Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models

Adrien Auclert[*] Bence Bardóczy[†] Matthew Rognlie[‡] Ludwig Straub[§]

June 2019

Abstract

We propose a general and highly efficient method for solving and estimating general equilibrium heterogeneous-agent models with aggregate shocks in discrete time. Our approach relies on the rapid computation and composition of *sequence-space Jacobians*—the derivatives of perfect-foresight equilibrium mappings between aggregate sequences around the steady state. We provide a fast algorithm for computing Jacobians for heterogeneous agents, a technique to substantially reduce dimensionality, a rapid procedure for likelihood-based estimation, a determinacy condition for the sequence space, and a method to solve nonlinear perfect-foresight transitions. We apply our methods to three canonical heterogeneous-agent models: a neoclassical model, a New Keynesian model with one asset, and a New Keynesian model with two assets.

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

[†]Northwestern University. Email: bardoczy@u.northwestern.edu.

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

[§]Harvard University and NBER. Email: ludwigstraub@fas.harvard.edu.

For helpful comments, we thank Riccardo Bianchi-Vimercati, Luigi Bocola, Jesus Fernández-Villaverde, Joao Guerreiro, Kurt Mitman, Ben Moll, Laura Murphy, Martin Souchier, and Christian Wolf. This research is supported by the National Science Foundation grant SES-1851717.

1 Introduction

A rapidly expanding literature at the frontier of macroeconomics incorporates rich heterogeneity into dynamic general equilibrium models. A central challenge in this literature is that equilibrium involves the time-varying, high-dimensional distribution of agents over their state variables.

In this paper, we propose a general, systematic, and highly efficient method to deal with this challenge. The core idea is to perturb the model to first order in aggregates, exploit certainty equivalence, and write equilibrium as a linear system in the space of perfect-foresight sequences—the *sequence space*. This builds on important work by Reiter (2002, 2009), who reduces equilibrium to a system of linear equations in the state space, and by Boppart, Krusell and Mitman (2018), who iteratively solve for the same equilibrium in the sequence space.

Because we bypass iterations and work directly with a small linear system, our approach makes it feasible to solve and estimate even very high-dimensional models. We demonstrate its power by solving and estimating three models, with increasing degrees of complexity, at unparalleled speed. A code repository accompanies this paper and provides general-purpose routines that automate many of the new algorithms we introduce.[1]

The central objects in our method are *sequence-space Jacobians*: the derivatives of equilibrium mappings between aggregate sequences around the steady state. These Jacobians summarize every aspect of the model that is relevant for general equilibrium. To understand why our approach is able to deal with arbitrary amounts of heterogeneity, consider a model in which households face income uncertainty and incomplete markets. Any such model features a Jacobian $\mathcal{J}^{C,r}$ that maps, to first order, changes in the sequence of real interest rates $\{r_t\}$ to changes in the sequence of aggregate consumption $\{C_t\}$. Under the hood, this mapping includes the heterogeneous responses of households to changes in r, as well as the evolution over time of the distribution of agents that follows from this change in r. But to know the aggregate effects of r on C, all we need to know is $\mathcal{J}^{C,r}$: it is a *sufficient statistic*. Our method exploits this property. We compute all relevant sequence-space Jacobians, and then compose and invert these Jacobians to obtain the model's full set of impulse responses to shocks, perform likelihood-based estimation, test for equilibrium determinacy, and solve for nonlinear perfect-foresight transitions very rapidly.

We make a series of five specific methodological contributions that operationalize the use of sequence-space Jacobians for solving and estimating a large class of heterogeneous-agent models.

First, we show how to compute Jacobians for the aggregate behavior of heterogeneous agents, such as the Jacobian $\mathcal{J}^{C,r}$ described above. We offer a straightforward direct method to compute Jacobians truncated to a horizon of $T \times T$. We then introduce a new "fake news" algorithm, which lowers the time it takes to compute these Jacobians by a factor of about T relative to this direct method. This provides a dramatic speed improvement, since T is typically at least equal to 100 in practice, and often as large as 1000.

Second, we introduce a method to reduce the dimensionality of the system used to solve for

[1] See https://github.com/shade-econ/sequence-jacobian, which we are continuing to update.

equilibrium. We start by characterizing equilibrium as a solution to a certain nonlinear system

$$\mathbf{H}(\mathbf{U}, \mathbf{Z}) = 0 \tag{1}$$

where $\mathbf{U}$ represents the time path $\mathbf{U}_0, \mathbf{U}_1, \ldots$ of n_u unknown aggregate sequences (usually aggregate prices and quantities) and $\mathbf{Z}$ represents the time path $\mathbf{Z}_0, \mathbf{Z}_1, \ldots$ of n_z exogenous shocks. With a truncation horizon of T, the linearized version of (1) is a system of $n_u T$ equations in $n_u T$ unknowns, which for models with many aggregate series can be large enough to become a bottleneck. To reduce n_u, we represent equilibrium computation as a directed acyclic graph (DAG), where each node represents a "block" with an easy-to-compute Jacobian, and we show how to efficiently compose Jacobians along this DAG. We also make the composition of blocks very efficient by exploiting sparsity. This allows us to compute the impulse response of any endogenous variable to any transitory shock $d\mathbf{Z}_0, d\mathbf{Z}_1, \ldots$ in a matter of milliseconds.

Third, we propose a method for rapid likelihood-based estimation.[2] As pointed out by Boppart, Krusell and Mitman (2018), the impulse response we solve for is equivalent to the $MA(\infty)$ representation of the model with aggregate shocks. We can therefore obtain autocovariances directly from impulse responses, and the likelihood function directly from autocovariances. We show how precomputing Jacobians makes estimating shock processes a matter of a few seconds, and estimating parameters a matter of a few minutes, even for the most complex model we consider.

Fourth, we introduce a novel local determinacy criterion for the sequence space, analogous to the Blanchard and Kahn (1980) condition for models of the form (1). We first we prove that the Jacobians $\mathbf{H}_\mathbf{U}$ and $\mathbf{H}_\mathbf{Z}$ of $\mathbf{H}$ have a special *asymptotically time-invariant* structure, such that all their far-out columns look alike. Our criterion involves a simple "winding number" condition on any far-out column of $\mathbf{H}_\mathbf{U}$, substantially generalizing results by Onatski (2006) to the case of heterogeneous-agent models. This criterion can be evaluated almost instantly once $\mathbf{H}_\mathbf{U}$ is known.

Finally, we propose a Newton-based method to solve equation (1) nonlinearly very rapidly. This nonlinear solution is relevant for the literature exploring size dependence and sign asymmetries in models with perfect foresight (see e.g. Kaplan and Violante 2018 for fiscal policy and Berger, Guerrieri, Lorenzoni and Vavra 2018 for house price changes), or to test for the accuracy of the linearization with respect to aggregates (Boppart, Krusell and Mitman 2018). We show how, with knowledge of sequence-space Jacobians, we can reliably converge to the nonlinear solution in just a few iterations.

Our methods apply to a large class of models, which we call *SHADE models*, for "Sequence-space Heterogeneous-Agent Dynamic Equilibrium". These models include most heterogeneous-agent neoclassical and New Keynesian models, most models of firm life cycles (Hopenhayn 1992, Hopenhayn and Rogerson 1993), lumpy investment models as in Khan and Thomas (2008), pric-

[2] Here we refer to estimation as the process of finding the posterior mode and computing standard errors using the Laplace approximation. We leave the exploration of the full posterior distribution using Markov-Chain Monte Carlo methods to future work.

Table 1: Summary of computing times.

Computing times for:	Krusell-Smith	HD Krusell-Smith	one-asset HANK	two-asset HANK
Heterogeneous-agent Jacobians	0.10 s	8.4 s	0.65 s	5.7 s
One impulse response	0.0012 s	0.0012 s	0.017 s	0.120 s
All impulse responses ($\mathbf{G}$)	0.0068 s	0.0068 s	0.097 s	0.400 s
Bayesian estimation (shocks)				
single likelihood evaluation	0.00088 s	0.00088 s	0.0021 s	0.058 s
entire estimation	0.12 s	0.12 s	0.50 s	21 s
Bayesian estimation (shocks + model)				
single likelihood evaluation	—	—	0.011 s	0.18 s
entire estimation	—	—	16 s	570 s
Determinacy test	252 μs	252 μs	631 μs	631 μs
Nonlinear impulse responses	0.18 s	13.76 s	0.96 s	27 s
No. of idiosyncratic states	3,500	250,000	3,500	10,500
Time horizon (T)	300	300	300	300
No. of shock parameters in estimation	3	3	6	14
No. of model parameters in estimation	0	0	3	5

Notes. Our Krusell-Smith model and its "high-dimensional" (HD) version are described in Section 2 and Appendix A.1. Our one-asset HANK model is described in Appendix A.2. Our two-asset HANK model is described in Appendix A.3. Bayesian estimation refers to finding the posterior mode and computing standard errors using the Laplace approximation. All calculations were performed on a 2017 MacBook Pro laptop with a 2.8 GHz processor and four cores.

ing models as in Golosov and Lucas (2007), or overlapping generation models as in Conesa and Krueger (1999). To illustrate the breadth of our methods, we apply them to three canonical heterogeneous-household models of increasing complexity: a neoclassical model in the spirit of Krusell and Smith (1998), a one-asset New Keynesian model in the spirit of McKay, Nakamura and Steinsson (2016), and a two-asset New Keynesian model in the spirit of Kaplan, Moll and Violante (2018).

Table 1 illustrates the speeds that our algorithms are able to achieve on a laptop computer.[3] For each of our main models (including a high-dimensional version of the Krusell-Smith model), it takes less than 8 seconds to compute the Jacobians $\mathbf{H_U}$ and $\mathbf{H_Z}$. Once these Jacobians are known, it is immediate to test for determinacy and calculate impulse responses. Posterior-mode estimation takes less than 10 minutes for every model, and is, for simpler models, a matter of seconds or milliseconds. By contrast, the leading computational techniques existing today find it challenging to estimate a two-asset HANK model at all.

Related literature. Since the early breakthroughs of Krusell and Smith (1998) and den Haan (1997), the literature on solution methods for heterogeneous-agent models has grown tremendously. Part of the literature has developed nonlinear methods, which are well-suited to address

[3] All computations were performed on a 2017 MacBook Pro laptop with a 2.8 GHz processor and four cores.

questions that inherently involve higher-order aggregate moments, such as risk premia or the effects of volatility shocks.[4] Our paper follows another class of methods, which linearize with respect to aggregates but preserve nonlinearities with respect to idiosyncratic shocks. These methods are growing in popularity, since they are typically able to accommodate more complex models while retaining the rich interactions between the distribution of agents and macroeconomic outcomes that are the hallmark of the recent heterogeneous-agent literature (see, for example, Krueger, Mitman and Perri 2016 and Kaplan and Violante 2018).

As discussed at the top of the paper, we specifically leverage the main strengths of two leading linearization approaches in the literature.

The first approach we build on has become known as the "Reiter method", following the influential contributions of Reiter (2002, 2009).[5] This method incorporates individual policy functions and discretized distributions as part of the state space, solves the model without aggregate shocks nonlinearly, and then linearizes the model with respect to aggregates. The resulting linear system can then be solved and estimated using standard linear rational expectations methods. The main limitation of this approach is that the dimension of the linear system grows with the dimension of the state space of the heterogeneous-agent model. For many complex applications, standard rational expectations solution methods become too costly. To address this challenge, the literature has developed model reduction techniques (see Reiter 2010, Ahn et al. 2018, Winberry 2018, Bayer and Luetticke 2018). These bring down the dimension of the linear system, but their effectiveness depends on the nature of the heterogeneity, and the resulting size can remain prohibitive.[6]

The second approach we build on, which we will refer to as the "BKM method", follows the lead of Boppart, Krusell and Mitman (2018) and solves for perfect-foresight impulse responses to small unanticipated aggregate shocks, starting from the steady state with no aggregate risk ("MIT shocks").[7] The key insight of Boppart, Krusell and Mitman (2018) is that this method solves for the same local impulse response as the Reiter method: once a model is linearized to first order in aggregates, it features aggregate certainty equivalence, and the impulse responses to MIT shocks are the same as those of the full stochastic model (Simon 1956, Theil 1957). In practice, to obtain these impulses, the BKM method then follows the standard nonlinear approach in the literature: write aggregate equilibrium conditions as functions of certain unknown aggregate sequences, and then iterate over guesses for these sequences until convergence. This approach is intuitive and bypasses the need for a large state-space system. Its major limitation is that it often requires a

[4] See the survey by Algan, Allais, Den Haan and Rendahl (2014) and recent work by Brumm and Scheidegger (2017), Mertens and Judd (2018), Proehl (2019), and Fernández-Villaverde, Nuño and Hurtado (2019), among many others.

[5] This is the method used in McKay and Reis (2016), Winberry (2018), Bayer, Luetticke, Pham-Dao and Tjaden (2019), Mongey and Williams (2017), Ahn, Kaplan, Moll, Winberry and Wolf (2018), and Plagborg-Møller and Liu (2019), among many others.

[6] For example, absent model reduction, the two asset model of Ahn et al. (2018) requires solving a linear rational expectations system of 120,000 in 120,000 variables. With model reduction, they are able to cut this size to a 2,445 by 2,445 system, but cannot lower the dimension further without altering the shape of their impulse responses.

[7] The nonlinear MIT shock method dates back to Auerbach and Kotlikoff (1987) for OLG models and to Conesa and Krueger (1999) for models with idiosyncratic risk. It was recently used by Guerrieri and Lorenzoni (2017), McKay et al. (2016) and Kaplan et al. (2018), among many others.

large number of iterations, making it too slow for advanced applications such as estimation—and convergence is not even guaranteed in general.

Our paper synthesizes these two approaches. As in Reiter (2009), we perturb the model to first order in aggregates and write equilibrium as a linear system, so that solving for equilibrium becomes a simple matter of linear algebra, bypassing the need for any iteration. But this linear system is in the *sequence space*, like BKM, rather than the state space, and it characterizes the first-order impulse response to unexpected shocks around the steady state. And as in the BKM method, the size of the system only scales with the macro complexity of the model (the number of aggregate variables), not with the micro complexity (the number of idiosyncratic states, the presence of occasionally binding borrowing constraints, and so on).[8]

The contributions of this paper relate to four other branches of the existing literature.

First, there is a growing literature on sufficient statistics for general equilibrium (Auclert and Rognlie 2018, Auclert, Rognlie and Straub 2018, Guren, McKay, Nakamura and Steinsson 2018, Koby and Wolf 2018, Wolf 2019). But, while this literature uses sufficient statistics as an empirical strategy to connect models to the data, or as a conceptual strategy to shed light on equilibrium adjustment mechanisms, in this paper we use them as a *computational* strategy to solve a large class of models very efficiently.

Second, the idea of using sequence-space Jacobians to solve for perfect-foresight equilibria relates to a literature that uses Newton-based methods to solve for nonlinear perfect-foresight transition paths in representative agent models (see Laffargue 1990, Boucekkine 1995, and Juillard 1996). A version of this approach is currently implemented in Dynare. The algorithms in this literature exploit the sparsity of representative-agent sequence-space Jacobians to accelerate the computation of matrix inverses. As we show, the heterogeneous-agent component implies that SHADE model Jacobians are never entirely sparse. This requires us to develop alternative methods.

Third, our methods for reducing dimensionality have two precedents. The idea of breaking down equilibrium computation into a series of small blocks formalizes a common procedure from the literature solving MIT shocks.[9] The idea of representing the flow of computation as a directed acyclic graph (DAG), and of accumulating Jacobians along this DAG, relates to what is known as the "forward mode" in the automatic differentiation literature (see Griewank and Walther 2008). While automatic differentiation has been used in the heterogeneous-agent literature for models in state-space form (e.g. Ahn et al. 2018, Childers 2018), to the best of our knowledge, we are the first to apply this idea to compute Jacobians in sequence space, for which it is particularly well-suited.[10]

[8] We share with all aggregate linearization methods the drawback that the model does not generate risk premia, portfolio choice is indeterminate, and optimal Ramsey policy is ill-defined. For these applications, higher-order perturbations or global solution methods are more appropriate (see for example Fernández-Villaverde, Rubio-Ramírez and Schorfheide 2016.)

[9] It also relates to ideas developed in an older, static, computational general equilibrium literature (e.g. Mansur and Whalley 1982, van der Laan 1985).

[10] Dimensionality reduction is also a key focus of the model reduction literature (e.g. Reiter 2010, Ahn et al. 2018,

Finally, the idea of using impulse responses directly to perform model estimation relates to an early time series literature on the estimation of MA processes (e.g. Box and Jenkins 1970, Hamilton 1994). The vast majority of the DSGE literature uses the Kalman filter to estimate models in state-space form (e.g. Herbst and Schorfheide 2015 and Fernández-Villaverde et al. 2016). Our approach instead relates to the small DSGE literature that estimates models using their $MA(\infty)$ representation (Mankiw and Reis 2007, Schmitt-Grohé and Uribe 2010, Meyer-Gohde 2010, Lan and Meyer-Gohde 2013). To the best of our knowledge, we are the first to apply these ideas to the estimation of heterogeneous-agent models.

Layout. The rest of the paper proceeds as follows. Section 2 introduces our computational method with an example. Section 3 provides our new algorithm for computing heterogeneous-agent Jacobians. Section 4 shows how to combine these Jacobians to compute impulse responses in SHADE models, and introduces our DAG-based dimension reduction method. Section 5 shows how to use these impulse responses to perform fast likelihood-based estimation, and estimates our three example economies with US data. Section 6 provides our new determinacy criterion for the sequence space. Section 7 shows how to use sequence-space Jacobians to compute nonlinear perfect-foresight transition paths. Section 8 concludes.

2 The sequence-space Jacobian: an example

We introduce our methods by means of an example: Krusell and Smith (1998)'s celebrated extension of the real business cycle model to heterogeneous households. This model is a natural starting point, since it well-known and there exist many well-established algorithms for solving it.

We set up the model in the sequence space, that is, assuming perfect foresight with respect to aggregates. We then show how to use the sequence-space Jacobian to solve for the impulse response of the model to a total factor productivity (TFP) shock in a fraction of a second.

2.1 Model description

The economy is populated by a mass 1 of heterogeneous households that maximize the time-separable utility function $\mathbb{E}\left[\sum \beta^t u\left(c_t\right)\right]$, where u has the standard constant relative risk aversion form, $u\left(c\right) = \frac{c^{1-\sigma}}{1-\sigma}$. There exist n_e idiosyncratic states, and in any period t, agents transition between any two such states e and e' with exogenous probability $P\left(e, e'\right)$. We denote by π the stationary distribution of P and assume that the mass of agents in each state e is always equal to $\pi\left(e\right)$.[11] Agents supply an exogenous number of hours l, and earn wage income $w_t el$, where w_t is

Winberry 2018, Bayer and Luetticke 2018). These methods reduce dimension in the state space by approximating distributions with lower-dimensional objects. By contrast, we make no approximation to the distribution when solving for equilibrium. Our main source of approximation error comes from truncation at finite T. We evaluate the inaccuracies these create in section 4.4.

[11] In the original Krusell and Smith (1998) model, the transition probabilities depend on the aggregate state, that is, $P\left(e, e', Z_t\right)$. Our methods can be applied to this case as well (see the general formulation in section 3.1).

the wage per efficient hour. Agents can only trade in capital k, which pays a rental rate r_t net of depreciation, and are subject to a no-borrowing constraint. The value function of an agent in state (e, k_-) at time t is therefore

$$\begin{aligned} V_t(e, k_-) = \quad & \max_{c,k} \quad u(c) + \beta \sum_{e'} V_{t+1}(e', k) P(e, e') \\ & \text{s.t.} \quad c + k = (1 + r_t) k_- + w_t e l \\ & \qquad k \geq 0 \end{aligned} \tag{2}$$

Denote by $c_t^*(e, k_-)$ and $k_t^*(e, k_-)$ the policy functions that solve the Bellman equation (2). Also denote by $D_t(e, K_-) \equiv \Pr(e_t = e, k_{t-1} \in K_-)$ the measure of households in state e that own capital in a set K_- at the start of date t. The distribution D_t has law of motion

$$D_{t+1}(e', K) = \sum_e D_t\left(e', k_t^{*-1}(e, K)\right) P(e, e') \tag{3}$$

where $k_t^{*-1}(e, \cdot)$ denotes the inverse of $k_t^*(e, \cdot)$. We assume that prior to $t = 0$, the economy is in a steady state with constant wage w_{ss} and net rental rate r_{ss}, corresponding to a steady state of the general equilibrium economy discussed momentarily.[12] In this steady state, there is a unique value function and decision rule solving (2), and then a unique stationary distribution D_{ss} solving (3). We suppose that agents start in this stationary distribution at date 0, so that $D_0 = D_{ss}$.

Equation (2) shows that, for any t, the policy $k_t^*(e, k_-)$ is a function of the future path $\{r_s, w_s\}_{s \geq t}$. Given $D_0 = D_{ss}$, through (3), the distribution $D_t(e, K)$ at any t is a function of the entire path $\{r_s, w_s\}_{s \geq 0}$.[13] It follows that aggregate household capital holdings are characterized by a *capital function* $\mathcal{K}_t\left(\{r_s, w_s\}_{s \geq 0}\right)$, where

$$\mathcal{K}_t(\{r_s, w_s\}) = \sum_e \int_{k_-} k_t^*(e, k_-) D_t(e, dk_-) \tag{4}$$

The ability to reduce interactions between heterogeneous agents to functions such as $\mathcal{K}_t$, which maps aggregate sequences into aggregate sequences, is key to the sequence-space Jacobian method. We now combine this $\mathcal{K}_t$ function with equations describing production and market-clearing conditions to describe the entire Krusell-Smith economy. Production in this economy is carried out by a competitive representative firm, which has a Cobb-Douglas technology $Y_t = Z_t K_{t-1}^{\alpha} L_t^{1-\alpha}$, rents capital and labor from workers at rates $r_t + \delta$ and w_t, and faces the sequence of total factor

[12] Achdou, Han, Lasry, Lions and Moll (2017) show that the steady state is unique when $\sigma \leq 1$.

[13] This can be shown recursively: given $D_0 = D_{ss}$, D_1 is a function of $\{r_s, w_s\}_{s \geq 0}$, and therefore so is D_2, through its dependence on D_1. In section 3 we elicit explicitly the first-order dependence of D_t, k_t^*, and $\mathcal{K}_t$ on the sequence $\{r_s, w_s\}_{s \geq 0}$.

productivity Z_t. The firm's first-order conditions

$$r_t = \alpha Z_t \left(\frac{K_{t-1}}{L_t}\right)^{\alpha-1} - \delta \tag{5}$$

$$w_t = (1-\alpha) Z_t \left(\frac{K_{t-1}}{L_t}\right)^{\alpha} \tag{6}$$

relate the paths of prices $\{r_t, w_t\}$ to the exogenous paths $\{Z_t, L_t = \sum \pi(e)\, el\}$ and the endogenous path for capital $\{K_t\}$. Combining (4)–(6), we can express the capital market clearing condition at each point in time as a function H,

$$H_t(\mathbf{K}, \mathbf{Z}) \equiv \mathcal{K}_t\left(\left\{\alpha Z_s \left(\frac{K_{s-1}}{\sum \pi(e)\, el}\right)^{\alpha-1} - \delta, (1-\alpha) Z_s \left(\frac{K_{s-1}}{\sum \pi(e)\, el}\right)^{\alpha}\right\}\right) - K_t = 0 \tag{7}$$

where $\mathbf{K} = (K_0, K_1, \ldots)'$. Given initial capital K_{-1} and the exogenous path for productivity, $\mathbf{Z} = (Z_0, Z_1, \ldots)'$, equation (7) pins down the equilibrium path of capital. Given $\mathbf{K}$, it is then immediate to obtain the value of all other endogenous variables, Y_t, L_t, w_t, and r_t at every t.[14]

2.2 Impulse responses

Applying the implicit function theorem to (7), the linear impulse response of capital to a transitory technology shock $d\mathbf{Z} = (dZ_0, dZ_1, \ldots)'$ is given by

$$d\mathbf{K} = -\mathbf{H}_{\mathbf{K}}^{-1} \mathbf{H}_{\mathbf{Z}} d\mathbf{Z} \tag{8}$$

where $\mathbf{H}_{\mathbf{K}}$ and $\mathbf{H}_{\mathbf{Z}}$ denote the Jacobians of $\mathbf{H}$ with respect to $\mathbf{K}$ and $\mathbf{Z}$, evaluated at the steady state. Given $d\mathbf{K}$, the impulse responses of other variables, e.g. $\{r_s, w_s\}$, follow immediately. In practice, (8) is solved up to a given (large) horizon T such that K and Z have approximately returned to steady state by time T.

