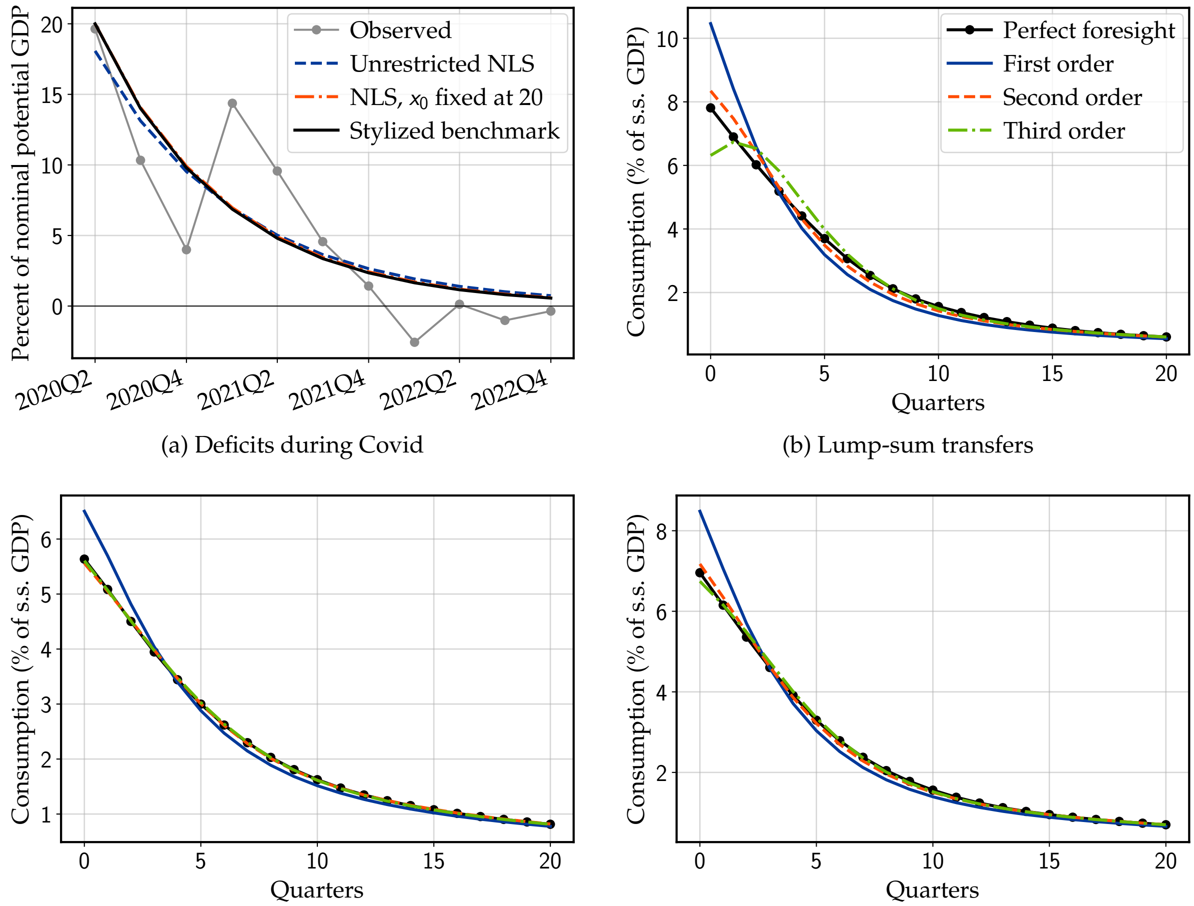
Figure C.5: Response to a lump-sum transfer with persistence 0.7 (GE model)

Covid-sized shock. Covid was a unique historical episode, with the largest-ever fiscal transfers seen in the United States. This provides a natural setting to judge the plausible magnitude and persistence of fiscal shocks.

The left panel of figure C.6 shows U.S. federal net borrowing over potential GDP, net of its pre-pandemic baseline.[A-17] The figure overlays an AR(1) shock with impact size 0.2 and persistence 0.7: this is very close to the best nonlinear least squares AR(1) fit (with impact 0.181 and persistence 0.726, plotted in the dashed blue line). Indeed, the quarterly fiscal deficit was very high in both 2020Q2 and 2021Q1; it only fell gradually back to its pre-Covid level over the following two years. A first takeaway from this exercise is that a fiscal shock persistence of $\rho = 0.7$, as in figures C.5 and C.3, is a natural approximation, and we have seen that this kind of persistence mitigates the nonlinearities in the impact effect of shocks.

A second takeaway is that a shock of 20% of quarterly (5% of annual) GDP is as large an impact shock as we have ever experienced historically. This is a 4σ shock for our baseline calibration,

[A-17] Federal net borrowing is from the BEA (FRED: AD02RC1Q027SBEA), potential GDP is from the CBO (FRED: NGDPPOT), and we take the baseline to be 2019Q4 value of the ratio.



(a) Deficits during Covid

(b) Lump-sum transfers

(c) Proportional taxes

(d) 50/50 mix

Figure C.6: Response to Covid shock

which seems reasonable relative to the size of ongoing shocks. In fact, to simulate the effect of a shock of this magnitude, nonlinear perfect foresight is likely to be a better route than a global solution taking into account ongoing risks.

It is unclear whether we should model the effect of this shock as a lump-sum or a proportional transfer. Very far from all of the deficit went to pay for direct fiscal transfers to households, for which a lump sum would be appropriate: for instance, a significant fraction also financed transfers to firms to make payroll, which eventually benefited households in proportion to their income. A 50/50 mix of lump-sum and proportional transfers seem like a realistic assumption, though we also consider pure lump-sum and pure proportional transfers for completeness.

Panels b, c and d of figure C.6 shows the nonlinear impulse response of consumption to a shock of this size. For pure lump-sum transfers of this size, even the third order solution has trouble matching the perfect-foresight solution on impact. However, third and even second-order appear quite accurate when we consider pure proportional transfers, or even a 50/50 mix of proportional and lump-sum transfers.

Takeaway. We find that third order perturbation used correctly on a smooth model is likely to be reliable in most settings where a failure of Ricardian equivalence from high MPCs is important. Significant errors in the third order approximation relative to the nonlinear perfect foresight solution only appear when have a combination of 1) pure lump-sum rather than proportional (or a mix of proportional and lump-sum) transfers, 2) transitory rather than persistent transfers, 3) looking at the impact effect, rather than cumulative effect over a few quarters, and 4) shocks that outside of a plausible range for business-cycle fluctuations, such as a Covid-sized shock.