We can in turn use the chain rule to relate the Jacobians $\mathbf{H}_{\mathbf{K}}$ and $\mathbf{H}_{\mathbf{Z}}$ to the derivatives of the $\mathcal{K}$ function defined in equation (4), all evaluated at the steady state. For example, differentiating (7) with respect to K_s, we find that the t, s entry of $\mathbf{H}_{\mathbf{K}}$ is

$$[\mathbf{H}_{\mathbf{K}}]_{t,s} = \frac{\partial \mathcal{K}_t}{\partial r_{s+1}} \frac{\partial r_{s+1}}{\partial K_s} + \frac{\partial \mathcal{K}_t}{\partial w_{s+1}} \frac{\partial w_{s+1}}{\partial K_s} - 1_{\{s=t\}} \tag{9}$$

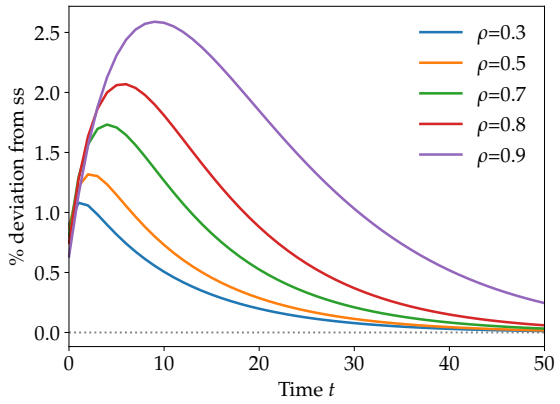
and a similar expression applies to $\mathbf{H}_{\mathbf{Z}}$. In addition, the derivatives $\frac{\partial r_{s+1}}{\partial K_s}$, $\frac{\partial w_{s+1}}{\partial K_s}$, $\frac{\partial r_{s+1}}{\partial Z_s}$ and $\frac{\partial w_{s+1}}{\partial Z_s}$ at $(\mathbf{K}_{ss}, \mathbf{Z}_{ss})$ can all be computed analytically: for example,

$$\frac{\partial r_{s+1}}{\partial K_s} = \alpha(\alpha-1) Z_{ss} \left(\frac{K_{ss}}{L_{ss}}\right)^{\alpha-2}$$

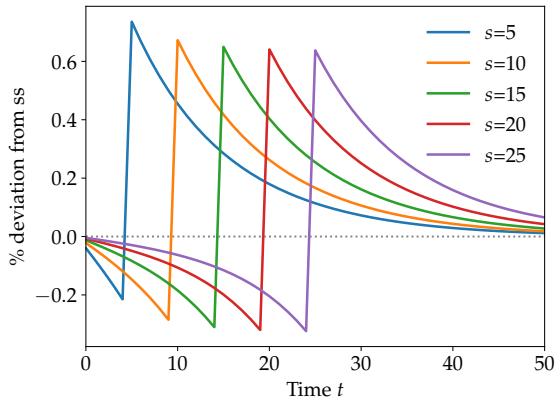
[14] Note that we were able to drop goods market clearing, owing to Walras's law.

Figure 1: Impulse responses of capital to 1% TFP shocks in the "high-dimensional" Krusell-Smith model

(a) AR(1) shock with persistence ρ



(b) News shock at time s



Hence, to obtain $\mathbf{H_K^{-1}H_Z}$ in (8), all we need are the Jacobians of the $\mathcal{K}$ function with respect to its two inputs r and w. Obtaining these Jacobians is perhaps the single biggest challenge in applying our methods. In section 3, we introduce two algorithms, a *direct* and a *fake news algorithm*, to compute them. As table 1 reveals, for a standard calibration of the Krusell-Smith model detailed in appendix A.1, our fake news algorithm takes 100 milliseconds to calculate Jacobians of $\mathcal{K}$, truncated to a horizon of 300×300. In a "high-dimensional" calibration that increases the dimensionality of the state space from 3,500 to 250,000, it still only takes 8 seconds.

Once we have these Jacobians, the underlying heterogeneity no longer matters: the Jacobians tell us everything that we need to know, to first order, about the household side of the model. This feature of our method is apparent in table 1, where we see that most other computing times are identical between the two calibrations of the Krusell-Smith model, despite the large disparity in the size of their state spaces.

Impulse responses and news-shock interpretation. Once we have the Jacobians of $\mathcal{K}$, we can immediately calculate $-\mathbf{H_K^{-1}H_Z}$. Given (8), applying this matrix to any path for $d\mathbf{Z}$ gives us the impulse response $d\mathbf{K}$ of capital with just a single matrix-vector multiplication, which is almost instantaneous. Panel (a) of figure 1 does this for a variety of $d\mathbf{Z}$, representing 1% AR(1) shocks to TFP with different persistences ρ in our high-dimensional Krusell-Smith model.

It is even more immediate to obtain the effect of the "news" at date 0 that TFP will be higher by 1% at time s, as in panel (b) of figure 1. By definition, the impulse responses to s-period ahead news are equal to the s^{th} column of the matrix $-\mathbf{H_K^{-1}H_Z}$. This "news shock" interpretation of the columns provides a useful way of understanding their role in the computation of generic impulse responses. For example, the impulse responses to AR(1) TFP paths of persistence ρ in panel (a) can be reinterpreted as the effect of the simultaneous news, at date 0, of an increase of ρ^s in TFP at times $s = 0, 1, \ldots$

3 Computing Jacobians for heterogeneous-agent blocks

In the previous section we established the usefulness of knowing the Jacobians $\partial\mathcal{K}/\partial r$ and $\partial\mathcal{K}/\partial w$ for computing the impulse responses of the Krusell-Smith model. The spirit of equation (9) should make clear that the advantages of Jacobians extend well beyond this specific example. Indeed, in section 4 we will show that, in a large class of models, the Jacobians of the $\mathbf{H}$ function can be related, via the chain rule, to derivatives of one or more heterogeneous-agent *outputs* with respect to one or more heterogeneous-agent *inputs*.

These derivatives generalize the concept of the $\mathcal{K}$ function Jacobians. In the general case, *outputs* can describe aggregate savings, consumption, investment, or other decisions by heterogeneous households or firms, while *inputs* are the aggregates relevant to the decision-making of individual agents, such as interest rates or wages. We now describe two algorithms for computing the Jacobians of outputs o with respect to inputs i, which we denote generically by $\mathcal{J}^{o,i}$.

3.1 Heterogeneous-agent blocks

We first formally introduce the functions whose Jacobians we would like to solve for. We refer to these functions as *heterogeneous-agent blocks*. In section 4, we will combine these blocks with other "simple" blocks to form general equilibrium models.

Definition 1. A *heterogeneous-agent block* maps sequences $\mathbf{X}^i$, for $i = 1, \ldots, n_x$, which we call *inputs*, to sequences $\mathbf{Y}^o$, for $o = 1, \ldots, n_y$, which we call *aggregate outcomes* or *outputs*. Stacking the $\mathbf{X}^i$s into a vector $\mathbf{X}$, and the $\mathbf{Y}^o$s into a vector $\mathbf{Y}$, we represent this map with the function

$$\mathbf{Y} = h(\mathbf{X}) \tag{10}$$

whose time-t output must take the form

$$\mathbf{Y}_t = \sum_{e \in \mathcal{E}} \int_{k \in \mathbb{K}} y\left(e, k_-; V_{t+1}, \mathbf{X}_t\right) D_t\left(e, dk_-\right) \tag{11}$$

for a given *individual outcome* function $y : \mathcal{E} \times \mathbb{K} \times \mathbb{R}^{\mathcal{E}\times\mathbb{K}} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$. In (11), D_t denotes a measure of agents over a set of exogenous and endogenous idiosyncratic states $(e, k_-) \in \mathcal{E} \times \mathbb{K}$, associated with a dynamic programming problem

$$V_t\left(e, k_-\right) = \max_{k \in \Gamma(e, k_-, \mathbf{X}_t)} u\left(e, k_-, k, \mathbf{X}_t\right) + \beta \sum_{e' \in \mathcal{E}} V_{t+1}\left(e', k\right) P\left(e, e', \mathbf{X}_t\right) \tag{12}$$

and whose law of motion is described, for any $e' \in \mathcal{E}$ and $K \subseteq \mathbb{K}$, by[15]

$$D_{t+1}(e', K) = \sum_{e \in \mathcal{E}} D_t(e, k_t^{*-1}(e, K)) P\left(e, e', \mathbf{X}_t\right) \tag{13}$$

[15]Throughout, we assume that $P(e, e', \mathbf{x}_t)$ is a stochastic transition matrix for every $\mathbf{x}_t$; $\mathbb{K}$ is a measurable space and $K \subseteq \mathbb{K}$ refers to a measurable set K.

where $k_t^{*-1}(e,K) \equiv \{k_-|k_t^*(e,k_-) \in K\}$ and $k_t^*(e,k_-)$ denotes a maximizer of (12).

The aggregate savings function $\mathcal{K}_t(\{r_s,w_s\})$ of the Krusell-Smith model fits the scope of definition 1. Indeed, given agents' continuation utility V_{t+1} and the pair $\mathbf{X}_t = (r_t,w_t)$, one can obtain the savings policy function as a solution to the Bellman equation (2), and therefore write our individual outcome function of interest as $y(e,k_-;V_{t+1},\mathbf{X}_t) \equiv k_t^*(e,k_-)$. Moreover, equation (2) clearly fits the general class of Bellman equations in (12), and the laws of motion for the distributions, (3) and (13), are identical.

The formulation in definition 1 is more general, however, in that it can accommodate multidimensional choices, as in heterogeneous-household problems with endogenous labor supply (see appendix A.2); multidimensional endogenous state variables, as in heterogeneous-household problems with two assets (appendix A.3); models with fixed costs, as in heterogeneous-firm problems with investment frictions (e.g. Khan and Thomas 2008), or menu costs in price-setting (e.g. Golosov and Lucas 2007); as well as models with search and matching where aggregate employment prospects affect the job finding and job destruction rate (e.g. Gornemann, Kuester and Nakajima 2016).[16] The main limitation is that (12) cannot depend on the full distribution directly, such as in OLG models with an endogenous distribution of bequests that are received in mid-life (e.g. De Nardi 2004, Straub 2017), or models of wage posting with on-the-job search a la Burdett and Mortensen (1998).

To evaluate heterogeneous-agent blocks $h(\mathbf{X})$ in practice, it is necessary to discretize the problem. There are many ways to do this. Any discretization routine uses a parameterization of the distribution $\mathbf{D}_t$ and a finite-dimensional forward-looking variable $\mathbf{v}_t$. We assume that $\mathbf{D}_t$ represents the distribution mass at each one of n_g gridpoints (the "histogram" of the distribution). For concreteness, in our baseline case we also let $\mathbf{v}_t$ be the value function V_t at each point on the same grid. However, we note that it is often more efficient computationally to include in the vector $\mathbf{v}_t$ information about derivatives of the value function in addition to, or instead of, the level of V_t, and our algorithm below applies to such formulations as well.[17] A discretization of (12), (13) and (11) results in the following system of equations:

$$\mathbf{v}_t = v(\mathbf{v}_{t+1},\mathbf{X}_t) \tag{14}$$

$$\mathbf{D}_{t+1} = \Lambda(\mathbf{v}_{t+1},\mathbf{X}_t)'\mathbf{D}_t \tag{15}$$

$$\mathbf{Y}_t = y(\mathbf{v}_{t+1},\mathbf{X}_t)'\mathbf{D}_t \tag{16}$$

where $\mathbf{v}_t$ and $\mathbf{D}_t$ are $n_g \times 1$ vectors representing the value function and distribution at each grid-

[16] Equations (12) and (13) do not allow forward-looking choices to affect $P(e,e',\mathbf{X}_t)$ directly (e.g. endogenous search effort), but this is for notational simplicity and would pose no difficulty for our algorithm.

[17] In particular, as implemented in our online code, our preferred solution methods for one-asset models combine Carroll (2006)'s endogenous grid method for policy function iteration with Young (2010)'s non-stochastic simulation method to translate optimal policies in between gridpoints to transition probabilities on the grid. In this case, $\mathbf{v}_t$ is the marginal value function $V'(k_t)$ at each point on the grid. Similarly, when applying our two-asset model algorithm from appendix B.1, $\mathbf{v}_t$ includes the marginal values with respect to both the liquid and illiquid assets at each point on the grid.

point, $\mathbf{X}_t$ is an $n_x \times 1$ vector of inputs, $\mathbf{Y}_t$ is an $n_y \times 1$ vector of aggregate outcomes, $\Lambda(\mathbf{v}_{t+1}, \mathbf{X}_t)$ is an $n_g \times n_g$ transition matrix that discretizes the law of motion (13), and $y(\mathbf{v}_{t+1}, \mathbf{X}_t)$ is a $n_g \times n_y$ matrix representing, for each aggregate outcome, the value of the individual outcome function y in equation (11) at each point on the grid.

Starting from (14)–(16), we can use a standard approach to obtain the steady-state $(\mathbf{v}_{ss}, \mathbf{D}_{ss}, \mathbf{y}_{ss}, \mathbf{Y}_{ss})$ given steady-state inputs $\mathbf{X}_{ss}$. First, iterate (14) backward from some initial guess for $\mathbf{v}_{ss}$ until approximate convergence to $(\mathbf{v}_{ss}, \mathbf{y}_{ss})$. With the interpretation of $\mathbf{v}_t$ as the value function at each gridpoint, this procedure is just value function iteration on a discretized version of (12). Second, iterate (15) forward from some initial guess for $\mathbf{D}_{ss}$, by repeatedly applying the transition matrix $\Lambda(\mathbf{v}_{ss}, \mathbf{X}_{ss})$, until approximate convergence to $\mathbf{D}_{ss}$. Third, apply (16) to obtain $\mathbf{Y}_{ss}$.

In general, to solve (14)–(16) for sequences $\{\mathbf{v}_t, \mathbf{D}_t, \mathbf{y}_t, \mathbf{Y}_t\}$ given an arbitrary time-varying sequence $\{\mathbf{X}_t\}$ of inputs, it is necessary to truncate at some time T. Suppose that $\mathbf{X}_t = \mathbf{X}_{ss}$ for all $t \geq T$, and that we are interested in obtaining the first T entries in the sequence $\{\mathbf{Y}_t\} = h(\{\mathbf{X}_t\})$. We can then apply a standard algorithm that parallels the steady-state computation. First, iterate (14) backward, starting with $\mathbf{v}_T = \mathbf{v}_{ss}$, and compute forward-looking variables $\mathbf{v}_t$ as well as individual outcomes $y(\mathbf{v}_{t+1}, \mathbf{X}_t)$ and the transition matrix $\Lambda(\mathbf{v}_{t+1}, \mathbf{X}_t)$ for $t = T-1, \ldots, 0$. We call this process a *backward iteration*, consisting of T *backward steps*. Second, iterate (15) forward, starting with $\mathbf{D}_0 = \mathbf{D}_{ss}$, to solve for the distributions $\mathbf{D}_t$ for $t = 1, \ldots, T-1$, by applying the transition matrices $\Lambda(\mathbf{v}_{t+1}, \mathbf{X}_t)$. We call this process a *forward iteration*, consisting of T *forward* steps. Finally, for each t, take the distribution-weighted sum $y(\mathbf{v}_{t+1}, \mathbf{X}_t)'\mathbf{D}_t$ of individual outcomes to obtain $\mathbf{Y}_t$ in (16).

Denote by $\mathcal{J}^{o,i}$ the Jacobian of $h(\mathbf{X})$ for output o and input i evaluated at steady state, so that $\mathcal{J}^{o,i}_{t,s} \equiv \frac{\partial Y^o_t}{\partial X^i_s}(\mathbf{X}_{ss})$. We now describe two methods for computing $\mathcal{J}^{o,i}$ up to a truncation horizon T.

3.2 Direct algorithm

The first way to compute the Jacobian $\mathcal{J}^{o,i}$ is by directly applying a finite-difference method. For each s, define the sequence $\mathbf{e}^s$ to have 0's everywhere except the sth entry, where there is a 1. Then we can compute

$$\mathcal{J}^{\cdot,i}_{\cdot,s} = \frac{1}{dx}\left[h(\{\mathbf{X}^i_{ss} + \mathbf{e}^s dx, \mathbf{X}^{-i}_{ss}\}) - \mathbf{Y}_{ss})\right] \tag{17}$$

for some small $dx > 0$, using the standard approach discussed above to calculate h given (14)–(16).[18] This calculates the s-th column of the Jacobian $\mathcal{J}^{o,i}$ for all $o = 1, \ldots, n_y$ at once.

This direct algorithm provides a general way of obtaining the Jacobians of the heterogeneous-agent block. Its advantage is that it is straightforward to implement in any model. Its drawback is that it can sometimes be slow. For example, as table 2 reveals, computing the two Jacobians

[18] In practice, it is useful to correct for the fact that unless the exact steady state is found, there is a slight difference numerically between $\mathbf{v}_{ss}$ and $v(\mathbf{v}_{ss}, \mathbf{X}_{ss})$, which can be blown up by repeated iteration and dividing by dx. One simple way to implement such a correction is to subtract $h(\mathbf{X}_{ss})$ rather than $\mathbf{Y}_{ss}$ in (17). Even better—but slightly more involved—is to do separate numerical differentiation of each backward step, always subtracting by $v(\mathbf{v}_{ss}, \mathbf{X}_{ss})$ rather than $\mathbf{v}_{ss}$ (and likewise for forward steps).

$\mathcal{J}^{\mathcal{K},r}$ and $\mathcal{J}^{\mathcal{K},w}$ with horizon T in the benchmark version of our Krusell and Smith (1998) model takes 26 seconds when $T = 300$. For the high-dimensional version, it takes more than 30 minutes. Fundamentally, the problem is that, with n_x inputs, the direct algorithm requires $n_x T^2$ backward steps and $n_x T^2$ forward steps. If each step takes time because the heterogeneous-agent block has many idiosyncratic states or is otherwise complex, the overall computation can be slow. This is especially likely to be a bottleneck if one needs to compute the Jacobian multiple times for different parameter values. This points to the need for an alternative, faster algorithm, which we introduce next.

3.3 Fake news algorithm

The second way to compute $\mathcal{J}^{o,i}$ is a new method we call the *fake news* algorithm.

To begin, since our objective is to compute Jacobians at the steady state, it is natural to work directly with the equations (14)–(16) linearized around the steady state. For notational simplicity, we start with the case of a one-dimensional input sequence $\mathbf{X} = (X_0, X_1, \ldots)$ and one-dimensional output sequence $\mathbf{Y} = (Y_0, Y_1, \ldots)$. The linearized system is then:

$$d\mathbf{v}_t = \mathbf{v}_{\mathbf{v}} d\mathbf{v}_{t+1} + \mathbf{v}_x dX_t \tag{18}$$

$$d\mathbf{D}_{t+1} = \left(\Lambda_{\mathbf{v}} d\mathbf{v}_{t+1} + \Lambda_x dX_t\right)' \mathbf{D}_{ss} + \Lambda'_{ss} d\mathbf{D}_t \tag{19}$$

$$dY_t = \left(\mathbf{y}_{\mathbf{v}} d\mathbf{v}_{t+1} + \mathbf{y}_x dX_t\right)' \mathbf{D}_{ss} + \mathbf{y}'_{ss} d\mathbf{D}_t \tag{20}$$

with terminal condition $d\mathbf{v}_T = 0$ and initial condition $d\mathbf{D}_0 = 0$.[19]

We now turn to the fake news algorithm. We will start by formally solving the system (18)-(20) for the Jacobian in a way that provides the foundation for rapid computation, and then proceed to discuss implementation details.

Two contributors to aggregate Y: individual and distribution effects. We can write a condensed version of (20) by defining $d\mathbf{y}_t \equiv \mathbf{y}_{\mathbf{v}} d\mathbf{v}_{t+1} + \mathbf{y}_x dX_t$. This results in:

$$dY_t = \underbrace{d\mathbf{y}'_t \mathbf{D}_{ss}}_{\text{individual}} + \underbrace{\mathbf{y}'_{ss} d\mathbf{D}_t}_{\text{distribution}} \tag{21}$$

Equation (21) splits the change dY_t in aggregate outcome into two effects. The *individual effect* is from the change $d\mathbf{y}_t$ in individual outcomes, while the *distribution effect* is from the change $d\mathbf{D}_t$ in the incoming distribution. We now study each of these effects separately, in response to a shock dX_s at time s, using (18)-(20).

Individual effect. If $t = s$, i.e. if the shock is in the current period, dX_s affects $d\mathbf{y}_t$ directly via $\mathbf{y}_x$. If $t < s$, i.e. if the shock is anticipated, then dX_s affects $d\mathbf{y}_t$ indirectly via $\mathbf{y}_{\mathbf{v}}$, through changing the

[19] Here, $\mathbf{v}_{\mathbf{v}}$ and $\mathbf{y}_{\mathbf{v}}$ are the $n_g \times n_g$ derivatives of v and y with respect to $\mathbf{v}$; $\mathbf{v}_x$ and $\mathbf{y}_x$ are the $n_g \times 1$ derivatives of v and y with respect to X; and $\Lambda_{\mathbf{v}}$ and Λ_x are the $n_g \times n_g \times n_g$ and $n_g \times n_g \times 1$ derivatives of Λ with respect to $\mathbf{v}$ and X.

value function tomorrow by $d\mathbf{v}_{t+1} = (\mathbf{v}_{\mathbf{v}})^{s-t-1}\mathbf{v}_x dX_s$. If $t > s$, i.e. if the shock is in the past, then there is no effect on $d\mathbf{y}_t$. Overall, if we define the scalars

$$\mathcal{Y}_u \equiv \begin{cases} 0 & \text{if } u < 0 \\ \mathbf{y}_x' \mathbf{D}_{ss} & \text{if } u = 0 \\ \left(\mathbf{y}_{\mathbf{v}}(\mathbf{v}_{\mathbf{v}})^{u-1}\mathbf{v}_x\right)' \mathbf{D}_{ss} & \text{if } u > 0 \end{cases} \tag{22}$$

then the individual effect is given by

$$d\mathbf{y}_t' \mathbf{D}_{ss} = \mathcal{Y}_{s-t} dX_s \tag{23}$$

Notice that this effect only depends on the distance between s and t, and not on s and t separately.

Distribution effect. The distribution effect is more complicated, since $d\mathbf{D}_t$ is given by the recursive relationship (19). We will proceed in two parts: we first derive an expression for the non-recursive term $\left(\Lambda_{\mathbf{v}} d\mathbf{v}_{t+1} + \Lambda_x dX_t\right)' \mathbf{D}_{ss}$ in (19), and we then proceed recursively to characterize the sequence $\{d\mathbf{D}_t\}$.

Start with the non-recursive term. If $t = s$, there is the direct effect $\Lambda_x' \mathbf{D}_{ss} dX_s$. If $t < s$, then the value function tomorrow changes by $d\mathbf{v}_{t+1} = (\mathbf{v}_{\mathbf{v}})^{s-t-1}\mathbf{v}_x dX_s$, and there is the indirect effect $\left(\Lambda_{\mathbf{v}} d\mathbf{v}_{t+1}\right)' \mathbf{D}_{ss}$. If $t > s$, there is no effect. If we define the $n_g \times 1$ vectors

$$\mathcal{D}_u \equiv \begin{cases} 0 & \text{if } u < 0 \\ \Lambda_x' \mathbf{D}_{ss} & \text{if } u = 0 \\ \left(\Lambda_{\mathbf{v}}(\mathbf{v}_{\mathbf{v}})^{u-1}\mathbf{v}_x\right)' \mathbf{D}_{ss} & \text{if } u > 0 \end{cases} \tag{24}$$

then the non-recursive term is $\mathcal{D}_{s-t} dX_s$.

Now, starting with the initial condition $d\mathbf{D}_0 = 0$ and recursively substituting the expression for the non-recursive term into (19), we can derive the sequence $\{d\mathbf{D}_t\}$:

$$\begin{aligned} d\mathbf{D}_1 &= \mathcal{D}_s dX_s \\ d\mathbf{D}_2 &= (\mathcal{D}_{s-1} + \Lambda_{ss}' \mathcal{D}_s) dX_s \\ d\mathbf{D}_3 &= (\mathcal{D}_{s-2} + \Lambda_{ss}' \mathcal{D}_{s-1} + (\Lambda_{ss}')^2 \mathcal{D}_s) dX_s \\ \cdots &= \cdots \\ d\mathbf{D}_t &= (\mathcal{D}_{s-t+1} + \Lambda_{ss}' \mathcal{D}_{s-t+2} + \ldots + (\Lambda_{ss}')^{t-1} \mathcal{D}_s) dX_s \end{aligned} \tag{25}$$

If we define the $1 \times n_g$ vectors $\mathcal{P}_u'$ as

$$\mathcal{P}_u' \equiv \begin{cases} 0 & \text{if } u < 0 \\ \mathbf{y}_{ss}'(\Lambda_{ss}')^u & \text{if } u \geq 0 \end{cases} \tag{26}$$

then we can combine (25) with (26) to write the general distribution effect as

$$\mathbf{y}'_{ss} d\mathbf{D}_t = \left(\mathcal{P}'_0 \mathcal{D}_{s-t+1} + \mathcal{P}'_1 \mathcal{D}_{s-t+2} + \ldots + \mathcal{P}'_{t-1} \mathcal{D}_s\right) dX_s \tag{27}$$

We can interpret the $\mathcal{P}'_u$ vectors as linear functionals, which when applied to some perturbation dD_τ to the distribution at time τ, give the effect $dY_{\tau+u}$ on the aggregate outcome u periods later.[20] In equation (27), these functionals are applied to the perturbations to $dD_1, \ldots, dD_t$ from anticipating the shock in periods $0, \ldots, t-1$.

Combining individual and distribution effects. Substituting both (23) and (27) into (21) gives us an expression for the Jacobian entries $\mathcal{J}_{t,s}$:

$$dY_t = \mathcal{J}_{t,s} dX_s = \left(\mathcal{Y}_{s-t} + \mathcal{P}'_0 \mathcal{D}_{s-t+1} + \mathcal{P}'_1 \mathcal{D}_{s-t+2} + \ldots + \mathcal{P}'_{t-1} \mathcal{D}_s\right) dX_s \tag{28}$$

This expression *only* uses the three objects defined in (22)–(26): the scalars $\mathcal{Y}_u$, the $n_g \times 1$ vectors $\mathcal{D}_u$, and the $1 \times n_g$ vectors $\mathcal{P}'_u$, for all $u = 0, \ldots, T-1$. In short, it reduces the problem of finding T^2 entries of $\mathcal{J}_{t,s}$ to the problem of finding these $3T$ objects.

Fake news matrix and recursion. Inspecting (28), we note that $\mathcal{J}_{t-1,s-1}$ and $\mathcal{J}_{t,s}$ are identical, except that the latter adds an additional term $\mathcal{P}'_{t-1} \mathcal{D}_s$. This motivates us to define the matrix $\mathcal{F}$ as:

$$\mathcal{F}_{t,s} \equiv \begin{cases} \mathcal{Y}_s & t = 0 \\ \mathcal{P}'_{t-1} \mathcal{D}_s & t > 0 \end{cases} \tag{29}$$

We can then use this matrix to recursively construct the Jacobian, as:

$$\mathcal{J}_{t,s} = \begin{cases} \mathcal{F}_{t,s} & t = s = 0 \\ \mathcal{F}_{t,s} + \mathcal{J}_{t-1,s-1} & \text{otherwise} \end{cases} \tag{30}$$

We call $\mathcal{F}$ the *fake news matrix*, since it can be interpreted as the Jacobian for "fake news" shocks. In a fake news shock, agents learn at date 0 about some shock dX_s at date s, but then learn at date 1 that dX_s will never actually materialize (except in the special case $s = 0$, where the shock has already occurred). A fake news shock generally has consequences for all periods after date 0, since agents' date-0 behavior responds to the fake news, and this alters the distribution going forward.

An appealing feature of fake news shocks is that at any given t, there is only an individual effect or distribution effect, but never both. At date 0, as always, there is only the individual effect $\mathcal{Y}_s$, since the distribution has not yet changed. From date 1 onward, no shock is expected, implying that forward-looking variables revert to steady state and the only effect is through the distribu-

[20] In the literature on control theory, the matrix consisting of rows $\mathcal{P}'_0, \mathcal{P}'_1, \ldots$ is sometimes called the *observability matrix*. This concept is also used by Reiter (2010) and Ahn et al. (2018). We avoid this terminology because observability in a different sense—observability to the econometrician—is central to our discussion of estimation in section 5.

tion. This effect comes from the change in date-1 distribution $\mathcal{D}_s$ caused by agents' anticipatory behavior at date 0, which then propagates and determines dY_t according to $\mathcal{P}'_{t-1}$.

As (30) shows, the fake news matrix $\mathcal{F}$ is all that is necessary to construct the Jacobian $\mathcal{J}$. An element $\mathcal{J}_{t,s}$ contains the response at date t to an *actual* date-s news shock: where agents learn at date 0 about a shock dX_s, and the shock does indeed happen at s. This is the same as the response at date $t-1$ to news shock about date $s-1$, except that it is anticipated for an extra period, and the effect of this anticipation persists through the distribution. This extra effect is exactly the effect of a date-s fake news shock, and hence $\mathcal{J}_{t,s} = \mathcal{F}_{t,s} + \mathcal{J}_{t-1,s-1}$.

Algorithm. We are now ready to describe our general algorithm. First, we generalize our notation to the case with multiple outputs o and inputs i. The $\mathcal{Y}_u$ defined in (22) depends on both the input shock dX^i and the output of interest dY^o, so we write it as $\mathcal{Y}_u^{o,i}$. In contrast, the $\mathcal{D}_u$ defined in (24) depends only on the input shock dX^i, and the $\mathcal{P}'_u$ defined in (26) depends only on the output of interest dY^o, so we write them as $\mathcal{D}_u^i$ and $(\mathcal{P}_u^o)'$ respectively.

We can then obtain $\mathcal{J}$ with the following four steps:

1. For each input i, perform a *backward iteration* with T steps, starting from a small shock $dX^i \equiv dX^i_{T-1}$ in the final period, calculating for each $u = 0, \ldots, T-1$

$$\mathcal{D}_u^i = \frac{\Lambda(\mathbf{v}_{T-u}, \mathbf{X}_{T-u-1})'\mathbf{D}_{ss} - \mathbf{Y}_{ss}}{dX^i} \tag{31}$$

$$\mathcal{Y}_u^{o,i} = \frac{y^o(\mathbf{v}_{T-u}, \mathbf{X}_{T-u-1})'\mathbf{D}_{ss} - \mathbf{Y}_{ss}}{dX^i} \quad \forall o \tag{32}$$

   and stacking the results in a $n_g \times T$ matrix $\mathcal{D}^i$ and a length-T vector $\mathcal{Y}^{o,i}$.

2. For each output o, perform a *"transpose" forward iteration* with $T-1$ steps, initializing to $\mathcal{P}_0^o = \mathbf{y}_{ss}^o$, recursively calculating for each $u = 1, \ldots, T-2$

$$\mathcal{P}_u^o = \Lambda_{ss}\mathcal{P}_{u-1}^o \tag{33}$$

   and stacking the results in an $n_g \times (T-1)$ matrix $\mathcal{P}^o$.

3. For each output-input pair (o,i), construct the fake news matrix $\mathcal{F}^{o,i}$, with $\mathcal{Y}^{o,i}$ in row $t=0$, and the matrix product $(\mathcal{P}^o)'\mathcal{D}^i$ in rows $t = 1, \ldots, T-1$.

4. For each output-input pair (o,i), apply recursion (30) to construct the Jacobian $\mathcal{J}^{o,i}$ from $\mathcal{F}^{o,i}$.

Step 1 is simply using numerical differentiation to apply (22) and (24).[21] Step 2 follows directly from (26), and we call it a "transpose" forward iteration since it has the same form as (15), except that here we multiply by $(\Lambda'_{ss})' = \Lambda_{ss}$ rather than by Λ'_{ss}.

[21] As discussed for the direct method in footnote 18, it can be useful to correct for any slight numerical difference between $\mathbf{v}_{ss}$ and $v(\mathbf{v}_{ss}, \mathbf{X}_{ss})$.

Table 2: Direct and fake news algorithms to compute 300 × 300 Jacobians.

	Krusell-Smith	HD Krusell-Smith	one-asset HANK	two-asset HANK
Direct	**26 s**	**1939 s**	**176 s**	**2107 s**
step 1 (backward)	16 s	1338 s	150 s	1291 s
step 2 (forward)	10 s	601 s	27 s	815 s
Fake news	**0.104 s**	**8.429 s**	**0.646 s**	**5.697 s**
step 1 (backward)	0.067 s	5.433 s	0.525 s	5.206 s
step 2 (forward)	0.010 s	1.546 s	0.021 s	0.122 s
step 3	0.023 s	1.445 s	0.092 s	0.346 s
step 4	0.004 s	0.004 s	0.008 s	0.023 s
Gridpoints n_g	3,500	250,000	3,500	10,500
Inputs n_x	2	2	4	5
Outputs n_y	2	2	4	4
Jacobians $n_x \times n_y$	4	4	16	20

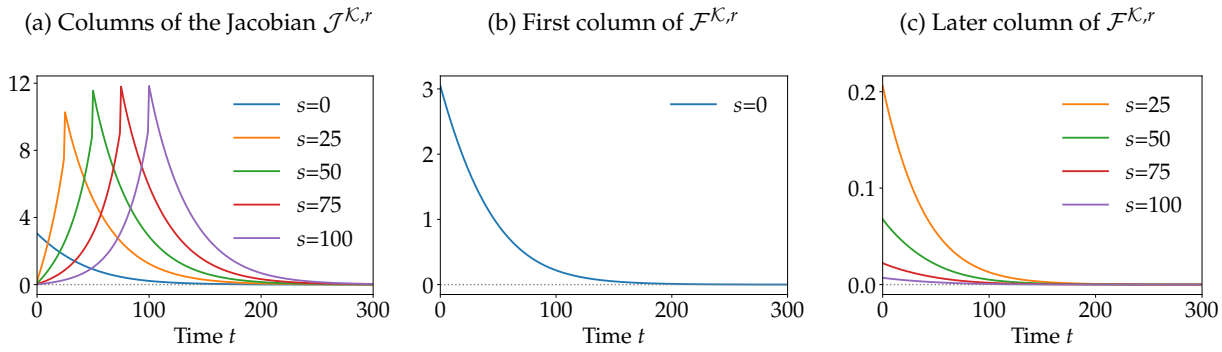
Efficiency. Table 2 displays the time it takes to compute $\mathcal{J}$s for the heterogeneous-agent block of each of our three benchmark models: the Krusell-Smith model already introduced, a one-asset HANK model with endogenous labor described in Appendix A.2, and a two-asset HANK model described in Appendix A.3. The speed-up from using the fake news rather than the direct algorithm is very large in all cases: a factor of over 200 for Krusell-Smith and one-asset HANK, and a factor of over 300 for two-asset HANK.

What is the source of this very large efficiency gain? As we discussed in section 3.2, when there are n_x inputs and n_y outputs, the direct algorithm requires $n_x T^2$ backward steps and $n_x T^2$ forward steps. By contrast, the fake news algorithm requires $n_x T$ backward steps and $n_y(T-1)$ "transpose" linearized forward steps, reducing effort in steps 1 and 2 by a factor of around T, which in our application is $T = 300$.[22]

There are two additional steps required for the fake news algorithm, steps 3 and 4. Step 3 involves the multiplication of $T \times n_g$ and $n_g \times T$ matrices, which has a cost proportional to $n_g T^2$ for each input-output pair—but since matrix multiplication is implemented extremely efficiently by standard numerical libraries, this is less of a bottleneck overall than the backward iteration in step 1, especially for models like the two-asset HANK where backward iteration is especially complex. Step 4 is even faster, since it is a simple recursion on $T \times T$ matrices.

[22]The transpose forward iteration in step 2 takes far less time than the backward iteration in step 1, especially for the more complex models, because it only requires repeatedly multiplying by Λ_{ss}—which can be split into multiplication by a small transition matrix for the exogenous state, and multiplication by a highly sparse matrix with policies for endogenous states, both of which we implement efficiently. In principle, similar improvements could be possible in step 1 if we used automatic rather than numerical differentiation in a way that took advantage of the sparsity of the backward operation, but here we focus on numerical differentiation to make the method immediately applicable to a wide variety of complex models.

Figure 2: Jacobian $\mathcal{J}^{\mathcal{K},r}$ and fake news matrix $\mathcal{F}^{\mathcal{K},r}$ in the Krusell and Smith (1998) model.



Accuracy. In practical implementations, one may wish to check that both the fake news and the direct algorithm compute the correct Jacobian $\mathcal{J}$. A simple test of this property is to compare numerically the outcomes of both algorithms. We conduct such an accuracy check in appendix B.2, with results in figure B.1. This check makes clear that our two algorithms compute exactly the same Jacobians—the fake news algorithm just does so much faster. Furthermore, the code we provide largely automates the fake news algorithm, so that it should be easy to apply for any given model.

3.4 Structure of Jacobians

In addition to being a useful computational tool, the fake news matrix sheds light on an important structural property of the Jacobians of heterogeneous-agent blocks, which we will now discuss.

Panel (a) of figure 2 displays several columns of the Jacobian $\mathcal{J}^{\mathcal{K},r}$ for the Krusell-Smith model of section 2. One striking feature is that these columns converge to a regular pattern around the main diagonal: the $s = 50$ impulse response around $t = 50$ is almost the same as the $s = 75$ and $s = 100$ impulse responses around $t = 75$ and $t = 100$. In other words, if the shock is anticipated far enough in advance, all impulse responses are just shifted versions of each other.

This turns out to be a general property of Jacobians for heterogeneous-agent blocks, which we call *asymptotic time invariance* and formalize in the following proposition.

Proposition 1 (Asymptotic time invariance)**.** *Consider a heterogeneous-agent block written using* (14)-(16)*. Assume that there is a unique ergodic distribution, and that* $\mathbf{v_v}$ *has all eigenvalues inside the unit circle. Then* $\mathcal{F}_{t,s}$ *converges exponentially to 0 as* $s + t \to \infty$*. Moreover, the Jacobian* $\mathcal{J}$ *is asymptotically time-invariant, that is, there exists a two-sided sequence* $A = (A_\tau)$, $\tau \in \mathbb{Z}$ *such that*

$$\mathcal{J}_{t+\tau,t} \to A_\tau \quad \text{as } t \to \infty$$

for any $\tau \in \mathbb{Z}$*.*

The key to the proof is equation (30).[23] Since $\mathcal{J}_{t,s} = \mathcal{F}_{t,s} + \mathcal{J}_{t-1,s-1}$, as we move down the

[23]Observe that the conditions for proposition 1 are weak: it only requires ergodicity of the distribution and the

diagonals of the Jacobian, we pick up terms from the fake news matrix. Panels (b) and (c) of figure 2 illustrate these terms in our example: (b) the large responses $\mathcal{F}_{\cdot 0}$, where households actually earn a high r at date 0, and (c) some other $\mathcal{F}_{\cdot s}$ columns, where households save at date 0 due to the fake promise of high future r. We see that $\mathcal{F}_{t,s}$ goes to zero both for high t (the effect of date-0 behavior through the distribution dies away) and for high s (the effect of far-out shocks on date-0 behavior dies away). It follows that $\mathcal{J}_{t,s} \approx \mathcal{J}_{t-1,s-1}$ for high t and s.

Proposition 1 provides a useful theoretical benchmark against which check the quality of any given implementation of a (direct or fake news) algorithm in practice. We will also return to it in section 6, where it will form the basis of our determinacy criterion.

4 SHADE models and their Jacobians

So far we have shown how, in the Krusell and Smith (1998) model, one can apply the chain rule to reduce the problem of finding impulse responses to that of finding the Jacobian of heterogeneous-agent blocks, and we have written down a fast algorithm to compute these Jacobians. We now show how to apply this idea systematically to a large class of models that encompasses many commonly-used heterogeneous-agent models. We will call these *SHADE models*.

We continue to write $\mathbf{Z}$ for the path of exogenous shocks, but now allow $\mathbf{Z}_t$ to be a $n_z \times 1$ vector at each t. We replace the endogenous variable $\mathbf{K}$ with the more generic $\mathbf{U}$, where $\mathbf{U}_t$ is an $n_u \times 1$ vector of "unknowns" at each t. Equilibrium is characterized by a system of equations in sequence space:

$$\mathbf{H}(\mathbf{U}, \mathbf{Z}) = 0 \tag{34}$$

As in section 2, impulse responses to first order around the steady state are given by $d\mathbf{U} = -\mathbf{H}_\mathbf{U}^{-1}\mathbf{H}_\mathbf{Z} d\mathbf{Z}$, which requires that we calculate the Jacobians $\mathbf{H}_\mathbf{U}$ and $\mathbf{H}_\mathbf{Z}$.

One difficulty that emerges in the general case is that the dimensionality can grow large enough that solving the linear system is itself a bottleneck. When the system is truncated at time T, $\mathbf{H}_\mathbf{U}$ has dimension $n_u T \times n_u T$. Quantitative DSGE models often have dozens of endogenous variables, and if all these are included as unknowns in (34), then $n_u T$ can potentially be as high as 10,000. This brings back the high dimensionality that our approach is intended to circumvent.

The literature solving MIT shocks typically reduces dimensionality by iterating on only a subset of endogenous variables, and expressing all other variables as functions of these. We use a similar approach to reduce n_u. In our case, however, we also need to calculate the Jacobian of the resulting system. To do so efficiently, we represent the model underlying $\mathbf{H}$ explicitly as a directed acyclic graph (DAG), and compute Jacobians using *forward accumulation* along the graph, a technique from the algorithmic differentiation literature (e.g. Griewank and Walther 2008).

eigenvalues of $\mathbf{v}_\mathbf{v}$ to lie inside the unit circle—the latter of which is satisfied whenever $\mathbf{v}$ is derived from a Bellman equation (12) that is a contraction mapping.

4.1 SHADE models

SHADE models consist of a combination of heterogeneous-agent blocks, as per definition 1, and *simple blocks*, as per the following definition:

Definition 2. A *simple block* maps input sequences $\mathbf{X}^i$, for $i = 1, \ldots, n_x$ to output sequences $\mathbf{Y}^o$ for $o = 1, \ldots, n_y$. Stacking the $\mathbf{X}_t^i$'s into a vector $\mathbf{X}_t$, and the $\mathbf{Y}_t^o$'s into a vector $\mathbf{Y}_t$, there must exist $k, l \in \mathbb{N}$ and a time-invariant function h such that $\mathbf{Y}_t$ is only a function of neighboring $\mathbf{X}_t$'s, that is,

$$\mathbf{Y}_t = h(\mathbf{X}_{t-k}, \ldots, \mathbf{X}_{t+l})$$

Simple blocks capture typical aggregate relationships in dynamic macro models. For instance, a neoclassical firm sector can be represented as a simple block mapping $\mathbf{X}_t = (K_t, Z_t)$ to $\mathbf{Y}_t = (Y_t, r_t, w_t)$. Combining such a sector with a heterogeneous-agent block mapping $\mathbf{X}_t = (r_t, w_t)$ to $\mathbf{Y}_t = \mathcal{K}_t(\{r_s, w_s\})$, as well as a simple block mapping $\mathbf{X}_t = (\mathcal{K}_t, K_t)$ to market clearing $\mathbf{Y}_t = \mathcal{K}_t - K_t$, we obtain the Krusell-Smith model of section 2.

We define SHADE models as a generalization of this representation. These models map *shocks* (like Z_t) and *unknowns* (like K_t) to *targets* (like asset market clearing) along a directed acyclic graph.

Definition 3. A *Sequence-space Heterogeneous-Agent Dynamic-Equilibrium (SHADE) model* is:

1. A set of sequence indices $\mathcal{N} = \mathcal{Z} \cup \mathcal{U} \cup \mathcal{O}$, where $\mathcal{Z}$ are *exogenous shocks*, $\mathcal{U}$ are *unknowns*, $\mathcal{O}$ are *outputs*, and $\mathcal{H} \subset \mathcal{O}$ are *targets.*

2. A set of *blocks*, each either simple or heterogeneous-agent blocks, indexed by $\mathcal{B}$, where each block $b \in \mathcal{B}$ has *inputs* $\mathcal{I}_b \subset \mathcal{N}$ and *outputs* $\mathcal{O}_b \subset \mathcal{O}$. Each output $o \in \mathcal{O}$ belongs to exactly one block. For each output $o \in \mathcal{O}_b$, block b provides a function $h^o(\{\mathbf{x}^i\}_{i \in \mathcal{I}_b})$ mapping the block's input sequences to this output sequence. The *directed graph* of blocks, formed by drawing an edge from b to b' whenever some output $o \in \mathcal{O}_b$ is used as an input $o \in \mathcal{I}_{b'}$, must be *acyclic*.

3. The number of unknowns and targets must be equal, that is, $n_u = n_h$.

An *equilibrium* of a SHADE model, given sequences $\{\mathbf{X}^i\}_{i \in \mathcal{Z}}$ for the exogenous shocks, is a set of sequences $\{\mathbf{X}^i\}_{i \in \mathcal{U} \cup \mathcal{O}}$ such that

1. $\mathbf{X}^o = h^o(\{\mathbf{X}^i\}_{i \in \mathcal{I}_b})$ for any output $o \in \mathcal{O}$.

2. $\mathbf{X}^o = 0$ for any target $o \in \mathcal{H}$.

A *steady state equilibrium* is an equilibrium in which all sequences are constant over time, $\mathbf{X}_t^i = \mathbf{X}_{ss}^i$ for all $i \in \mathcal{N}$.

An important part of this definition is that the blocks form a *directed acyclic graph (DAG)* if we draw an edge from b to b' to represent the dependency of b' on the output of b. A directed

acyclic graph always has a *topological sort*, meaning that there is some ordering $b^1, \ldots, b^{n_b}$ of the blocks—not necessarily unique—such that the inputs of each later block in the ordering are either in $\mathcal{Z} \cup \mathcal{U}$ or outputs of earlier blocks. If we start with sequences $\{\mathbf{X}^i, i \in \mathcal{Z} \cup \mathcal{U}\}$ for all shocks and unknowns, we can evaluate the blocks along this ordering: first applying h^o for each $o \in \mathcal{O}_{b^1}$, then for each $o \in \mathcal{O}_{b^2}$, and so on. When we are done, we will have calculated all outputs, including the targets $\mathcal{H}$.

We can view this procedure as a mapping from exogenous shocks and unknowns $\{\mathbf{X}^i\}_{i \in \mathcal{Z} \cup \mathcal{U}}$ to targets $\{\mathbf{X}^o\}_{o \in \mathcal{H}}$. We write this mapping in more condensed form as $\mathbf{H}(\mathbf{U}, \mathbf{Z})$, where $\mathbf{U}$ is defined as the stacked vector of unknown sequences $\{\mathbf{X}^i\}_{i \in \mathcal{U}}$, $\mathbf{Z}$ is defined as the stacked vector of exogenous sequences $\{\mathbf{X}^i\}_{i \in \mathcal{Z}}$, and $\mathbf{H}(\mathbf{U}, \mathbf{Z})$ itself is the implied stacked vector of targets $\{\mathbf{X}^i\}_{i \in \mathcal{H}}$. Since the procedure satisfies $\mathbf{X}^o = h^o(\{\mathbf{X}^i\}_{i \in \mathcal{I}_b})$ by construction, equilibrium is then equivalent to

$$\mathbf{H}(\mathbf{U}, \mathbf{Z}) = 0 \tag{35}$$

Figure 3a visualizes the DAG for the Krusell-Smith model, which has two blocks (neoclassical firms, and heterogeneous households "HA"), one exogenous shock (productivity Z), one unknown (capital K), four outputs (capital return r, wage w, household savings $\mathcal{K}$, asset market clearing H), and one target (asset market clearing H).[24]

One-asset HANK model. Figure 3b illustrates the DAG of our second example: a one-asset HANK model similar to McKay et al. (2016). This model combines standard NK elements—sticky prices, flexible wages, and a Taylor rule for monetary policy, but no capital—with a one-asset incomplete market HA household sector where labor supply is endogenous. It is introduced formally in Appendix A.2.

As one would expect, the DAG for this model is more complicated than for Krusell-Smith. There are three unknowns (wages w, output Y, and inflation π) and three targets (a Phillips curve condition H_1, labor market clearing H_2, and asset market clearing H_3). We introduce two exogenous shocks (productivity Z and Taylor rule intercept r^*).

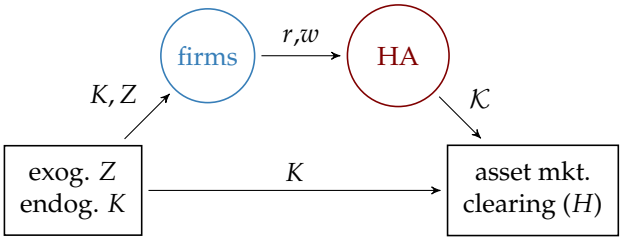
The DAG makes it easy to visualize some of the dependencies embedded in the model: for instance, the dividends from firms are distributed to households (according to a certain rule), so the output d of the firm block is an input to the HA block. Similarly, the real interest rate r affects the taxes required for the government to achieve its balanced-budget target, so r is an input to the fiscal block, which has an output τ that is an input to the HA block.

There are many equivalent DAG representations of the same model, depending on which model equations we solve out to write as simple blocks, and which equations we include only as targets. This particular representation is chosen to minimize the number of targets $n_u = n_h$. A lower n_u lowers the size of the $\mathbf{H_U}$ matrix, and this generally makes the system easier to solve.

[24] Not visualized are firm production Y or household consumption C, which could be additional outputs of the firm and HA blocks, respectively, but are not strictly necessary since we are using asset rather than the goods market clearing to define equilibrium.

Figure 3: DAG representations of our Krusell-Smith and one-asset HANK economies

(a) KS economy



(b) One-asset HANK economy

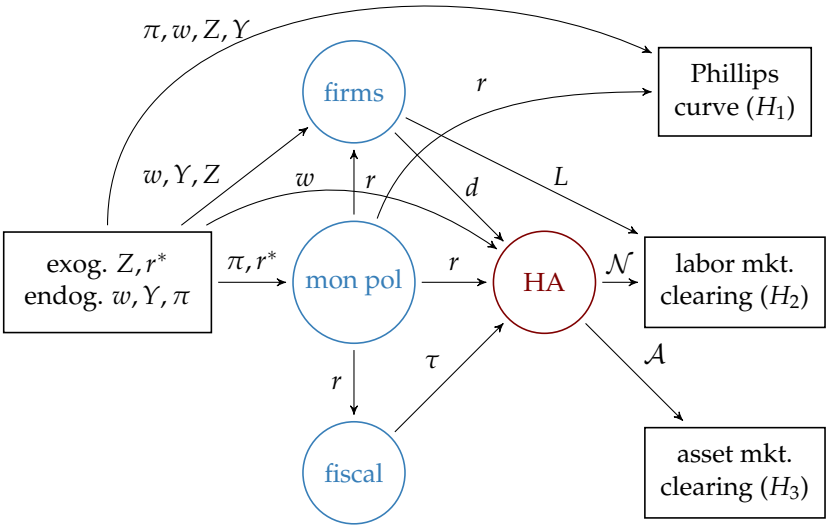
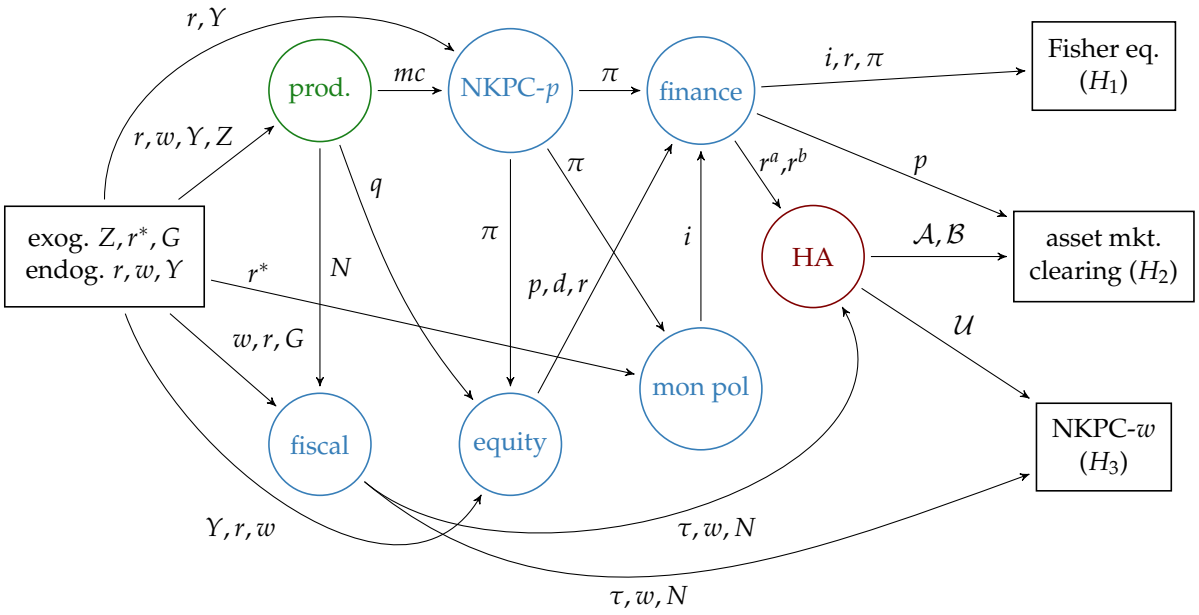


Figure 4: DAG representation of our two-asset HANK economy



Two-asset HANK model. Figure 3 illustrates the DAG for our third example: a two-asset HANK model with household side similar to Kaplan et al. (2018), described in detail in appendix A.3. For households, the model features liquid and illiquid assets with convex adjustment costs of portfolio adjustment. On the supply side, it features wage as well as price rigidities, as well as capital with quadratic adjustment costs. Hence investment follows the standard q theory equations.[25]

Monetary policy follows a standard interest rate rule. The government levies a distortionary labor income tax to finance its debt and its expenditure on the final good. We assume a balanced budget. Some government bonds are held by households directly (along with firm equity) as illiquid assets, and the rest are transformed into liquid assets by a competitive financial intermediary. This liquidity transformation incurs a proportional cost, which determines the equilibrium spread between liquid and illiquid assets in all periods along perfect-foresight paths.

Scope of SHADE models. Our definition of SHADE models is general enough to encompass the vast majority of models in the emerging heterogeneous-agent literature. This includes, for instance, most heterogeneous-agent New Keynesian models, most models of firm life cycles (Hopenhayn 1992, Hopenhayn and Rogerson 1993), lumpy investment models as in Khan and Thomas (2008), pricing models as in Golosov and Lucas (2007), or overlapping generation models as in Conesa and Krueger (1999). The main omissions are in the class of models where decentralized arrangements mean that agents in different parts of the distribution interact directly with each

[25] These q theory equations are embedded in the "production" block of the DAG. Technically, this block (displayed in green) is what we called a solved block, which internally solves for the path of q and capital K given its inputs. Solved blocks can speed up the evaluation of the model and simplify the DAG representation (see appendix B.3).

other, in a way that cannot be intermediated through a limited number of aggregate variables. For instance, in the model of on-the-job search in Burdett and Mortensen (1998), agents take the full distribution of wages as an input to their decision problem, and it is impossible to represent this via a DAG of feasible dimension.[26]

4.2 Jacobians of SHADE models and impulse responses

As in section 2, if $\mathbf{H}$ is differentiable, we can apply the implicit function theorem to equation (35) to solve for the first order response of $\mathbf{U}$, $d\mathbf{U}$, in response to a transitory shock to $\mathbf{Z}$, $d\mathbf{Z}$, as:

$$d\mathbf{U} = -\mathbf{H}_{\mathbf{U}}^{-1}\mathbf{H}_{\mathbf{Z}}d\mathbf{Z} \tag{36}$$

In general, $\mathbf{H}_{\mathbf{U}}$ and $\mathbf{H}_{\mathbf{Z}}$ are complicated objects that depend on the entire model. The crucial advantage of the DAG structure of SHADE models, however, is that we can use *forward accumulation*, a tool from the literature on algorithmic differentiation (Griewank and Walther 2008), to combine the Jacobians of individual blocks to build up these objects. The key idea behind forward accumulation is to apply the chain rule in the same order that we would evaluate a function itself.[27]

Total Jacobians J. To start, we need a new concept. For any exogenous shock or unknown $i \in \mathcal{Z} \cup \mathcal{U}$ and any output o, let the $\mathbf{J}^{o,i}$ denote the *total Jacobian* of o with respect to i when o is evaluated along the DAG. For instance, in the one-asset HANK model in figure 3b, the total Jacobian $\mathbf{J}^{\mathcal{N},w}$ of household labor supply with respect to wages combines two forces: the direct effect of w on household decisions, and the indirect effect working through the influence of w on firm profits and therefore the dividends d received by households. This is in contrast to $\mathcal{J}^{\mathcal{N},w}$, which is a partial Jacobian that captures only the direct effect.

To obtain $\mathbf{J}^{o,i}$ through forward accumulation, we follow a procedure similar to that of the previous section. We initialize $\mathbf{J}^{i,i}$ to the identity for each $i \in \mathcal{Z} \cup \mathcal{U}$. We then go through blocks following a topological sort $b^1, \ldots, b^{n_b}$, and for each block b we evaluate the following for all $o \in \mathcal{O}_b$ and $i \in \mathcal{Z} \cup \mathcal{U}$:

$$\mathbf{J}^{o,i} = \sum_{m \in \mathcal{I}_b} \mathcal{J}^{o,m}\mathbf{J}^{m,i} \tag{37}$$

Fundamentally, this is just the chain rule: for each input m, (37) takes the product of the partial Jacobian $\mathcal{J}^{o,m}$ with the already-calculated total derivative $\mathbf{J}^{m,i}$ of m with respect to i. (When $m = i$, then the latter is the identity and the term is just the partial Jacobian $\mathcal{J}^{o,i}$.) The benefit of building up the $\mathbf{J}^{o,i}$ progressively via forward accumulation, however, is that the chain rule is applied in an efficient way, without redundant computations.

[26] Note, however, that SHADE models do not require approximate aggregation: agents are allowed to take into account the effect on their decision problem from entire *sequences* of aggregate variables, and these sequences in turn may be determined by a high-dimensional subspace of the distribution.

[27] In actual computations, the methods in this section will be applied on Jacobians that are truncated to some horizon $T \times T$, except when using our sparse methods for simple blocks described in the next section, where they can be exact.

Table 3: Computing times for $\mathbf{G}$.

	Krusell-Smith	one-asset HANK	two-asset HANK
Without directed acyclic graph (DAG)	**44.3 ms**	**251.1 ms**	**2072.8 ms**
With DAG	**6.8 ms**	**96.9 ms**	**402.1 ms**
step 1 (forward accumulate $\mathbf{H_U}$ and $\mathbf{H_Z}$)	1.3 ms	15.1 ms	131.4 ms
step 2 (compute $\mathbf{G}^{\mathcal{U},\mathcal{Z}} = \mathbf{H_U^{-1}H_Z}$)	2.6 ms	43.4 ms	68.8 ms
step 3 (forward accumulate for all $\mathbf{G}^{\mathcal{U},\mathcal{Z}}$)	2.8 ms	38.5 ms	201.9 ms
No. of unknowns (without DAG)	3	7	18
No. of unknowns (with DAG)	1	3	3
No. of exogenous shocks	1	3	7

General equilibrium Jacobians G. Using the $\mathbf{J}$ matrices, we can now obtain a full linear characterization of general equilibrium in response to shocks $d\mathbf{Z}$. We will write $\mathbf{G}^{o,z}$ to denote the linear mapping from shock $z \in \mathcal{Z}$ to output $o \in \mathcal{O}$. These $\mathbf{G}$ matrices are the *general equilibrium* Jacobians that map from *any* sequence of exogenous shocks, up to the truncation horizon T, to impulse responses for all variables o of interest. Such a complete first-order characterization of equilibrium will prove invaluable when estimating shock processes in the next section.

In what follows, we adopt the notational convention that $\mathbf{G}^{\mathcal{U},z}$ is $\mathbf{G}^{u,z}$ stacked for all $u \in \mathcal{U}$, $\mathbf{G}^{o,\mathcal{Z}}$ is $\mathbf{G}^{o,z}$ stacked for all $z \in \mathcal{Z}$, and so on. We have $d\mathbf{X}^o = \mathbf{G}^{o,\mathcal{Z}} d\mathbf{Z}$ for all output variables $o \in \mathcal{O}$ and stacked shocks $d\mathbf{Z}$.

Algorithm for G. We now describe the algorithm to obtain $\mathbf{G}$. To start, note that by (36), $\mathbf{G}^{\mathcal{U},\mathcal{Z}} = -\mathbf{H_U^{-1}H_Z}$. The first step is therefore to use forward accumulation on $\mathbf{J}$'s (37) to obtain both $\mathbf{H_U} = \mathbf{J}^{\mathcal{H},\mathcal{U}}$ and $\mathbf{H_Z} = \mathbf{J}^{\mathcal{H},\mathcal{Z}}$. We then compute $\mathbf{G}^{\mathcal{U},\mathcal{Z}} = -\mathbf{H_U^{-1}H_Z}$. Finally, initializing with $\mathbf{G}^{\mathcal{U},\mathcal{Z}}$ (and $\mathbf{G}^{\mathcal{Z},\mathcal{Z}}$ equal to the identity), we perform forward accumulation[28] to obtain all other $\mathbf{G}^{o,\mathcal{Z}}$:

$$\mathbf{G}^{o,\mathcal{Z}} = \sum_{m \in \mathcal{I}_b} \mathcal{J}^{o,m} \mathbf{G}^{m,\mathcal{Z}} \tag{38}$$

Interpreting G in terms of individual impulse responses. Each $\mathbf{G}^{o,\mathcal{Z}}$ has $n_z T$ columns, each of which can be interpreted as the impulse response of o to some news shock. One way to think about this approach, therefore, is that we are *simultaneously* calculating $n_z T$ general equilibrium impulse responses. For our Krusell-Smith, one-asset HANK, and two-asset HANK models, $n_z T$ is $1 \times 300 = 300$, $3 \times 300 = 900$, and $3 \times 300 = 900$, respectively.

Appendix B.4 discusses a related method that only calculates an individual impulse response. Since $\mathbf{G}$ consists of hundreds of such impulse responses, it is no surprise that calculating $\mathbf{G}$ is

[28] An alternative approach for this final step is to directly write $\mathbf{G}^{o,\mathcal{Z}} = \mathbf{J}^{o,\mathcal{U}}\mathbf{G}^{\mathcal{U},\mathcal{Z}} + \mathbf{J}^{o,\mathcal{Z}}$, calculating $\mathbf{J}^{o,\mathcal{U}}$ and $\mathbf{J}^{o,\mathcal{Z}}$ for all o in the first forward accumulation step. Depending on the structure of the DAG, this may be more or less computationally intensive; in practice, we have generally found that it is costlier.

costlier. But interestingly, it is not too much more expensive: the times for our algorithm to calculate $\mathbf{G}$ in table 3 are only above 5 times higher than those in table B.1. This efficiency in calculating $\mathbf{G}$ is possible because we only need to calculate $\mathbf{H_U}$ once, independent of shocks, and do all other computations stacked together in matrices.

Comparing table 3 to table 2, we see that computing $\mathbf{G}$s is, in each of our cases, significantly cheaper than applying the fake news algorithm to obtain $\mathcal{J}$s for the heterogeneous-agent block: for instance, it takes about 0.1 seconds for the one-asset HANK model, while the fake news algorithm took 0.6 seconds. This shows the power of $\mathcal{J}$s as sufficient statistics: once we have them, it is just a matter of linear algebra to obtain a full characterization of equilibrium.

Comparison: times without DAG. How important is our directed acyclic graph (DAG) approach to writing and solving models? Table 3 also shows the time needed to compute $\mathbf{G}$ matrices *without* taking advantage of the structure of our DAGs—by solving a large system of equations in which every input in the DAG, aside from exogenous shocks $\mathcal{Z}$, is turned into an unknown, and every linkage in the DAG is turned into an equation. This turns out to be substantially slower, by a factor that ranges from 2 to 10. The reason is that the dimensionality of the linear system becomes so high that solving it is quite costly.

4.3 Efficient multiplication of simple Jacobians

One important detail underlying the speeds in table 3 is a set of special routines that efficiently handle the Jacobians of simple blocks. These simple blocks comprise the majority of our DAGs. Their Jacobians are easy to obtain to high accuracy (for instance, with symmetric numerical differentiation), and have a special sparse structure: they can be expressed as linear combinations of a few *shift* operators S_i on sequences.

For positive i, S_i maps $(x_0, x_1, \ldots) \to (0, \ldots, 0, x_0, x_1, \ldots)$, with i zeros inserted at the beginning, and for negative $-i$, S_{-i} maps $(x_0, x_1, \ldots) \to (x_i, x_{i+1}, \ldots)$. The former takes an i-period *lag* in sequence space, while the latter takes an i-period *lead* in sequence space.[29] For instance, in the one-asset HANK economy depicted in figure 3b, the Jacobian $\mathcal{J}^{H_1,\pi}$ of the Phillips curve condition with respect to price inflation π is $S_0 - \frac{1}{1+r}S_{-1}$.[30]

For the most part, these operators obey simple rules: if i and j are both positive, $S_i S_j = S_{i+j}$, and so on. However, as is well known from an older literature that works with the lag algebra (e.g. Whiteman 1983), the S are not quite closed under multiplication. To take the simplest example, $S_1 S_{-1}$, a one-period lag of a one-period lead, maps $(x_0, x_1, x_2, \ldots) \to (0, x_1, x_2, \ldots)$, zeroing out the first entry of a sequence and leaving everything else unchanged. Fortunately, we have found a more general set of operators that includes the S and is closed under multiplication following an easy-to-compute rule, as we derive in the following proposition.

[29] In matrix form, S_i has zeros everywhere, except for ones on the ith diagonal below the main diagonal.

[30] This corresponds to a linearized curve of the form $\pi_t = \ldots + \frac{1}{1+r}\mathbb{E}_t\pi_{t+1}$.

Proposition 2. *Let S_i be the shift operator on sequences, and Z_m be the "zero" operator that replaces the first m entries of a sequence with zeros. If we define*

$$Q_{i,m} \equiv \begin{cases} S_i Z_m & i > 0 \\ Z_m S_i & i < 0 \end{cases}$$

then $Q_{i,m} Q_{j,n} = Q_{k,l}$, where

$$k = i + j \tag{39}$$

and

$$l = \begin{cases} \max(m - j, n) & i, j \geq 0 \\ \max(m, n) + \min(i, -j) & i \geq 0,\ j \leq 0 \\ \max(m - i - j, n) & i \leq 0,\ j \geq 0,\ i + j \geq 0 \\ \max(n + i + j, m) & i \leq 0,\ j \geq 0,\ i + j \leq 0 \\ \max(m, n + i) & i, j \leq 0 \end{cases} \tag{40}$$

This proposition nests the shift operators S_i in a more general class of operators $Q_{i,m}$.[31] This has two advantages. First, it makes multiplying the Jacobians of simple blocks vastly more efficient: rather than doing matrix multiplication with large $T \times T$ matrices, we just need to apply rules (39) and (40) a few times. Second, it is computationally easy to multiply $Q_{i,m}$ and an ordinary matrix Jacobian (or vector), since this is a combination of shifting and zeroing elements. Together, these features make forward accumulation on the DAG, which consists mostly of simple blocks, vastly more efficient.

In our online code, we implement this by simply overriding the matrix multiplication operator, so that sparse linear combinations of $Q_{i,m}$ and ordinary matrices can be used interchangeably. With this in place, the methods of section 4.2 can be applied without any outwardly visible modification.

Exploiting sparsity has played a prominent role in both the heterogeneous-agent literature (e.g. Achdou et al. 2017) and the literature on solving for perfect-foresight paths using Newton's method (e.g. Juillard 1996). Our approach builds on the latter, but our much more compact representation of Jacobians offers additional efficiencies. For instance, to store $0.5 \cdot Q_{1,1}$, we only need a few numbers, while a conventional $T \times T$ sparse matrix representation not taking advantage of this structure would need $T - 2$ separate entries, and still create some truncation error.

4.4 Truncation error

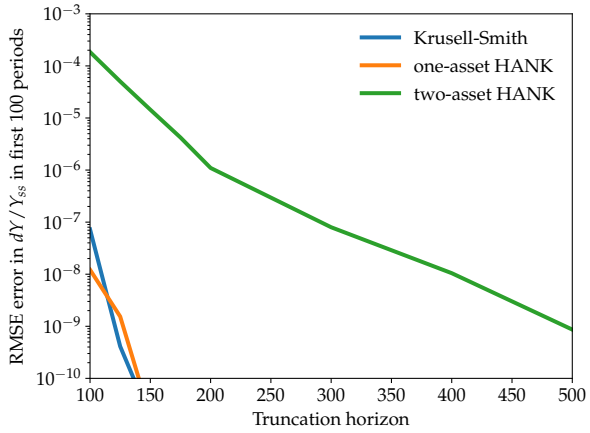
Unlike model reduction for state-space methods, our dimensionality reduction using the DAG does not itself involve any kind of approximation to the underlying model: it is just an efficient way of rewriting the same model equations.

The one approximation that our methods do require—as is generally true for the sequence

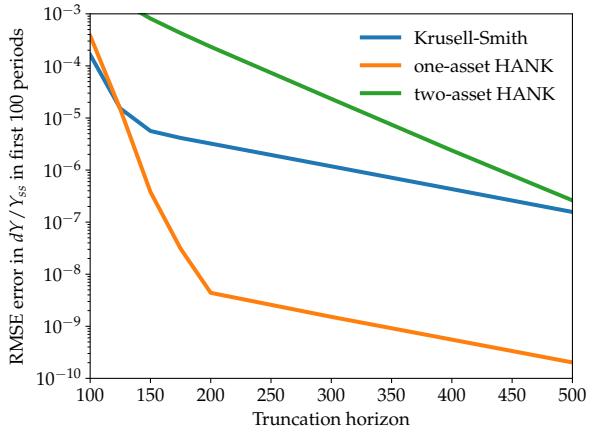
[31] The matrix representation of $T_{i,m}$ is the same as that of S_i, except that the first m entries on the diagonal are zeros.

Figure 5: Impulse response error as a function of the truncation horizon T

(a) TFP persistence $\rho = 0.9$



(b) TFP persistence $\rho = 0.99$



space—is that we truncate Jacobians at some finite horizon T, since it is infeasible to compute using the infinite-horizon Jacobians that are relevant in the theory.[32] For large enough T, the results in any given time period eventually converge to the true equilibrium, but it is important to verify that our results are valid for the T we use in practice.

Figure 5 performs one such exercise for our three models, showing the root mean squared error (RMSE) in dY/Y_{ss} in the first hundred periods, relative to the true equilibrium impulse responses, in response to a persistent 1% shock to TFP. As our benchmark for the true equilibrium, we use the result obtained with an extremely long truncation horizon of $T = 1000$ (beyond which we find minimal effect of T).

The left panel looks at the case of AR(1) persistence $\rho = 0.9$. We see that even for truncation horizons T far shorter than the $T = 300$ used in this paper, the Krusell-Smith and one-asset HANK models have almost perfect accuracy. The two-asset HANK, which has much greater internal persistence, requires a longer T, but by $T = 300$ it is accurate to over seven digits, and with longer truncation horizons steadily improves.

The right panel looks at an extreme case, with AR(1) persistence $\rho = 0.99$. For such a persistent shock—which at $t = 300$ is still at 5% of its level on impact—a much longer truncation horizon is appropriate. Still, the two-asset HANK has a RMSE of just over 10^{-5} with $T = 300$, and the other models are even more accurate.

Overall, figure 5 confirms that although truncation horizons must be chosen judiciously—extending beyond the period of interest and the persistence of the shock—horizons with excellent computational performance are also very accurate for all but the most persistent shocks.

[32] The major exception is the technique for composing Jacobians from simple blocks in section 4.3, where proposition 2 is the exact multiplication rule for the infinite-dimensional Jacobians.

5 Using Jacobians for fast estimation

A common way of confronting dynamic macroeconomic models with time-series data is to estimate them using likelihood-based methods (see e.g. Smets and Wouters 2007 or An and Schorfheide 2007). The standard approach in the DSGE literature is to compute the likelihood by applying the Kalman filter to the model's state space representation. This approach is appropriate for models with small state spaces. With the large state spaces that characterize heterogeneous-agent models, however, solving for the likelihood in this fashion can become prohibitively slow.

In this section, we show how to use the sequence-space Jacobians to rapidly estimate models, and we perform an estimation exercise on US data using our model examples. We proceed in three steps: we first identify impulse responses with the $MA(\infty)$ representation of the model, we then use this equivalence to compute autocovariances from impulse responses, and we finally compute the likelihood function from the autocovariances. Within the DSGE literature, this approach was previously followed by Mankiw and Reis (2007) and Schmitt-Grohé and Uribe (2010), among others. Our key innovation is a process of recyling Jacobians to reevaluate the likelihood which, together with our other innovations in the paper, results in an extremely fast estimation procedure.

5.1 SHADE models with aggregate risk

We first show that the impulse responses computed so far correspond to the $MA(\infty)$ representation of the model with aggregate shocks. The idea leverages the certainty equivalence property of linearization, as in Simon (1956), Theil (1957), Judd and Guu (1993), or Fernández-Villaverde et al. (2016). Boppart, Krusell and Mitman (2018) were the first to observe that this property also holds in heterogeneous agent models computed with small MIT shocks. In appendix C.3, we provide a formal proof of this result. To do this, we set up a version of a general SHADE model with aggregate risk. This *stochastic SHADE model* is still a collection of blocks, but these are now *stochastic blocks*, mapping stochastic processes $\{\tilde{\mathbf{X}}_t\}$ into one another, rather than sequences $\{\mathbf{X}_t\}$. Our results can be summarized as follows.

Assume that exogenous shocks in the model are given by a set of $MA(\infty)$ processes

$$d\tilde{Z}_t^z = \sum_{s=0}^{\infty} m_s^z \epsilon_{t-s}^z$$

where there are as many processes as there are exogenous shocks $z \in \mathcal{Z}$. Here, $\{\epsilon_t^z\}$ are mutually iid standard normally distributed innovations, and $\{m_s^z\}$ are the MA coefficients of the shock to z. We denote by $\mathbf{m}^z$ the column vector that results from stacking the m_s^z's.

For any output $o \in \mathcal{O}$ of the stochastic SHADE model, consider the $MA(\infty)$ representation of

of the equilibrium sequence $\{d\tilde{X}_t^o\}$

$$d\tilde{X}_t^o = \sum_{s=0}^{\infty} \sum_{z \in \mathcal{Z}} m_s^{o,z} \epsilon_{t-s}^z \tag{41}$$

Our key result is that the MA coefficients $\mathbf{m}^{o,z} \equiv (m_s^{o,z})$ of output o in response to shock z can be obtained by solving the associated *perfect foresight* SHADE model, that is, for any $o \in \mathcal{O}, z \in \mathcal{Z}$, we have:

$$\mathbf{m}^{o,z} = \mathbf{G}^{o,z} \mathbf{m}^z \tag{42}$$

where $\mathbf{G}^{o,z}$ is the general equilibrium Jacobian defined in section 4.2.

To understand why this is true, let us focus on a given shock $z \in \mathcal{Z}$ and assume that all shock innovations are zero, except for $\epsilon_t^z = 1$ in some period t. In that case, the expected path of $d\tilde{Z}_t^z$ going forward is given by

$$\mathbb{E}_t \left[d\tilde{Z}_{t+s}^z \right] = m_s^z$$

Hence, this expected path corresponds to a shock $\mathbf{m}^z = d\mathbf{Z}^z$ to z in the perfect-foresight model. In other words, the impulse response of output o is given by $\mathbf{G}^{o,z} d\mathbf{Z}^z = \mathbf{G}^{o,z} \mathbf{m}^z$. In light of (42), then, the $MA(\infty)$ representation of $\{d\tilde{X}_t^o\}$ is given by $\mathbf{m}^{o,z}$.[33]

We next use this equivalence property to perform estimation.[34]

5.2 Second moments and the likelihood function

Starting from the $MA(\infty)$ representation of the model, we first compute all second moments from the impulse responses, and then compute the likelihood function from these second moments.

Second moments. The first step consists of computing the model's autocovariance function. Let $\widehat{\mathcal{O}} \subset \mathcal{O}$ be the set of outputs whose second moments we would like to characterize, and denote by $d\tilde{\mathbf{X}}_t = (d\tilde{X}_t^o)_{o \in \widehat{\mathcal{O}}}$ the vector-valued stochastic process of all outputs in $\widehat{\mathcal{O}}$. Similarly, let $\mathbf{m}_t^{\widehat{\mathcal{O}},\mathcal{Z}} = (m_t^{o,z})_{o \in \widehat{\mathcal{O}}, z \in \mathcal{Z}}$ be the stacked $|\widehat{\mathcal{O}}| \times n_z$ matrix of MA coefficients of $d\tilde{\mathbf{X}}_t$. Then, the autocovariances of $d\tilde{\mathbf{X}}_t$ are given by

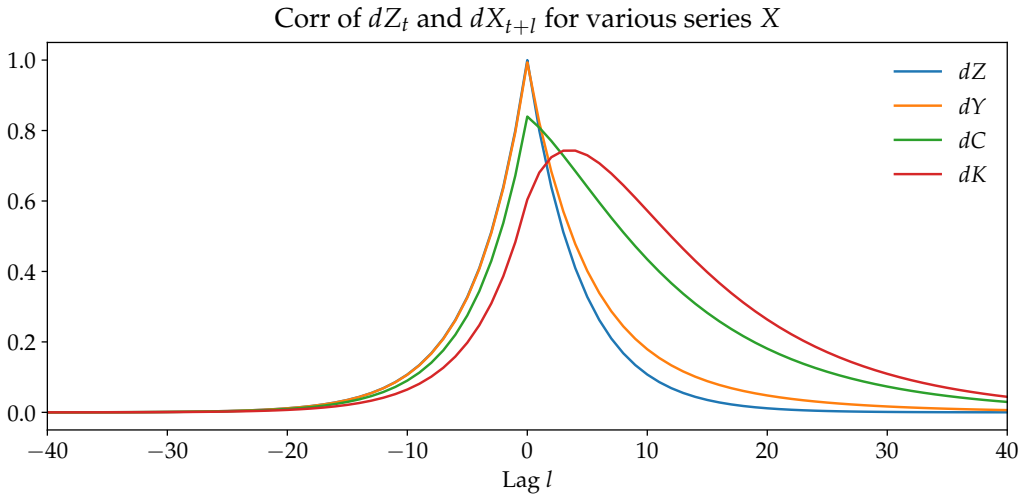
$$\text{Cov}(d\tilde{\mathbf{X}}_t, d\tilde{\mathbf{X}}_{t'}) = \sum_{s=0}^{T-(t'-t)} \left[\mathbf{m}_s^{\widehat{\mathcal{O}},\mathcal{Z}} \right] \left[\mathbf{m}_{s+t'-t}^{\widehat{\mathcal{O}},\mathcal{Z}} \right]' \tag{43}$$

Equation (43) is the standard expression for the autocovariance function of an MA process (see, for instance, Hamilton 1994). As is well-known, these autocovariances only depend on the distance $t' - t$, not on t and t' separately.

[33] Of course, since in practice the $\mathbf{G}$ matrix is truncated to $T \times T$ horizon, the process is an $MA(T-1)$.

[34] A corollary to our equivalence result is that our sequence-space impulse responses correspond to impulse responses in the state-space formulation of the model, *without any reduction in the dimensionality of the state space*. The ability to circumvent the "model reduction" step, which is a crucial but delicate aspect of state-space methods in large heterogeneous-agent models (see Reiter 2010, Ahn et al. 2018, Bayer and Luetticke 2018), is a major advantage of our approach.

Figure 6: Selected second moments of the Krusell and Smith (1998) model for AR(1) TFP, $\rho = 0.8$



In figure 6, we provide an illustration of second moments in a stochastic version of the Krusell and Smith (1998) model, where the TFP shock follows an AR(1) process

$$d\tilde{Z}_t = \sigma \sum_{s=0}^{\infty} \rho^s \epsilon_{t-s}$$

with persistence ρ and innovations with standard deviation σ. The figure shows the correlations of productivity, output, consumption, and capital with the underlying productivity process, at various lags. The figure shows that capital and consumption—and to a much lesser extent, output—tend to lag productivity. This reflects the typical transmission mechanism of TFP shocks in RBC models.

In appendix B.5, we describe how (43) can be calculated in an very efficient way using the fast Fourier transform (FFT). As table 7 reveals, for our estimation exercises in section 5.4, moving from the $MA(\infty)$ representation to a full set of autocovariances, which are stacked in the matrix $\mathbf{V}$, takes between 0.5 and 2.5 milliseconds.

Likelihood function. The second step to estimation is to evaluate the likelihood function. Let

$$d\tilde{\mathbf{X}}_t^{obs} = B\, d\tilde{\mathbf{X}}_t + \mathbf{u}_t$$

denote the vector of n_{obs} observables whose likelihood we would like to determine. Here $\{\mathbf{u}_t\}$ is iid normal with mean 0 covariance matrix Σ_u, and B is a $n_{obs} \times |\widehat{\mathcal{O}}|$ matrix. Since $d\tilde{\mathbf{X}}_t^{obs}$ is a linear combination of the ϵ_t^z and $\mathbf{u}_t$ terms, it has a multivariate normal distribution. Moreover, its second moments are a simple linear transformation of those of $d\tilde{\mathbf{X}}_t$:

$$\text{Cov}(d\tilde{\mathbf{X}}_t^{obs}, d\tilde{\mathbf{X}}_{t'}^{obs}) = 1_{t=t'} \cdot \Sigma_{\mathbf{u}} + B\,\text{Cov}(d\tilde{\mathbf{X}}_t, d\tilde{\mathbf{X}}_{t'})\, B' \tag{44}$$

We stack these covariances into a large symmetric $n_{obs}T_{obs} \times n_{obs}T_{obs}$ matrix $\mathbf{V}$, where T_{obs} is the number of time periods in our data. The log-likelihood function is then the conventional log multivariate density. Dropping the constant term, it can be expressed as a function of the observed data $d\tilde{\mathbf{X}}^{obs} = (d\tilde{\mathbf{X}}_t^{obs})$ (stacked as $n_{obs}T_{obs}$-dimensional vector) as

$$\mathcal{L} = -\frac{1}{2}\log\det\mathbf{V} - \frac{1}{2}\left[d\tilde{\mathbf{X}}^{obs}\right]'\mathbf{V}^{-1}\left[d\tilde{\mathbf{X}}^{obs}\right] \tag{45}$$

Appendix B.5 discusses efficient computation of (45), given $\mathbf{V}$ and $d\tilde{\mathbf{X}}^{obs}$. Our baseline approach is to perform a Cholesky decomposition of $\mathbf{V}$, from which we can quickly calculate both the log determinant $\log\det\mathbf{V}$ and the quadratic form $\left[d\tilde{\mathbf{X}}^{obs}\right]'\mathbf{V}^{-1}\left[d\tilde{\mathbf{X}}^{obs}\right]$. Table 7 reveals that this is quite efficient: calculating $\mathcal{L}$ takes about one millisecond or less in all applications except the two-asset HANK, where it takes about 10 milliseconds.[35]

5.3 Reusing Jacobians for rapid estimation

The idea of using the $MA(\infty)$ representation of the model to perform estimation is not new. Instead, our key methodological innovation is to reuse sequence-space Jacobians in evaluating $\mathbf{V}$ for different parameters. Almost any likelihood-based estimation methodology requires reevaluating the likelihood function many times. Our methods make this extremely simple, provided that the parameters that are changing keep the *steady state* intact. This is true of parameters that govern the shock processes, as well as parameters that are only relevant for the transition, such as adjustment cost parameters and policy rule coefficients.

To understand the benefit of reusing Jacobians, consider the DAG representation of equilibrium. In our method, the first evaluation of the likelihood computes the Jacobian of each block individually, stores these Jacobians, and then rapidly composes them along the DAG to generate impulse responses. Our key observation is that we can reuse many of these stored Jacobians when recomputing the likelihood. In particular, shock processes do not change any Jacobian—allowing us to calculate $\mathbf{G}$ once and reuse it for every evaluation—while most transition-relevant parameters only alter simple block Jacobians, which are fast to recompute and combine.

Recycling the Jacobians is extremely useful in practice. To illustrate, consider from table 7 that our two-asset HANK model estimation requires 8259 evaluations of the likelihood function. From table 2, computing the household block Jacobian takes 5.7 seconds when using our fake news algorithm, and 2107 seconds when using a direct method. Hence, estimation with no reuse of Jacobians would take at least 13 hours when using our fake news algorithm, and a full 200 days when using a direct algorithm.

We now provide details on our three estimation exercises.

[35] Alternatives to Cholesky for high T include Levinson recursion (e.g. Meyer-Gohde 2010) or the Whittle (1953) approximation to the likelihood (e.g. Plagborg-Møller 2019).

Table 4: Estimated parameters for our Krusell-Smith economy

Shock		Prior distribution	Mode	Posterior std. dev
	s.d.	Invgamma(0.4, 4)	0.178	(0.010)
TFP shock	AR-1	Beta(0.5, 0.2)	0.908	(0.035)
	AR-2	Beta(0.5, 0.2)	0.330	(0.088)

5.4 Three estimation exercises

In this section, we perform Bayesian estimation exercises using our three example models. Our primary objective here is to illustrate that this can be done very efficiently. In this spirit, our procedure is only to maximize the posterior to determine its mode, and to use a local Laplace approximation around the mode to compute standard errors. We leave a full exploration of the posterior distribution using Markov chain Monte Carlo methods, as well as a detailed understanding of the economics behind the estimation results, to future research.

Our assumptions are the following. We assume that the priors for the standard deviations of all shocks are Inverse-Gamma distributed with mean 0.4 and standard deviation 4. We assume that the priors for persistence parameters are Beta distributed with mean 0.5 and standard deviation 0.2. We also assume no measurement error, $\Sigma_{\mathbf{u}} = 0$. Initial conditions for parameters are set at their prior modes. For each of our models, we use the data from Smets and Wouters (2007). We linearly detrend the logs of all growing variables (output, consumption, investment, wages, hours) and take out the sample mean of inflation and nominal interest rates. The individual models are then estimated as follows.

Estimating Krusell-Smith. We estimate our Krusell and Smith (1998) model with just a single shock, TFP, and a single time series $\{dX_t^{obs}\}$, output. We assume TFP shocks to follow an $AR(2)$ process, $(1-\rho_1 L)(1-\rho_2 L)d\tilde{Z}_t = \sigma\epsilon_t$,where L is the lag operator. We estimate the autoregressive roots ρ_1, ρ_2 as well as the standard deviation σ. Table 4 shows the estimates, finding one persistent root, $\rho_1 \approx 0.9$, and one less persistent root $\rho_2 \approx 0.3$.

Estimating the one-asset HANK model. We estimate our one-asset HANK model both with only shock parameters, and with shock and model parameters together. In both cases, we use three shocks (monetary policy shocks, government spending shocks, and price markup shocks) and three time series (output, inflation, and nominal interest rates). Each shock is modeled as an $AR(1)$ with its own standard deviation and persistence. Thus, there are six shock parameters for this model. The first two posterior columns in table 5 show our estimates when only estimating those shock parameters; we find that government spending shocks are the most volatile and persistent. The last two posterior columns in table 5 are the estimated shock and model parameters in the joint estimation. We find a Taylor coefficient ϕ just above 1, a modest responsiveness of the Taylor

Table 5: Estimated parameters for our one-asset HANK economy

Parameter / shock		Prior distribution	Posterior (shocks) Mode	Posterior (shocks) std. dev	Posterior (shocks + model) Mode	Posterior (shocks + model) std. dev
Monetary policy shock	s.d.	Invgamma(0.4, 4)	0.443	(0.025)	0.257	(0.015)
	AR-1	Beta(0.5, 0.2)	0.678	(0.021)	0.103	(0.052)
G shock	s.d.	Invgamma(0.4, 4)	0.684	(0.044)	0.595	(0.038)
	AR-1	Beta(0.5, 0.2)	0.816	(0.024)	0.878	(0.027)
P markup shock	s.d.	Invgamma(0.4, 4)	0.150	(0.012)	0.285	(0.023)
	AR-1	Beta(0.5, 0.2)	0.629	(0.046)	0.141	(0.069)
ϕ		Gamma(1.5, 0.25)			1.055	(0.017)
ϕ_y		Gamma(0.5, 0.25)			0.024	(0.010)
κ		Gamma(0.1, 0.1)			0.325	(0.056)

rule to output ϕ_y, and a Phillips curve slope parameter κ around 0.3. These are typical values in the literature.

Estimating the two-asset HANK model. We add all seven shocks from Smets and Wouters (2007) to the two-asset model: shocks to TFP, government spending, monetary policy, price and wage markups; the two exceptions are that we use discount factor shocks rather than "risk premium" shocks (both shock the Euler equation and are thus very similar), and we shock firms' first-order conditions for capital instead of investment-specific technology shocks (which have problematic implications for the relative price of investment, see Justiniano, Primiceri and Tambalotti 2010, 2011 and Schmitt-Grohé and Uribe 2012). We use those seven shocks to perfectly match the time series of output, consumption, investment, hours, wages, nominal interest rates and price inflation. As with the one-asset model, we estimate two versions of the model, one with only shock parameters and one with shock and model parameters (table 6). Compared to the one-asset model, we find here more responsive Taylor rule parameters ϕ, ϕ_y, and smaller Phillips curve slope parameters κ^p, κ^w. We also estimate the degree of capital adjustment costs ϵ_I and find it to be in line with standard estimates from the literature.

Estimation times. Table 7 lists the overall computing times for each of our five estimation exercises, as well as times for each likelihood evaluation and their composition across the three steps. Once the $\mathbf{G}^{o,z}$ matrices are computed (table 3), the Krusell-Smith model's likelihood can be evaluated in less than one millisecond, with an overall estimation time of around 120 milliseconds. We attain similar speeds estimating the shock processes for the one-asset HANK model. Since we allow for seven shocks when estimating the two-asset HANK model, estimating the shocks' parameters is somewhat slower than in the other two models; this has nothing to do with the complexity or micro heterogeneity of the two-asset model. Still, a single likelihood evaluation takes a few milliseconds, and the entire estimation a few seconds.

Table 6: Estimated parameters for our two-asset HANK economy

			Posterior (shocks)		Posterior (shocks + model)	
Parameter / shock		**Prior distribution**	**Mode**	**std. dev**	**Mode**	**std. dev**
TFP shock	s.d.	Invgamma(0.4, 4)	0.222	(0.013)	0.223	(0.013)
	AR-1	Beta(0.5, 0.2)	0.070	(0.043)	0.134	(0.063)
G shock	s.d.	Invgamma(0.4, 4)	0.538	(0.030)	1.357	(0.218)
	AR-1	Beta(0.5, 0.2)	0.885	(0.005)	0.830	(0.012)
β shock	s.d.	Invgamma(0.4, 4)	1.079	(0.061)	1.077	(0.060)
	AR-1	Beta(0.5, 0.2)	0.941	(0.011)	0.944	(0.007)
r_I (investment) shock	s.d.	Invgamma(0.4, 4)	0.674	(0.056)	0.881	(0.093)
	AR-1	Beta(0.5, 0.2)	0.708	(0.028)	0.356	(0.073)
Monetary policy shock	s.d.	Invgamma(0.4, 4)	0.610	(0.056)	0.469	(0.046)
	AR-1	Beta(0.5, 0.2)	0.561	(0.038)	0.139	(0.062)
P markup shock	s.d.	Invgamma(0.4, 4)	0.145	(0.010)	0.176	(0.027)
	AR-1	Beta(0.5, 0.2)	0.042	(0.031)	0.206	(0.103)
W markup shock	s.d.	Invgamma(0.4, 4)	1.944	(0.110)	2.042	(0.252)
	AR-1	Beta(0.5, 0.2)	0.015	(0.011)	0.034	(0.024)
ϕ		Gamma(1.5, 0.25)			1.407	(0.110)
ϕ_y		Gamma(0.5, 0.25)			1.378	(0.257)
κ^p		Gamma(0.1, 0.1)			0.075	(0.043)
κ^w		Gamma(0.1, 0.1)			0.125	(0.035)
ϵ_I		Gamma(4, 2)			2.998	(1.731)

Table 7: Estimation times.

	Krusell-Smith	one-asset HANK		two-asset HANK	
	shocks	shocks	model + shocks	shocks	model + shocks
Single likelihood evaluation	**0.877 ms**	**2.151 ms**	**58.440 ms**	**11.422 ms**	**179.371 ms**
step 1 ($MA(\infty)$: $\{\mathbf{m}_t^{\widehat{O},\mathcal{Z}}\}$)	0.020 ms	0.234 ms	56.279 ms	1.876 ms	167.094 ms
step 2 (autocovariances $\mathbf{V}$)	0.673 ms	0.985 ms	0.963 ms	2.209 ms	2.436 ms
step 3 (log likelihood $\mathcal{L}$)	0.184 ms	0.932 ms	1.198 ms	7.337 ms	9.842 ms
Full estimation	**0.12 s**	**0.50 s**	**15.82 s**	**21.22 s**	**566.55 s**
no. of evaluations	109	188	554	1544	8249
No. of shocks	1	3	3	7	7
No. of estimated shock parameters	3	6	6	14	14
No. of estimated model parameters	0	0	3	0	5
Total no. of estimated parameters	3	6	9	14	19

When model parameters are also estimated, the likelihood takes a bit longer to be re-evaluated. This is entirely due to step 1—the computation of impulse responses. The single likelihood evaluation for the two-asset model, for instance, takes 180 ms rather than 11 ms when model parameters change. As noted above, we gain speed by reusing Jacobians, because many reevaluations of the likelihood do not require updating model parameters and maintain Jacobians intact. This allows us to still keep the *average* time it takes to evaluate likelihood below 70 ms, and for an overall estimation time of around 9 minutes. To the best of our knowledge, these are much faster speeds for estimation of such models than what any other method has been able to achieve.

6 Using Jacobians to evaluate determinacy

So far we have discussed model simulation and estimation without formally addressing the question of determinacy: we have simply assumed a locally determinate solution.

This turns out to be necessary for our sequence space methods to work. If, instead, there is local multiplicity, then there is a nonzero bounded solution $d\mathbf{U}^{mult}$ to $\mathbf{H_U} d\mathbf{U}^{mult} = 0$, and this singularity in $\mathbf{H_U}$ implies that $d\mathbf{U} = -\mathbf{H_U}^{-1}\mathbf{H_Z} d\mathbf{Z}$ in (36) is not well-defined.

In practice, when $\mathbf{H_U}$ is truncated to some finite T, this generically shows up as *near*-singularity of the truncated $\mathbf{H_U}$. Then, when we calculate $d\mathbf{U} = -\mathbf{H_U}^{-1}\mathbf{H_Z} d\mathbf{Z}$, the near-infinite entries of $\mathbf{H_U}^{-1}$ cause $d\mathbf{U}$ to explode, and its entries are meaningless. Eyeballing $d\mathbf{U}$ can thus provide a useful heuristic test of multiplicity. Directly testing $\mathbf{H_U}$ for near-singularity—for instance, by looking at the singular values—provides additional evidence.

These are only, however, *approximate* criteria. In this section, we provide instead an *exact* determinacy criterion that makes use of the asymptotic structure of the $\mathbf{H_U}$ matrix of SHADE models, and is simple to evaluate. This provides a sequence-space counterpart to the well-known determi-

nacy tests (e.g. Blanchard and Kahn 1980, Klein 2000, or Sims 2002) for linear rational expectations models in state-space form, filling an important gap in the literature. And since our criterion can be evaluated much more quickly than state-space tests for large heterogeneous-agent models, it may prove useful for the rapidly growing literature on determinacy in this area (see, for example, Ravn and Sterk 2017, Acharya and Dogra 2019, and Bilbiie 2019).[36]

Throughout this section, we focus on SHADE models whose heterogeneous-agent blocks admit the discretized representation in (14)–(16). We first define local determinacy of a SHADE model, paralleling the definition in Woodford (2003).

Definition 4 (Local determinacy)**.** A generic SHADE model is locally determinate if, to first order, it admits at most a single bounded equilibrium.

Our condition for determinacy is based on the asymptotic behavior of the Jacobian $\mathbf{H_U}$, which is characterized by the following lemma.

Lemma 1. *The Jacobian $\mathbf{H_U}$ of a SHADE model is asymptotically time-invariant, that is,*

$$\mathbf{H_U} \simeq \begin{pmatrix} * & * & \ddots & \ddots & & \\ * & * & A_{-1} & A_{-2} & \ddots & \\ \ddots & A_1 & A_0 & A_{-1} & A_{-2} & \ddots \\ \ddots & A_2 & A_1 & A_0 & A_{-1} & \ddots \\ & \ddots & A_2 & A_1 & A_0 & \ddots \\ & & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{46}$$

where $A_j \equiv \lim_{s\to\infty} \mathbf{H}_{\mathcal{U}s,s+j} = \lim_{s\to\infty} \frac{\partial \mathbf{H}_{s+j}}{\partial \mathbf{U}_s}$, $j \in \mathbb{Z}$, are $k \times k$ matrices. The sequence (A_j) is absolutely summable entry-by-entry.

Lemma 1 follows from the fact that both simple and heterogeneous-agent blocks admit asymptotically time-invariant Jacobians. This is true by definition for simple blocks, and follows from Proposition 1 for heterogeneous-agent blocks. Moreover, the composition of blocks into a SHADE model preserves asymptotic time invariance.

Lemma 1 is useful because it allows us to characterize the behavior of the model solution $d\mathbf{U}_t$ for large t. To explain where our criterion comes from and how it relates to the standard state-space criteria, we begin with a heuristic derivation.

Heuristic derivation. Suppose that A_j is only nonzero for finitely many $j \in \{-l, \dots, m\}$, and that dU_t has only one dimension. Suppose further that t is large enough that the impact of the

[36] For example, Ahn et al. (2018) after model reduction requires performing a Schur decomposition on a 2445×2445 matrix, which takes us approximately 6 seconds, compared to 631 microseconds in Table 1 for evaluating determinacy in the two-asset model once given $\mathbf{H_U}$ (which requires 110 milliseconds to compute in Table B.1).

shock on the model $\mathbf{H}$, $\mathbf{H}_Z d\mathbf{Z}$, has vanished. Under these simplifying assumptions, the equation $\mathbf{H}_U d\mathbf{U} = -\mathbf{H}_Z d\mathbf{Z}$ can be written, for large enough time t, as:

$$\sum_{j=-l}^{m} A_j dU_{t-j} = 0 \tag{47}$$

Equation (47) is a linear difference equation for dU_t, expressing dU_t at every t as a function of m predetermined values $\{dU_{t-m}, \ldots, dU_{t-1}\}$ and l future values $\{dU_{t+1}, \ldots, dU_{t+l}\}$. Let $\{z_i\}_{i=1}^{m+l}$ be the $m+l$ complex solutions to the polynomial equation

$$\sum_{j=-l}^{m} A_j z^{m-j} = 0$$

and assume for simplicity that all these roots are distinct and ordered such that $|z_0| \leq \cdots \leq |z_n| \leq 1 < |z_{n+1}| \leq \ldots |z_{m+l}|$, that is, n is the number of roots inside the unit circle. Then, by standard results, the set of solutions to (47) is characterized by

$$dU_t = \sum_{i=1}^{m+l} C_i z_i^t \tag{48}$$

Moreover, the boundedness requirement in definition 4 imposes that $C_i = 0$ for $i \geq n+1$. The resulting system is determinate when the n remaining unknown coefficients $\{C_i\}$ can be solved for, given the m predetermined values $\{dU_{t-m}, \ldots, dU_{t-1}\}$. Since all the z_i are distinct, this is possible if and only if the number of roots inside the unit circle is equal to the number of predetermined variables, that is, if and only if $n = m$. If $n > m$, then there is multiplicity.

The idea of counting roots vs. predetermined variables also underlies existing state-space determinacy criteria, such as Blanchard and Kahn (1980). For SHADE models that are only made of simple blocks, both l and m are finite and this criterion can be applied exactly. But for general SHADE models that include heterogeneous-agent blocks, l and m are generally infinite and an alternative criterion is needed.

An alternative way to state the $n = m$ condition is to say that the Laurent polynomial

$$A(z) \equiv \sum_{j=-l}^{m} A_j z^j$$

has exactly as many zeros inside the unit circle as it has poles (it has l poles and $m + l - n$ zeros). If, on the other hand, there are fewer zeros than poles, then $n > m$ and there is multiplicity.

According to the argument principle, the number of zeros and poles coincide if and only if

$$\frac{1}{2\pi i} \oint_C \frac{A'(z)}{A(z)} dz = \#\text{ zeros in unit circle} - \#\text{ poles in unit circle} = 0 \tag{49}$$

where C is the counterclockwise unit circle in the complex plane. Substituting $w = f(z)$, the

integral in (49) becomes $\frac{1}{2\pi i}\oint_{A(C)}\frac{1}{w}dw$, which is the *winding number* of the curve $A(C)$ around the origin—the number of times it winds counterclockwise around the origin. We can parameterize the curve $A(C)$ as

$$\mathcal{A}(\lambda) \equiv \sum_{j=-l}^{m} A_j e^{ij\lambda}; \quad \lambda \in [0, 2\pi] \tag{50}$$

Our criterion then states that there is a unique solution if and only if the winding number of $\mathcal{A}(\lambda)$ is zero. A negative winding number corresponds to multiplicity.

In principle, one can evaluate $\mathcal{A}(\lambda)$ and compute its winding number even if m and l are infinite, and it turns out that the criterion continues to work correctly in this case. It also extends to the case where $d\mathbf{X}$ is multidimensional, provided that one uses the winding number of $\det \mathcal{A}(\lambda)$. We have the following result, formally proven in the appendix.

Proposition 3. *A SHADE model is locally determinate if and only if the winding number of* $\det \mathcal{A}(\lambda)$ *is equal to zero.*

This result is powerful since it provides us with a direct and efficient way to check for determinacy of a SHADE model:

1. First, after constructing the Jacobian $\mathbf{H_U}$, store A_j's as the entries above and below the main diagonal for a far-out column of $\mathbf{H}_U$. For instance, if the truncation horizon is $T = 300$, we can store the entries in rows 0 through 299 of column 150 as A_{-150} through A_{149}.[37]

2. Second, using these A_j, calculate the winding number of $\det \mathcal{A}(\lambda)$, i.e. the number of times the graph of $\det \mathcal{A}(\lambda)$ wraps counterclockwise around the origin as λ goes from 0 to 2π.[38]

Table 1 shows that this method, given $\mathbf{H_U}$, is almost instantaneous, taking less than one millisecond for each of our models.

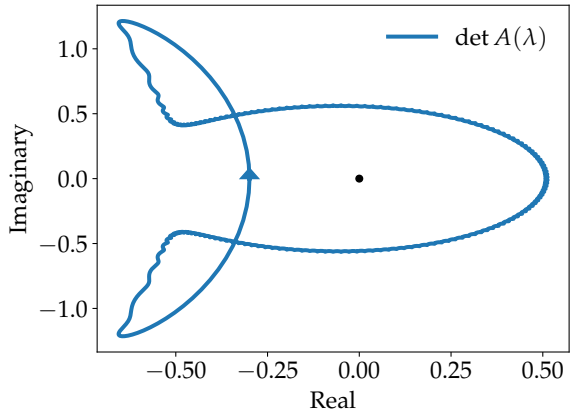
A direct predecessor to proposition 3 is a winding-number based determinacy condition for linear rational expectations equilibria proposed by Onatski (2006). Onatski's condition is stated identically in terms of the winding number of $\det \mathcal{A}(\lambda)$, but it can only be applied in the case where $\mathbf{H}_U$ is *exactly* time invariant, that is $\mathbf{H}_{Us,s+j} = A_j$ for *every* s. In the standard formulation of SHADE models, this is almost never the case when there are heterogeneous-agent blocks. However, the proof of proposition 3 shows how to transform any SHADE model into a much larger but equivalent model that has an exactly time invariant Jacobian, so that the Onatski (2006) criterion can be applied and connected to the asymptotically time-invariant representation (46).

[37] An alternative approach, which bypasses needing the full $\mathbf{H_U}$ and can in principle deliver greater accuracy, is to construct the asymptotic A_js in (46) for the model from those of individual blocks by convolution, which can be done efficiently with the FFT. We use the $\mathbf{H_U}$ approach for simplicity, since it is equally accurate for our examples.

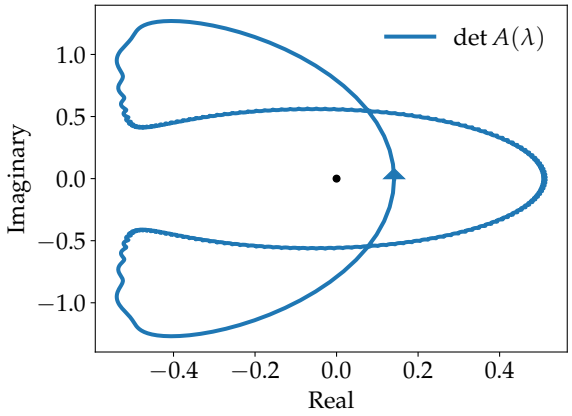
[38] One efficient way to do so is to sample (50) at a large number of equispaced points on the interval $[0, 2\pi]$ using the fast Fourier transform (FFT)—which allows all these values to be computed simultaneously—and then count how many times the resulting piecewise linear curve crosses the ray from 0 to positive real infinity from below.

Figure 7: $\det \mathcal{A}(\lambda)$ for determinate and indeterminate cases of the one-asset HANK model

(a) Indeterminate case: $\phi = 1.001$



(b) Determinate case: $\phi = 1.007$



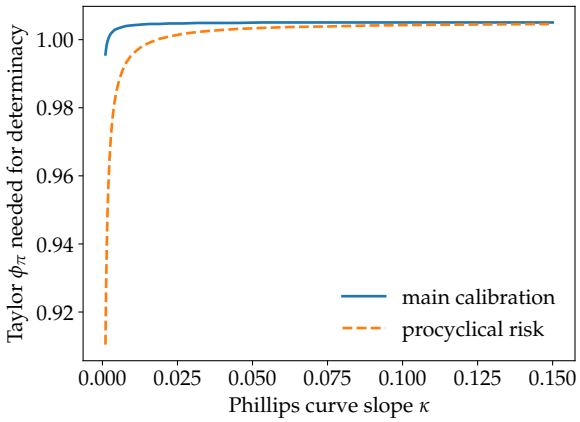
One-asset HANK example. We illustrate proposition 3 by applying it to our one-asset HANK model, using the parameterization described in appendix A.2. Since the model has 3 unknowns and 3 targets, the A_j's are 3×3 matrices. The two panels in figure 7 plot $\det \mathcal{A}(\lambda)$ for two values of ϕ, the slope of the Taylor rule with respect to inflation. In panel (a), $\phi = 1.001$. As can be seen, $\det \mathcal{A}(\lambda)$ wraps once counterclockwise around the origin, implying local *indeterminacy*. This shows that the standard Taylor principle for representative agent New Keynesian models fails for this HANK model. However, in our baseline parameterization, the Taylor coefficient threshold for determinacy is very close to 1. Indeed, panel (b) of figure 7 shows that raising ϕ to 1.007 shifts the contour such that it no longer wraps around the origin, indicating a locally *determinate* model. Using bisection, we find that the determinacy threshold ϕ^* is about $\phi^* \approx 1.005$.

How does ϕ^* depend on the slope κ of the Phillips curve? Given our determinacy criterion, it is straightforward to bisect and find ϕ^* for many values of κ. Panel (a) of figure 8 plots the resulting ϕ^* as a function of the slope of the Phillips curve κ. For our main calibration, it turns out that ϕ^* is increasing in price flexibility but remains close to 1 in all instances, never moving far from the representative agent benchmark. For an alternative calibration, where we follow McKay et al. (2016) and make idiosyncratic income risk procyclical by evenly rebating countercyclical dividends to all agents, ϕ^* is similar for more flexible prices but is much lower when prices are very sticky. This latter finding is consistent with a growing analytical literature (see Ravn and Sterk 2017, Acharya and Dogra 2019, or Bilbiie 2019) finding that procyclical risk makes determinacy easier to achieve, while countercyclical risk pushes toward indeterminacy.

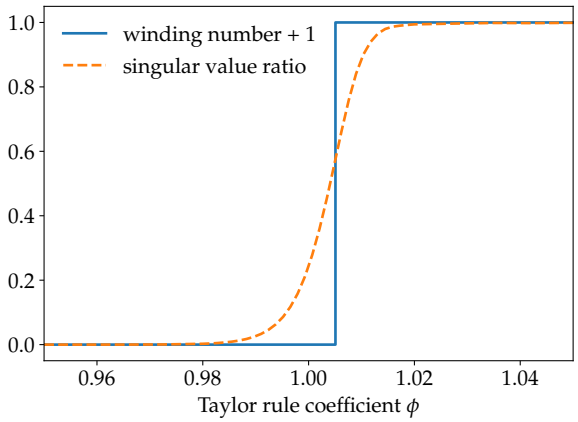
Comparison to numerical determinacy criterion. As discussed at the top of this section, an alternative way to check determinacy is to evaluate the $\mathbf{H_U}$ matrix numerically: in cases of indeterminacy, the true $\mathbf{H_U}$ is singular, and the truncated $\mathbf{H_U}$ is nearly singular.

Figure 8: Checking determinacy

(a) Determinacy threshold vs. price flexibility



(b) Winding number criterion vs. singular value ratio



When determinacy is uncertain, often there is at most a single dimension of indeterminacy, as in our application above. This suggests a simple test involving singular values: if the smallest singular value is far below the second-smallest, then the model is likely indeterminate, while if they are similar, the model is likely determinate. Panel (b) of figure 8 implements this test, plotting the ratio of smallest and second-smallest singular values ("singular value ratio") as we vary ϕ in the baseline calibration. This is plotted against the winding number plus one, which steps up from 0 to 1 at exactly the point where the model becomes determinate. We see that although the ratio of singular values gives a consistent answer—increasing from 0 to 1 near the determinacy threshold—it offers less precise guidance, making clear the advantage of our winding number criterion.[39]

7 Using Jacobians to solve nonlinear MIT shocks

Up until now, we have demonstrated that the sequence-space Jacobian is sufficient to capture all first-order effects of heterogeneity. Once we compute its value at the steady state, we can obtain impulse responses, evaluate local determinacy, and compute the model likelihood just as easily as if we were dealing with a representative-agent model. Our methods preserve the nonlinear effects of idiosyncratic uncertainty, such as heterogeneous MPCs. However, linearization with respect to aggregates eliminates any dependence on the size and sign of aggregate shocks: the response to large shocks is just a scaled-up version of the response to small shocks.

[39]One benefit of using singular values is that the right singular vector associated with the smallest singular value corresponds to a $d\mathbf{U}$ that approximately solves $\mathbf{H}_\mathbf{U} d\mathbf{U} = 0$. This allows us to characterize and study the multiplicity. This points to a hybrid approach: first, use the efficient and precise winding number criterion to test determinacy. Second, if multiplicity is revealed, use a singular value decomposition to solve for the approximate null vector $d\mathbf{U}$. This approach can therefore in principle be applied to solve indeterminate models, as in e.g. Lubik and Schorfheide (2003).

In this section, we provide a method for solving nonlinear perfect-foresight transitions starting from steady state. Technically, we look for a nonlinear solution to equation (34). The interpretation of such a solution requires caution. The resulting impulse response can no longer be viewed as emerging from a model with aggregate shocks. Instead, it requires a strict interpretation as the dynamic effect of a large unexpected shock, starting from steady state—a pure "MIT shock".

On the other hand, this nonlinear solution has some important benefits. First, it allows the researcher to explore size dependence and sign asymmetries, which have gathered some attention in the literature (see e.g. Kaplan and Violante 2018 in the context of fiscal policy and Berger et al. 2018 for house price changes). Second, it allows the researcher to perform a test of the accuracy of the linearization, as recommended by Boppart et al. (2018).

Since we have already computed the sequence-space Jacobian, it is natural to look for nonlinear solutions to (34) by using a version of Newton's method.[40] To find the $\mathbf{U}$ that solves $\mathbf{H}(\mathbf{U}, \mathbf{Z}) = 0$ for a given $\mathbf{Z}$, with all sequences truncated to $T = 300$, we proceed as follows:

1. Starting from $j = 0$, guess a path $\mathbf{U}^0$ (typically, $\mathbf{U}^0 = \mathbf{U}_{ss}$)
2. Calculate $\mathbf{H}(\mathbf{U}, \mathbf{Z})$
3. Update the guess according to

$$\mathbf{U}^{j+1} = \mathbf{U}^j - \left[\mathbf{H}_{\mathbf{U}}(\mathbf{U}_{ss}, \mathbf{Z}_{ss})\right]^{-1} \mathbf{H}\left(\mathbf{U}^j, \mathbf{Z}\right)$$

This algorithm falls in the class of quasi-Newton methods, since the steady-state sequence space Jacobian $\mathbf{H}_{\mathbf{U}}(\mathbf{U}_{ss}, \mathbf{Z}_{ss})$ is used instead of the actual Jacobian $\mathbf{H}_{\mathbf{U}}\left(\mathbf{U}^j, \mathbf{Z}\right)$.[41]

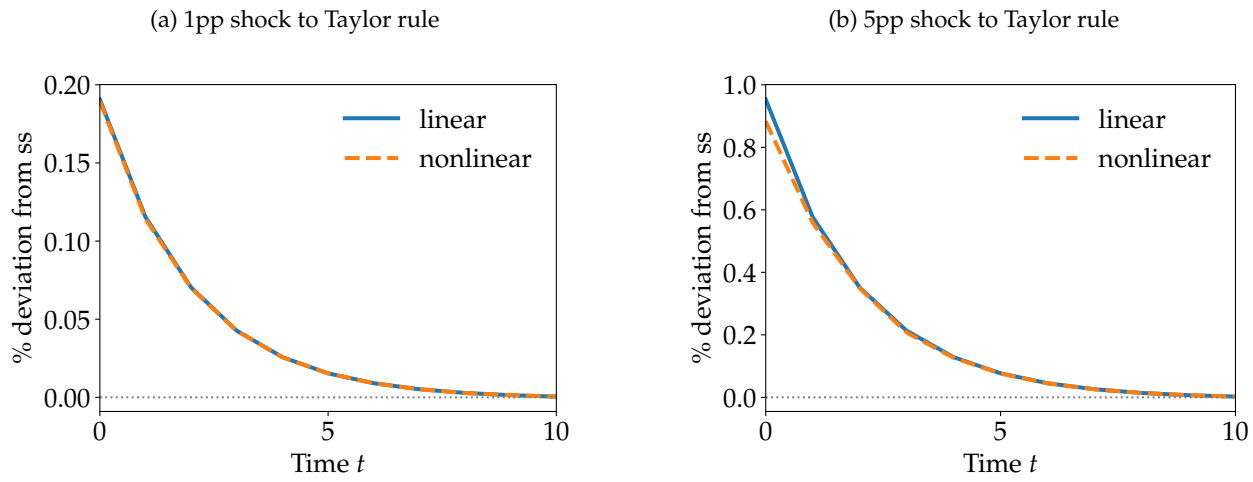
We illustrate this method in the case of our one-asset HANK model. Because it has flexible wages and endogenous labor supply, it is the model that features the biggest nonlinearities among those that we are considering. Figure 9 presents the nonlinear impulse response of the model to a monetary policy shock starting from the steady state at $t = 0$. Panel (a) plots the impulse response of consumption after a -1pp shock to the Taylor rule. Consumption increases, as is typical following a monetary policy shock, and the magnitude is essentially identical to the linear impulse response. Panel (b) plots the impulse response of consumption with a -5pp shock. Here, in the nonlinear solution, consumption increases by slightly less than in the linear solution.

There are two potential economic reasons for nonlinearities in this type of model. First, nonlinearities in adjustment costs, such as the resource costs of price adjustment in the Rotemberg model, start to kick in for large shocks. Second, large shocks can move constrained households away from the borrowing limit. Here, a combination of these two forces is likely at work. However, note that the nonlinearities are very slight, even though the shock is implausibly large. Our

[40] This general idea dates back to Laffargue (1990), Boucekkine (1995), and Juillard (1996), and is implemented in Dynare's perfect-foresight solution method, whose command is `perfect_foresight_solver`.

[41] For heterogeneous-agent models, previous versions of this method were used by generating approximations to the $\mathbf{H}_{\mathbf{U}}(\mathbf{U}_{ss}, \mathbf{Z}_{ss})$ matrix. Auclert and Rognlie (2018) used an auxiliary model, while Straub (2017) used interpolations to approximate the Jacobian from a limited number of partial equilibrium transitions.

Figure 9: Linear vs. nonlinear impulse responses of consumption to monetary shocks



model therefore passes the Boppart, Krusell and Mitman (2018) test, suggesting that there is little reason to be concerned about the accuracy of linearization for a typical monetary shock.

The algorithm above converges to $|\mathbf{H}| < 10^{-8}$ in 5 iterations for the 1pp shock, and in 8 iterations for the 5pp shock—the latter taking slightly longer due to the larger role of nonlinearities. By contrast, MIT shock methods that rely on ad-hoc adjustment criteria frequently require hundreds of iterations before convergence of equilibrium sequences. This faster convergence to equilibrium with our quasi-Newton based method constitutes our final methodological contribution.[42]

8 Conclusion

This paper presents a highly efficient method for computing heterogeneous-agent models. The core idea is that *sequence-space Jacobians* are sufficient statistics that summarize all we need to know about the heterogeneity in order to determine general equilibrium dynamics. We make five methodological contributions that turn this idea into a systematic algorithm for solving and estimating a large class of models. We provide a fast "fake news" method for computing Jacobians for heterogeneous agents, a technique to substantially reduce dimensionality by representing equilibrium as a directed acyclic graph, a rapid procedure for likelihood-based estimation based on the $MA(\infty)$ representation of the model, a winding-number based determinacy condition for the sequence space, and a Newton-type method to solve nonlinear perfect-foresight transitions.

Our methods allow us to estimate a two-asset HANK model in under ten minutes, which had so far been out of reach for the literature. We hope that they will prove useful to solve and estimate alternative heterogeneous-agent models and facilitate new developments in the field.

[42] Table 1 also reports times for nonlinear impulse responses for all models, which range from 0.18 s for Krusell-Smith to 27 s for two-asset HANK. For comparability, these numbers are for a 1% shock to TFP, which is available in every model.

References

Acharya, Sushant and Keshav Dogra, "Understanding HANK: Insights from a PRANK," *Manuscript*, March 2019.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *Manuscript*, November 2017.

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf, "When Inequality Matters for Macro and Macro Matters for Inequality," *NBER Macroeconomics Annual*, April 2018, *32* (1), 1–75.

Algan, Yann, Olivier Allais, Wouter J. Den Haan, and Pontus Rendahl, "Chapter 6 - Solving and Simulating Models with Heterogeneous Agents and Aggregate Uncertainty," in Karl Schmedders and Kenneth L. Judd, eds., *Handbook of Computational Economics*, Vol. 3 of *Handbook of Computational Economics Vol. 3*, Elsevier, January 2014, pp. 277–324.

An, Sungbae and Frank Schorfheide, "Bayesian Analysis of DSGE Models," *Econometric Reviews*, April 2007, *26* (2-4), 113–172.

Auclert, Adrien and Matthew Rognlie, "Inequality and Aggregate Demand," Working Paper 24280, National Bureau of Economic Research February 2018.

_ , _ , and Ludwig Straub, "The Intertemporal Keynesian Cross," Working Paper 25020, National Bureau of Economic Research September 2018.

Auerbach, Alan J. and Laurence J. Kotlikoff, *Dynamic Fiscal Policy*, Cambridge Cambridgeshire ; New York: Cambridge University Press, April 1987.

Bayer, Christian and Ralph Luetticke, "Solving Heterogeneous Agent Models in Discrete Time with Many Idiosyncratic States by Perturbation Methods," *Manuscript*, July 2018.

_ , _ , Lien Pham-Dao, and Volker Tjaden, "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," *Econometrica*, 2019, *87* (1), 255–290.

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra, "House Prices and Consumer Spending," *Review of Economic Studies*, July 2018, *85* (3), 1502–1542.

Bilbiie, Florin Ovidiu, "Monetary Policy and Heterogeneity: An Analytical Framework," January 2019.

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

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*, April 2018, *89*, 68–92.

Boucekkine, Raouf, "An Alternative Methodology for Solving Nonlinear Forward-Looking Models," *Journal of Economic Dynamics and Control*, May 1995, *19* (4), 711–734.

Box, George E. P. and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

Brumm, Johannes and Simon Scheidegger, "Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models," *Econometrica*, 2017, *85* (5), 1575–1612.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, May 1998, *39* (2), 257–273.

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

Chan, Tony F and Julia A Olkin, "Circulant preconditioners for Toeplitz-block matrices," *Numerical Algorithms*, 1994, *6* (1), 89–101.

Childers, David, "Automated Solution of Heterogeneous Agent Models," *Manuscript*, December 2018.

Conesa, Juan C. and Dirk Krueger, "Social Security Reform with Heterogeneous Agents," *Review of Economic Dynamics*, October 1999, *2* (4), 757–795.

den Haan, Wouter J., "Solving Dynamic Models with Aggregate Shocks and Heterogeneous Agents," *Macroeconomic Dynamics*, June 1997, *1* (2), 355–386.

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

Fernández-Villaverde, Jesús, Galo Nuño, and Samuel Hurtado, "Financial Frictions and the Wealth Distribution," *Manuscript*, March 2019.

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

Gornemann, Nils, Keith Kuester, and Makoto Nakajima, "Doves for the Rich, Hawks for the Poor? Distributional Consequences of Monetary Policy," *Manuscript*, April 2016.

Griewank, Andreas and Andrea Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, Second Edition*, SIAM, November 2008.

Guerrieri, Veronica and Guido Lorenzoni, "Credit Crises, Precautionary Savings, and the Liquidity Trap," *Quarterly Journal of Economics*, August 2017, *132* (3), 1427–1467.

Guren, Adam M., Alisdair McKay, Emi Nakamura, and Jón Steinsson, "What Do We Learn From Cross-Sectional Empirical Estimates in Macroeconomics?," *Manuscript*, November 2018.

Hamilton, James D., *Time Series Analysis*, 1 ed., Princeton University Press, January 1994.

Herbst, Edward P. and Frank Schorfheide, *Bayesian Estimation of DSGE Models*, Princeton University Press, December 2015.

Hopenhayn, Hugo A., "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60* (5), 1127–1150.

Hopenhayn, Hugo and Richard Rogerson, "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, October 1993, *101* (5), 915–938.

Judd, Kenneth L. and Sy-Ming Guu, "Perturbation Solution Methods for Economic Growth Models," in Hal R. Varian, ed., *Economic and Financial Modeling with Mathematica®*, New York, NY: Springer New York, 1993, pp. 80–103.

Juillard, Michel, "Dynare: A Program for the Resolution and Simulation of Dynamic Models with Forward Variables Through the Use of a Relaxation Algorithm," *CEPREMAP Working Paper no 9602*, 1996.

Justiniano, Alejandro, Giorgio E. Primiceri, and Andrea Tambalotti, "Investment Shocks and Business Cycles," *Journal of Monetary Economics*, March 2010, *57* (2), 132–145.

_ , _ , and _ , "Investment Shocks and the Relative Price of Investment," *Review of Economic Dynamics*, 2011.

Kaplan, Greg and Giovanni L. Violante, "Microeconomic Heterogeneity and Macroeconomic Shocks," *Journal of Economic Perspectives*, August 2018, *32* (3), 167–194.

_ , Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, March 2018, *108* (3), 697–743.

Khan, Aubhik and Julia K. Thomas, "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics," *Econometrica*, March 2008, *76* (2), 395–436.

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

Koby, Yann and Christian Wolf, "Understanding Investment Stimulus: From Micro Data to Macro Outcomes," *Manuscript*, August 2018.

Krueger, D., K. Mitman, and F. Perri, "Chapter 11 - Macroeconomics and Household Heterogeneity," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, January 2016, pp. 843–921.

Krusell, Per and Anthony A. Smith, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, October 1998, *106* (5), 867–896.

Laffargue, Jean-Pierre, "Résolution D'un Modèle Macroéconomique Avec Anticipations Rationnelles," *Annales d'Économie et de Statistique*, 1990, (17), 97–119.

Lan, Hong and Alexander Meyer-Gohde, "Solving DSGE Models with a Nonlinear Moving Average," *Journal of Economic Dynamics and Control*, 2013, *37* (12), 2643–2667.

Lubik, Thomas A. and Frank Schorfheide, "Computing Sunspot Equilibria in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, November 2003, *28* (2), 273–285.

Mankiw, N. Gregory and Ricardo Reis, "Sticky Information in General Equilibrium," *Journal of the European Economic Association*, May 2007, *5* (2-3), 603–613.

Mansur, Ahsan and John Whalley, "A Decomposition Algorithm for General Equilibrium Computation with Application to International Trade Models," *Econometrica*, 1982, *50* (6), 1547–1557.

McKay, Alisdair and Ricardo Reis, "The Role of Automatic Stabilizers in the U.S. Business Cycle," *Econometrica*, January 2016, *84* (1), 141–194.

_ , Emi Nakamura, and Jón Steinsson, "The Power of Forward Guidance Revisited," *American Economic Review*, October 2016, *106* (10), 3133–3158.

Mertens, Thomas M. and Kenneth L. Judd, "Solving an Incomplete Markets Model with a Large Cross-Section of Agents," *Journal of Economic Dynamics and Control*, June 2018, *91*, 349–368.

Meyer-Gohde, Alexander, "Linear Rational-Expectations Models with Lagged Expectations: A Synthetic Method," *Journal of Economic Dynamics and Control*, May 2010, *34* (5), 984–1002.

Mongey, Simon and Jerome Williams, "Firm Dispersion and Business Cycles: Estimating Aggregate Shocks Using Panel Data," *Manuscript, New York University*, 2017.

Nardi, Mariacristina De, "Wealth Inequality and Intergenerational Links," *The Review of Economic Studies*, 2004, *71* (3), 743–768.

Onatski, Alexei, "Winding Number Criterion for Existence and Uniqueness of Equilibrium in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, February 2006, *30* (2), 323–345.

Plagborg-Møller, Mikkel, "Bayesian Inference on Structural Impulse Response Functions," *Quantitative Economics*, 2019, *10* (1), 145–184.

_ **and Laura Liu**, "Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data," *Manuscript*, 2019.

Proehl, Elisabeth, "Approximating Equilibria with Ex-Post Heterogeneity and Aggregate Risk," SSRN Scholarly Paper ID 2620937, Social Science Research Network, Rochester, NY February 2019.

Ravn, Morten O. and Vincent Sterk, "Macroeconomic Fluctuations with HANK & SAM: An Analytical Approach," January 2017.

Reiter, Michael, "Recursive Computation of Heterogeneous Agent Models," *Manuscript, UPF*, 2002.

_ , "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, March 2009, *33* (3), 649–665.

_ , "Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models," *IHS Working Paper 258*, 2010.

Schmitt-Grohé, Stephanie and Martín Uribe, "Evaluating the Sample Likelihood of Linearized DSGE Models Without the Use of the Kalman Filter," *Economics Letters*, December 2010, *109* (3), 142–143.

_ **and** _ , "What's News in Business Cycles," *Econometrica*, November 2012, *80* (6), 2733–2764.

Simon, Herbert A., "Dynamic Programming Under Uncertainty with a Quadratic Criterion Function," *Econometrica*, 1956, *24* (1), 74–81.

Sims, Christopher A., "Solving Linear Rational Expectations Models," *Computational Economics*, October 2002, *20* (1), 1–20.

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

Sowell, Fallaw, "A Decomposition of Block Toeplitz Matrices with Applications to Vector Time Series," 1989.

Straub, Ludwig, "Consumption, Savings, and the Distribution of Permanent Income," *Manuscript*, November 2017.

Theil, H., "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica*, 1957, *25* (2), 346–349.

van der Laan, Gerard, "The Computation of General Equilibrium in Economies with a Block Diagonal Pattern," *Econometrica*, 1985, *53* (3), 659–665.

Whiteman, Charles, *Linear Rational Expectations Models: A User's Guide*, Univ Of Minnesota Press, April 1983.

Whittle, Peter, "The analysis of multiple stationary time series," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1953, *15* (1), 125–139.

Winberry, Thomas, "A Method for Solving and Estimating Heterogeneous Agent Macro Models," *Quantitative Economics*, November 2018, *9* (3), 1123–1151–1151.

Wolf, Christian, "The Missing Intercept: A Sufficient Statistics Approach to General Equilibrium Effects," *Manuscript*, May 2019.

Woodford, Michael, *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, August 2003.

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*, January 2010, *34* (1), 36–41.

Appendix to "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models"

A Model descriptions

A.1 Calibration of the Krusell-Smith economy

We follow a standard calibration. We assume that $P(e, e')$ discretizes a log AR(1) process,

$$\log e_t = \rho \log e_{t-1} + \sigma \epsilon_t$$

with normal innovations $\epsilon_t \sim \mathcal{N}(0, 1)$. We use the Rouwenhorst method for discretization. Table A.1 summarizes the rest of our calibration. "Baseline" refers to our baseline calibration with 3500 idiosyncratic states. "HD" refers to our high-dimensional calibration with 250,000 idiosyncratic states.

Table A.1: Calibration of our Krusell-Smith economy

Parameter		Value (baseline)	Value (HD)
r	Real interest rate	0.01	—
σ	Risk aversion	1	—
α	Capital share	0.11	—
δ	Depreciation rate	0.025	—
ρ	Skill mean reversion	0.966	—
$\sigma/\sqrt{1-\rho^2}$	Cross-sectional std of log earnings	0.5	—
n_e	Points in Markov chain for e	7	50
n_k	Points on asset grid	500	5000

A.2 One-asset HANK model

This is an economy without capital, but with nominal rigidities.

Households.

Households now get to choose labor supply in addition to consumption and savings to maximize a standard separable utility function. A household in state e working n_t hours earns labor income $w_t n_t e$, pays a lump-sum tax $\overline{\tau}(e)\,\tau_t$, and receives dividend $\overline{d}(e)\,d_t$ per period from his ownership of a nontradable firm share. The only asset households can trade is one-period nominal bond that

pays r_t in real terms. All in all, the Bellman equation can be written as

$$V_t(e, a_-) = \max_{c,n,a} \left\{ \frac{c^{1-\sigma}}{1-\sigma} - \varphi \frac{n^{1+\nu}}{1+\nu} + \beta \sum_{e'} V_{t+1}(e', a) \mathcal{P}(e, e') \right\}$$
$$c + a = (1 + r_t) a_- + w_t e n - \tau_t \bar{\tau}(e) + d_t \bar{d}(e)$$
$$a \geq 0$$

The solution is a collection of policy functions $c_t(e, a_-)$, $n_t(e, a_-)$ and $a_t(e, a_-)$ that depend on the paths $\{r_s, w_s, \tau_s, d_s\}_{s \geq t}$ that households take as given. Analogously to section 2, we can summarize the household block by aggregate consumption, labor supply and asset demand functions

$$\mathcal{C}_t\left(\{r_s, w_s, \tau_s, d_s\}_{s \geq 0}\right) \equiv \int c_t(a, e) \, dD_t(e, a_-) \tag{51}$$
$$\mathcal{N}_t\left(\{r_s, w_s, \tau_s, d_s\}_{s \geq 0}\right) \equiv \int e n_t(a, e) \, dD_t(e, a_-) \tag{52}$$
$$\mathcal{A}_t\left(\{r_s, w_s, \tau_s, d_s\}_{s \geq 0}\right) \equiv \int a \, dD_t(e, a_-) \tag{53}$$

Firms.

There is a continuum of identical firms that produce differentiated goods using labor only. To preserve symmetry, we assume that firm employs a representative workforce. They engage in monopolistic competition and set the price of their product subject to the usual iso-elastic demand curve and quadratic adjustment costs. The Bellman equation of firm j is

$$J_t(p_{jt-1}) = \max_{y_{jt}, p_{jt}, n_{jt}} \left\{ \frac{p_{jt}}{p_t} y_{jt} - w_t n_{jt} - \frac{\mu}{\mu - 1} \frac{1}{2\kappa} \Big[\log(1 + \pi_{jt}) \Big]^2 Y_t + \frac{J_{t+1}(p_{jt})}{1 + r_{t+1}} \right\}$$
$$\text{s.t. } y_{jt} = Z_t n_{jt}$$
$$y_{jt} = \left(\frac{p_{jt}}{p_t} \right)^{-\frac{\mu}{\mu - 1}} Y_t$$

This is a standard problem that yields the following equilibrium conditions:

- Phillips curve:

$$\log(1 + \pi_t) = \kappa \left(\frac{w_t}{Z_t} - \frac{1}{\mu} \right) + \frac{1}{1 + r_{t+1}} \frac{Y_{t+1}}{Y_t} \log(1 + \pi_{t+1}). \tag{54}$$

- Production:

$$Y_t = Z_t N_t \tag{55}$$

- Price adjustment cost:

$$\psi_t = \frac{\mu}{\mu - 1} \frac{1}{2\kappa} \Big[\log(1 + \pi_t) \Big]^2 Y_t. \tag{56}$$

Table A.2: Calibration of our one-asset HANK economy

Parameter		Value	Target
Households			
β	Discount factor	0.976	$r = 0.0125$
φ	Disutility of labor	0.786	$\mathcal{N} = 1$
σ	Inverse IES	2	
ν	Inverse Frisch	2	
$\underline{b}$	Borrowing constraint	0	
ρ_e	Autocorrelation of earnings	0.966	
σ_e	Cross-sectional std of log earnings	0.5	
Firms			
μ	Steady-state markup	1.2	
κ	Slope of Phillips curve	0.1	
Policy			
B	Bond supply	5.6	
ϕ	Taylor rule coefficient on inflation	1.5	
ϕ_y	Taylor rule coefficient on output	0	
Discretization			
n_e	Points in Markov chain for e	7	
n_a	Points on asset grid	500	

- Dividends:

$$d_t = Y_t - w_t N_t - \psi_t \tag{57}$$

Policy and market clearing

The government runs a balanced budget, maintaining a constant level of bonds B by adjusting taxes. Monetary policy sets the nominal rate according to a Taylor rule that is subject to a shock r_t^*

$$\tau_t = r_t B, \tag{58}$$

$$i_t = r_t^* + \phi \pi_t + \phi_y \left(Y_t - Y_{ss}\right) \tag{59}$$

$$1 + r_t = \frac{1 + i_{t-1}}{1 + \pi_t} \tag{60}$$

Market clearing requires that $B = \mathcal{A}_t, N_t = \mathcal{N}_t$ and $Y_t = C_t + \psi_t$.

Calibration.

The calibration mostly follows McKay et al. (2016) and is summarized in table A.2.

A.3 Two-asset HANK model

This embeds a two-asset household block, described in more detail in section B.1, into a New Keynesian model.

Households.

Income z_{it} is determined as

$$z_{it} = (1 - \tau_t) w_t N_t e_{it} \tag{61}$$

where e_{it} is individual productivity following a Markov process as in the other models. All in all, the household block takes as inputs the sequences of interest rates $\{r_s^a, r_s^b\}$, wage per efficiency units $\{w_s\}$, labor tax rate $\{\tau_s\}$ and labor demand $\{N_s\}$ as inputs. The relevant outputs are illiquid asset demand, liquid asset demand, productivity-weighted marginal utility, consumption, and portfolio adjustment costs:

$$\mathcal{A}_t\left(\{r_s^a, r_s^b, w_s, \tau_s, N_s\}\right) = \int a \, dD_t(e,b,a), \tag{62}$$

$$\mathcal{B}_t\left(\{r_s^a, r_s^b, w_s, \tau_s, N_s\}\right) = \int b_t(e,b,a) \, dD_t(e,b,a), \tag{63}$$

$$\mathcal{U}_t\left(\{r_s^a, r_s^b, w_s, \tau_s, N_s\}\right) = \int e \cdot c_t(e,b,a)^{-\sigma} \, dD_t(e,b,a), \tag{64}$$

$$\mathcal{C}_t\left(\{r_s^a, r_s^b, w_s, \tau_s, N_s\}\right) = \int c_t(e,b,a) \, dD_t(e,b,a), \tag{65}$$

$$\mathcal{P}_t\left(\{r_s^a, r_s^b, w_s, \tau_s, N_s\}\right) = \int \Phi(a_t(e,b,a), a) \, dD_t(e,b,a). \tag{66}$$

The last two are required only for checking the omitted goods market clearing condition.

Labor unions.

We assume that every household provides a continuum of differentiated labor services, each of which is represented by a labor union. Unions set hours and wages to maximize the average utility of members, taking as given their consumption-savings decisions as well as the decisions of other (identical) unions. Changing the nominal wage incurs quadratic adjustment costs. The Bellman equation of union k is

$$\begin{aligned} U_t(w_{kt-1}) = \max_{n_{kt}, w_{kt}} & \int u(c_{it}) - v(n_{kt}) \, dD_t - \frac{\mu_w}{1-\mu_w} \frac{1}{2\kappa_w} \left[\log(1+\pi_{kt}^w)\right]^2 N_t + \beta U_{t+1}(w_{kt}) \\ \text{s.t. } & n_{kt} = \left(\frac{w_{kt}}{w_t}\right)^{-\frac{\mu_w}{\mu_w - 1}} N_t \end{aligned}$$

where π_{kt}^w is wage inflation

$$\pi_{kt}^w = (1+\pi_t)\frac{w_{kt}}{w_{kt-1}} - 1. \tag{67}$$

This setup is almost identical to that in Auclert et al. (2018) and, as shown there, leads to a wage Phillips curve of the form

$$\log(1+\pi_t^w) = \kappa_w\Big[\varphi N_t^{1+\nu} - \mu_w(1-\tau_t)w_t N_t \mathcal{U}_t\Big] + \beta\log(1+\pi_{t+1}^w). \tag{68}$$

For convenience later on, let's introduce the following shorthand for wage adjustment costs

$$\psi_t^w = \frac{\mu_w}{1-\mu_w}\frac{1}{2\kappa_w}\Big[\log(1+\pi_t^w)\Big]^2 N_t. \tag{69}$$

Firms.

Let there be a continuum of identical firms that produce differentiated goods using labor and capital. They own capital, and hire a representative workforce from the labor union. They engage in monopolistic competition and set the price of their product subject to the usual iso-elastic demand curve and quadratic adjustment costs. The Bellman equation of firm j is

$$\begin{aligned} J_t(k_{jt-1}, p_{jt-1}) = \max_{y_{jt},p_{jt},k_{jt},i_{jt},n_{jt}} \Bigg\{ & \frac{p_{jt}}{p_t}y_{jt} - w_t n_{jt} - i_{jt} - \frac{\mu_p}{\mu_p - 1}\frac{1}{2\kappa_p}\Big[\log(1+\pi_{jt})\Big]^2 Y_t \\ & - \frac{1}{2\delta\epsilon_I}\left(\frac{k_{jt}-k_{jt-1}}{k_{jt-1}}\right)^2 k_{jt-1} + \frac{J_{t+1}(k_{jt},p_{jt})}{1+r_{t+1}}\Bigg\} \\ \text{s.t. } y_{jt} &= Z_t k_{jt-1}^{\alpha} n_{jt}^{1-\alpha} \\ y_{jt} &= \left(\frac{p_{jt}}{p_t}\right)^{-\frac{\mu_p}{\mu_p-1}} Y_t \\ k_{jt} &= (1-\delta)\,k_{jt-1} + i_{jt} \end{aligned}$$

This is a standard problem that yields the following equilibrium conditions

- Phillips curve:

$$\log(1+\pi_t) = \kappa_p\left(mc_t - \frac{1}{\mu_p}\right) + \frac{1}{1+r_{t+1}}\frac{Y_{t+1}}{Y_t}\log(1+\pi_{t+1}). \tag{70}$$

- Labor demand:

$$w_t = (1-\alpha)\frac{Y_t}{N_t}mc_t. \tag{71}$$

- Valuation:

$$(1+r_{t+1})Q_t = \alpha\frac{Y_{t+1}}{K_t}mc_{t+1} - \left[\frac{K_{t+1}}{K_t} - (1-\delta) + \frac{1}{2\delta\epsilon_I}\left(\frac{K_{t+1}-K_t}{K_t}\right)^2\right] + \frac{K_{t+1}}{K_t}Q_{t+1}. \tag{72}$$

- Investment:
$$Q_t = 1 + \frac{1}{\delta \epsilon_I} \frac{K_t - K_{t-1}}{K_{t-1}}. \tag{73}$$

- Capital law of motion:
$$I_t = K_t + (1 - \delta) K_{t-1} \tag{74}$$

- Production:
$$Y_t = Z_t K_{t-1}^{\alpha} N_t^{1-\alpha} \tag{75}$$

- Price adjustment cost:
$$\psi_t^p = \frac{\mu_p}{\mu_p - 1} \frac{1}{2\kappa_p} \Big[\log(1 + \pi_t) \Big]^2 Y_t \tag{76}$$

- Dividends:
$$d_t = Y_t - w_t N_t - I_t - \psi_t \tag{77}$$

Policy.

Fiscal policy follows a balanced-budget policy

$$\tau_t w_t N_t = r_t B^g + G_t, \tag{78}$$

Monetary policy follows an interest rate rule

$$i_t = r_t^* + \phi \pi_t + \phi_y \left(Y_t - Y_{ss} \right) \tag{79}$$

$$1 + r_t = \frac{1 + i_{t-1}}{1 + \pi_t} \tag{80}$$

$$r_t^* \text{ is exogenous.} \tag{81}$$

Assets.

The economy has two assets, nominal government bonds and firm equity, both illiquid by default. The ex-post real return on government bonds is simply r_t. Let p_t denote the ex-dividend price of equity. The real return on equity is

$$\frac{d_{t+1} + p_{t+1}}{p_t}.$$

In the absence of aggregate uncertainty, these assets have to earn the same return and so we get the following no arbitrage condition

$$p_t = \frac{d_{t+1} + p_{t+1}}{1 + r_{t+1}}. \tag{82}$$

If bonds and equity are both illiquid, where do liquid assets come from? We assume that there is a representative financial intermediary endowed with the technology to turn illiquid nominal

assets into liquid nominal assets. This process has a proportional cost ω. This simple setup implies that

$$r_t^b = r_t - \omega. \tag{83}$$

Market clearing.

The only market clearing conditions not implied implicitly by notation are that of goods and assets

$$Y_t = \mathcal{C}_t + G_t + I_t + \mathcal{P}_t + \omega \mathcal{B}_t + \psi_t^p + \psi_t^w, \tag{84}$$

$$\mathcal{A}_t + \mathcal{B}_t = p_t + B^g. \tag{85}$$

Accounting for surprise inflation and capital gains.

When an unanticipated shock hits, the no arbitrage condition (82) will fail for one period due to surprise inflation and capital gains. Accounting for these effects requires further assumptions on the illiquid portfolio of households. For simplicity, we assume that all households hold the same ratio of bonds and equity. This means that the ex-post illiquid return can be written as

$$1 + r_t^a = \frac{p_{t-1}}{\mathcal{A}_{t-1}} \cdot \frac{d_t + p_t}{p_{t-1}} + \frac{B^g - \mathcal{B}_{t-1}}{\mathcal{A}_{t-1}} \cdot (1 + r_t). \tag{86}$$

This equation holds both in periods with and without unexpected shocks. We could stop here, but then we would have to add r_t^a to the unknowns of the DAG. The reason is that it is an input of the household block that also depends on outputs of the household block. We can avoid this by replacing (86) with

$$1 + r_t^a = \frac{p}{\mathcal{A}} \cdot \frac{d_t + p_t}{p_{t-1}} + \frac{B^g - \mathcal{B}}{\mathcal{A}} \cdot (1 + r_t). \tag{87}$$

The difference is the use of steady-state portfolio shares. These are targeted in the calibration, and thus effectively exogenous. Importantly, they are the correct shares in period 0, when the unexpected shock hits. Although they are incorrect in periods $t > 0$, this does not matter, because by then the no arbitrage condition (82) is restored.

Calibration. Table A.3 summarizes the calibration of our two-asset model.

B Computational details

B.1 Two-asset household model algorithm

In this section we describe a generic two-asset household model with convex adjustment costs in an illiquid asset whose return is superior to that of a liquid asset, in the spirit of Kaplan et al. (2018). We then describe an efficient algorithm, based on the endogenous gridpoints approach of Carroll (2006), to solve this model.

Table A.3: Calibration of our two-asset HANK economy

Parameter		Value	Target
Households			
β	Discount factor	0.976	$r = 0.0125$
σ	Inverse IES	2	
χ_0	Portfolio adj. cost pivot	0.25	
χ_1	Portfolio adj. cost scale	6.416	$\mathcal{B} = 1.04Y$
χ_2	Portfolio adj. cost curvature	2	
$\underline{b}$	Borrowing constraint	0	
ρ_e	Autocorrelation of earnings	0.966	
σ_e	Cross-sectional std of log earnings	0.92	
Labor unions			
φ	Disutility of labor	2.073	$N = 1$
ν	Inverse Frisch elasticity	1	
μ_w	Steady state wage markup	1.1	
κ_w	Slope of wage Phillips curve	0.1	
Firms			
Z	TFP	0.468	$Y = 1$
α	Capital share	0.33	$K = 10Y$
μ_p	Steady-state markup	1.015	$\mathcal{A} + \mathcal{B} = 14Y$
δ	Depreciation	0.02	
κ_p	Slope of price Phillips curve	0.1	
Financial intermediary			
ω	liquidity premium	0.005	
Policy			
τ	Labor tax	0.356	budget balance
G	Government spending	0.2	
B^g	Bond supply	2.8	
ϕ	Taylor rule coefficient	1.5	
ϕ_y	Taylor rule coefficient on output	0	
Discretization			
n_e	Points in Markov chain for e	3	
n_b	Points on liquid asset grid	50	
n_a	Points on illiquid asset grid	70	

Generic setup. Households' individual state variables are (exogenous) income $z \in \{z_1, \ldots, z_m\}$, liquid assets $b \in [\underline{b}, \infty)$, and illiquid assets $a \in [0, \infty)$. Adjusting the illiquid account incurs a convex cost $\Phi(a_t, a_{t-1})$. The Bellman equation is

$$\begin{aligned} V_t(z_t, b_{t-1}, a_{t-1}) = \max_{c_t, b_t, a_t} \; & u(c_t) + \beta \mathbb{E}_t V_{t+1}(z_{t+1}, b_t, a_t) \\ \text{s.t. } & c_t + a_t + b_t = z_t + (1 + r_t^a) a_{t-1} + (1 + r_t^b) b_{t-1} - \Phi(a_t, a_{t-1}) \\ & a_t \geq 0, \quad b_t \geq \underline{b}. \end{aligned}$$

Let the adjustment cost function be specified as

$$\Phi(a_t, a_{t-1}) = \frac{\chi_1}{\chi_2} \left| \frac{a_t - (1 + r_t^a) a_{t-1}}{(1 + r_t^a) a_{t-1} + \chi_0} \right|^{\chi_2} [(1 + r_t^a) a_{t-1} + \chi_0] \tag{88}$$

with $\chi_0, \chi_1 > 0$ and $\chi_2 > 1$. Note that $\Phi(a_t, a_{t-1})$ is bounded, differentiable, and convex in the choice a_t.[43]

First-order and envelope conditions. The Bellman equation can be rewritten more compactly as

$$\begin{aligned} V_t(z_t, b_{t-1}, a_{t-1}) = \max_{b_t, a_t} \; & u\left(z_t + (1 + r_t^a) a_{t-1} + (1 + r_t^b) b_{t-1} - \Phi(a_t, a_{t-1}) - a_t - b_t \right) \\ & + \lambda_t (b_t - \underline{b}) + \mu_t a_t + \beta \mathbb{E} V_{t+1}(z_{t+1}, b_t, a_t) \end{aligned}$$

The first-order conditions with respect to b_t and a_t are

$$u'(c_t) = \lambda_t + \beta \mathbb{E} \partial_b V_{t+1}(z_{t+1}, b_t, a_t), \tag{89}$$

$$u'(c_t) \left[1 + \Phi_1(a_t, a_{t-1}) \right] = \mu_t + \beta \mathbb{E} \partial_a V_{t+1}(z_{t+1}, b_t, a_t), \tag{90}$$

and the envelope conditions are

$$\partial_b V_t(z_t, b_{t-1}, a_{t-1}) = (1 + r_t^b) u'(c_t), \tag{91}$$

$$\partial_a V_t(z_t, b_{t-1}, a_{t-1}) = \left[1 + r_t^a - \Phi_2(a_t, a_{t-1}) \right] u'(c_t). \tag{92}$$

It's convenient to define the *post-decision value function* $W_t(z_t, b_t, a_t) \equiv \beta \mathbb{E}_t V_{t+1}(z_t, b_t, a_t)$. Note that the partials of this are just what we have on the right-hand side of the Euler equations (89) and (90).

Algorithm. The algorithm is a variant of the endogenous gridpoint method of Carroll (2006) that we developed for this two-asset problem. The key trick is, whenever the household is partially constrained, to include Lagrange multipliers in the backward iteration. We also exploit the fact

[43] The functional form (88) is chosen for concreteness; more generally, we could have a mix of terms with different $\chi_2 > 1$.

that, endogenously, the constraint on the illiquid asset will never be binding unless the constraint on the liquid asset is also binding (otherwise, a simple variation will improve utility)—and if both are binding, then the policy is trivial.

Overall, we start from a guess for the (discretized) partials of the value function and iterate backward until convergence. Throughout, we will use (z', b', a') to refer to *tomorrow's grid* and (z, b, a) *today's grid*.

1. **Initial guess.** Guess $V_a(z', b', a')$ and $V_b(z', b', a')$.

2. **Common** $z' \to z$. By definition

$$W_b(z, b', a') = \beta \Pi V_b(z', b', a') \tag{93}$$
$$W_a(z, b', a') = \beta \Pi V_a(z', b', a') \tag{94}$$

3. **Unconstrained** $\underline{a' \to a}$. Assuming that no constraints bind, $\lambda_t = \mu_t = 0$, and (89) and (90) become

$$u'(c) = W_b(z, b', a'), \tag{95}$$
$$u'(c)\Big[1 + \Phi_1(a', a)\Big] = W_a(z, b', a'). \tag{96}$$

   Combine these to get

$$0 = F(z, b', a, a') \equiv \frac{W_a(z, b', a')}{W_b(z, b', a')} - 1 - \Phi_1(a', a) \tag{97}$$

   which characterizes $a'(z, b', a)$. Use this to map $W_b(z, b', a')$ into $W_b(z, b', a)$ by interpolation, then compute consumption as

$$c(z, b', a) = W_b(z, b', a)^{-\frac{1}{\sigma}}. \tag{98}$$

4. **Unconstrained** $\underline{b' \to b}$. Now using $a'(z, b', a)$ and $c(z, b', a)$ from the previous step, use the budget constraint to obtain

$$b(z, b', a) = \frac{c(z, b', a) + a'(z, b', a) + b' - (1 + r^a)a + \Phi(a'(z, b', a), a) - z}{1 + r^b}.$$

   We invert this function via interpolation to get $b'(z, b, a)$. The same interpolation weights can be used to do $a'(z, b', a) \to a'(z, b, a)$.

5. **Liquidity constrained** $\underline{a' \to a}$. This branch is analogous to the unconstrained case. Assum-

ing that the liquidity constraint is binding, $\lambda_t > 0$, and (89) and (90) become

$$u'(c) = \lambda + W_b(z, 0, a'),$$
$$u'(c)\Big[1 + \Phi_1(a', a)\Big] = W_a(z, 0, a').$$

To help with scaling, let us define $\kappa \equiv \lambda / W_b(z, 0, a')$ and rewrite the first equation as

$$u'(c) = (1 + \kappa) W_b(z, 0, a').$$

Divide and rearrange to get

$$0 = F(z, \kappa, a, a') \equiv \frac{1}{1+\kappa} \frac{W_a(z, 0, a')}{W_b(z, 0, a')} - 1 - \Phi_1(a', a). \tag{99}$$

We solve this for $a'(z, \kappa, a)$, and compute consumption as

$$c(z, \kappa, a) = \big[(1 + \kappa) W_b(z, \kappa, a)\big]^{-\frac{1}{\sigma}}. \tag{100}$$

6. **Liquidity constrained** $\underline{\kappa \to b}$. Now using $a'(z, \kappa, a)$ and $c(z, \kappa, a)$ from the previous step, use the budget constraint to obtain

$$b(z, \kappa, a) = \frac{c(z, \kappa, a) + a'(z, \kappa, a) + \underline{b} - (1 + r^a)a + \Phi(a'(z, \kappa, a), a) - z}{1 + r^b}.$$

We invert this function via interpolation to get $\kappa(z, b, a)$. The same interpolation weights can be used to map $a'(z, \kappa, a)$ into $a'(z, b, a)$. We already know that $b'(z, b, a) = \underline{b}$.

7. **Update guesses.** The final $b'(z, b, a)$ is the element-wise maximum of its unconstrained and liquidity-constrained counterparts. Replace the unconstrained $a'(z, b, a)$ with constrained one at the exact same points. Compute consumption from the budget constraint as

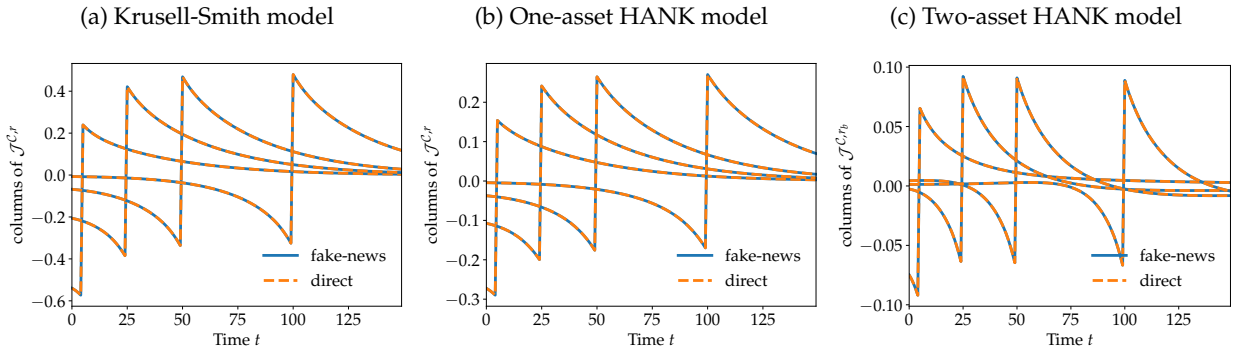
$$c(z, b, a) = z + (1 + r^a)a + (1 + r^b)b - \Phi(a'(z, b, a), a) - a'(z, b, a) - b'(z, b, a). \tag{101}$$

Finally use the envelope conditions (91) and (92) to update the guesses

$$V_b(z, b, a) = (1 + r^b) c(z, b, a)^{-\sigma}, \tag{102}$$
$$V_a(z, b, a) = \Big[1 + r^a - \Phi_2\Big(a'(z, b, a), a\Big)\Big] c(z, b, a)^{-\sigma}. \tag{103}$$

Go back to step 2, repeat until convergence.

Figure B.1: Jacobians $\mathcal{J}^{\mathcal{C},r}$ for all three models via the direct and fake news method.



B.2 Accuracy check for fake news and direct algorithm

Figure B.1 performs a visual accuracy check for each of our three models, by plotting different columns of the Jacobians $\mathcal{J}^{\mathcal{C},r}$ of the model's aggregate consumption with respect to real interest rates (in the two-asset model, this is an impulse to the liquid asset rate).

Recall that these columns correspond to the impulse responses of aggregate consumption to a small increase impulse in real interest rates at various dates. In each of our models we observe the standard intertemporal substitution pattern in models with incomplete markets—agents in the aggregate lower their consumption in anticipation of the shock, with the effect building up over time as agents become progressively more unconstrained, and raise their consumption after the shock occurs.

This check makes clear that our two algorithms compute exactly the same Jacobians—the fake news algorithm just does so much faster.

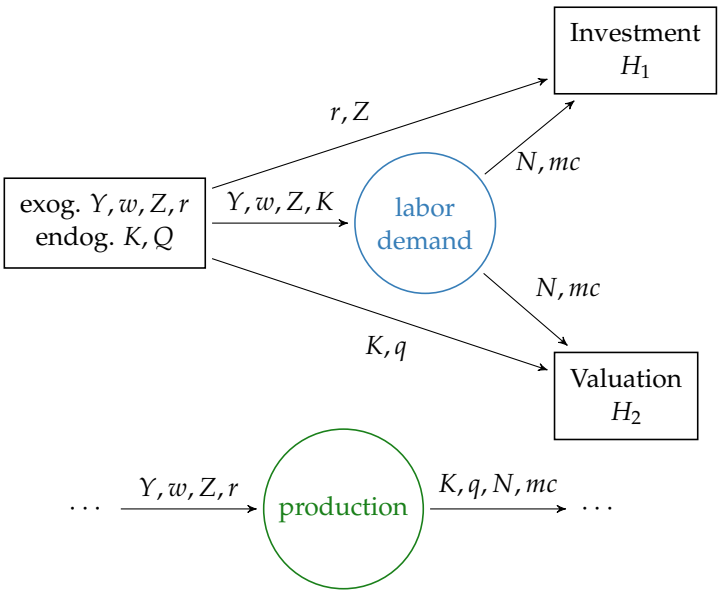
B.3 Solved blocks

The two-asset model introduced via the DAG in figure 4 included a green "production" block. Production with adjustment costs is well-known to involve the joint determination of investment and q, and it is natural to solve for these two jointly inside a block. This leads us to introduce a "solved block" concept, as follows:

Definition 5. A *solved block* b has an underlying SHADE model with shocks $\tilde{\mathcal{Z}}$, unknowns $\tilde{\mathcal{U}}$, outputs $\tilde{\mathcal{O}}$, and targets $\tilde{\mathcal{T}}$, and an equilibrium that is locally unique around the steady state, where we define:

1. The inputs of the solved block to be the shocks of the underlying SHADE model: $\mathcal{I}_b \equiv \tilde{\mathcal{Z}}$.
2. The outputs of the solved block to be the unknowns and outputs, minus targets, of the underlying SHADE model: $\mathcal{O}_b \equiv \tilde{\mathcal{U}} \cup (\tilde{\mathcal{O}} \setminus \tilde{\mathcal{T}})$.
3. For each output $o \in \mathcal{O}_b$, the function $h^o(\{\mathbf{x}^i\}_{i\in\mathcal{I}_b})$ is the locally unique equilibrium path of o in the underlying SHADE model given sequences $\{\mathbf{x}^i\}_{i\in\tilde{\mathcal{Z}}}$ for the exogenous shocks in that

Figure B.2: The concept of a solved block, applied to the production block of our two-asset HANK model



model (recalling that $\mathcal{I}_b = \tilde{\mathcal{Z}}$).

Informally, a solved block is a SHADE model, turned into a block. Figure B.2 illustrates how this concept works in the case of the production block of our two-asset HANK model. Given the exogenous inputs Y, w, Z, r, the solved block solves for the endogenous paths for K and Q that jointly satisfy the q theory equations, so that its outputs are K, Q as well as labor demand N and marginal costs mc.

B.4 Fast solution for individual impulse responses

In the case where we are only interested in a single impulse response, we only need to do full forward accumulation (37) for $i \in \mathcal{U}$ to obtain $o \in \mathcal{H}$, which gives $\mathbf{H_U} = \mathbf{J}^{\mathcal{H},\mathcal{U}}$. Then, to deal with shocks, we do forward accumulation on *vectors* rather than matrices, writing

$$\mathbf{J}^{o,\mathcal{Z}} d\mathbf{Z} = \sum_{m \in \mathcal{I}_b} \mathcal{J}^{o,m} \mathbf{J}^{m,\mathcal{Z}} d\mathbf{Z} \tag{104}$$

This gives $\mathbf{H_Z} d\mathbf{Z} = \mathbf{J}^{\mathcal{H},\mathcal{Z}} d\mathbf{Z}$. We then solve the linear system $\mathbf{H_U} d\mathbf{U} = -\mathbf{H_Z} d\mathbf{Z}$ to obtain $d\mathbf{U}$. Finally, to obtain equilibrium impulse responses $d\mathbf{X}^o$ for $o \notin \mathcal{Z} \cup \mathcal{U}$, we need to calculate

$$d\mathbf{X}^o = \mathbf{J}^{o,\mathcal{Z}} d\mathbf{Z} + \mathbf{J}^{o,\mathcal{U}} d\mathbf{U} \tag{105}$$

The first term, $\mathbf{J}^{o,\mathcal{Z}} d\mathbf{Z}$, has already been calculated in (104). For the second term, we do forward accumulation on vectors as in (104), just solving for $\mathbf{J}^{o,\mathcal{U}} d\mathbf{U}$ rather than $\mathbf{J}^{o,\mathcal{Z}} d\mathbf{Z}$.[44]

[44] Another approach is to use the $\mathbf{J}^{o,\mathcal{U}}$ that we already calculated as part of the initial forward accumulation to obtain $\mathbf{H_U} = \mathbf{J}^{\mathcal{H},\mathcal{U}}$, and directly apply these to $d\mathbf{U}$. This approach has similar (and low) cost, but is less useful in general

Table B.1: Computing times for impulse responses.

	Krusell-Smith	one-asset HANK	two-asset HANK
Without directed acyclic graph (DAG)	**25.6 ms**	**99.7 ms**	**1114.3 ms**
With DAG	**1.2 ms**	**17.4 ms**	**121.7 ms**
step 1 (forward accumulate $\mathbf{H_U}$)	0.5 ms	7.7 ms	109.8 ms
step 2 (forward accumulate $\mathbf{J}^{o,\mathcal{Z}}d\mathbf{Z}$)	0.1 ms	0.1 ms	0.8 ms
step 3 (solve linear system for $d\mathbf{U}$)	0.5 ms	9.1 ms	9.2 ms
step 4 (forward accumulate $\mathbf{J}^{o,\mathcal{U}}d\mathbf{U}$, get $d\mathbf{X}^o$)	0.1 ms	0.5 ms	1.8 ms
No. of unknowns (without DAG)	3	7	18
No. of unknowns (with DAG)	1	3	3
No. of exogenous shocks	1	3	7

Table B.1, under "with DAG", shows the time each step of this process takes for our three models, starting from the Jacobians $\mathcal{J}$ for each model block. In general, this process is very cheap, with the only costly parts being the steps that involve matrices rather than vectors: forward accumulation in step 1 to get $\mathbf{H_U} = \mathbf{J}^{\mathcal{H},\mathcal{U}}$, and second, solving the linear system $\mathbf{H_U}d\mathbf{U} = -\mathbf{H_Z}d\mathbf{Z}$ for $d\mathbf{U}$ in step 3.

Since these steps are costly because they involve the shock-independent matrix $\mathbf{H_U}$, there are clear economies of scale from computing the impulse response to multiple shocks. We can calculate $\mathbf{H_U}$ a single time, and then also calculate $\mathbf{H_U}^{-1}$ (or, better, an LU factorization of $\mathbf{H_U}$) a single time, at which point the marginal cost of computing additional impulse responses is very low. This is the approach we use in section 5 to evaluate the likelihood when redrawing model parameters, since this involves finding impulse responses to each shock simultaneously. Taking this idea to its fullest extent, we can calculate the impulse responses to all shocks simultaneously, which is the "$\mathbf{G}$ matrix" approach in section 4.2.

B.5 Efficient second moment and likelihood calculation

Fast Fourier transform (FFT) to compute autocovariances. To start, consider any sequences $a_0, \ldots, a_{T-1}$ and $b_0, \ldots, b_{T-1}$ of real scalars. If we define the sequences

$$\begin{aligned}(\hat{a}_0, \ldots, \hat{a}_{2T-2}) &= (a_0, \ldots, a_{T-1}, 0, \ldots, 0)\\ (\hat{b}_0, \ldots, \hat{b}_{2T-2}) &= (b_0, \ldots, b_{T-1}, 0, \ldots, 0)\end{aligned}$$

to be a and b each padded by $T-1$ zeros, then

$$a_0 b_u + a_1 b_{u+1} + \ldots + a_{T-1-u} b_{T-1} = \sum_{\ell=0}^{2T-2} \hat{a}_\ell \hat{b}_{u+\ell} \tag{106}$$

because it does give o that were not necessary in calculating $\mathbf{H_U}$.

where $\hat{b}_{u+\ell} \equiv \hat{b}_{u+\ell-(2T-2)}$ when $u+\ell \geq 2T-2$. It then follows from the standard properties of the discrete Fourier transform $\mathcal{F}$ that for any $u \in 0, \ldots, T-1$

$$\sum_{\ell=0}^{2T-2} \hat{a}_\ell \hat{b}_{u+\ell} = \left(\mathcal{F}^{-1}\left(\mathcal{F}(\hat{a})^* \cdot \mathcal{F}(\hat{b})\right)\right)_s \tag{107}$$

where $*$ denotes complex conjugation.[45]

Since the discrete Fourier transform is a linear operator, we can extend this method to apply to the matrices $\mathbf{m}_0^{\tilde{\mathcal{O}},\mathcal{Z}}, \ldots, \mathbf{m}_{T-1}^{\tilde{\mathcal{O}},\mathcal{Z}}$ in (43), where we interpret $\mathcal{F}$ as applying element-by-element to a sequence of matrices. Letting $\hat{\mathbf{m}}_0^{\tilde{\mathcal{O}},\mathcal{Z}}, \ldots, \hat{\mathbf{m}}_{2T-2}^{\tilde{\mathcal{O}},\mathcal{Z}}$ denote the sequence padded with zeros like above, we have from (106) and (107), substituting $u = t' - t$, that

$$[\mathbf{m}_0^{\tilde{\mathcal{O}},\mathcal{Z}}][\mathbf{m}_{t'-t}^{\tilde{\mathcal{O}},\mathcal{Z}}]' + \ldots + [\mathbf{m}_{T-1-(t'-t)}^{\tilde{\mathcal{O}},\mathcal{Z}}][\mathbf{m}_{T-1}^{\tilde{\mathcal{O}},\mathcal{Z}}]' = \left(\mathcal{F}^{-1}\left(\mathcal{F}(\widehat{\Phi})^* \mathcal{F}(\widehat{\Phi}')\right)\right)_{t'-t} \tag{108}$$

where the transpose is applied individually to each matrix in the sequence, and $\mathcal{F}(\widehat{\Phi})^* \mathcal{F}(\widehat{\Phi}')$ is the product of each pair of matrices in the frequency-by-frequency sequence.

We simply apply (108), using the fast Fourier transform for $\mathcal{F}$, to calculate the covariances in (43) for each $t' - t$. Since the two key operations—the FFT and matrix multiplication—have extremely efficient implementations widely available, this can be done very quickly, taking only a few milliseconds in table 7 for the examples in this paper. It is far faster than a naive calculation of the sum in (43).[46]

Calculating the log-likelihood. To go from $\text{Cov}(d\tilde{\mathbf{X}}_t, d\tilde{\mathbf{X}}_{t'})$ for each $t' - t$ to the log-likelihood $\mathcal{L}$ in (45), there are two steps. First, we need to obtain $\text{Cov}(d\tilde{\mathbf{X}}_t^{obs}, d\tilde{\mathbf{X}}_{t'}^{obs})$, which we can do easily for each $t' - t$ by directly applying (44). Then, given the matrix $\mathbf{V}$ comprised of blocks $\text{Cov}(d\tilde{\mathbf{X}}_t^{obs}, d\tilde{\mathbf{X}}_{t'}^{obs})$, we need to evaluate the log-likelihood (45). The key steps here are the calculation of the log-determinant $\log(\det(\mathbf{V}))$ and the quadratic form $\mathbf{w}'\mathbf{V}^{-1}\mathbf{w}$.

Standard approach: Cholesky. The simplest approach is the standard one: to explicitly form the matrix $\mathbf{V}$ and then to take a Cholesky decomposition $\mathbf{V} = LL'$ using standard routines.[47] The log-determinant is then just twice the sum of the log entries of L. We can also write $[d\tilde{\mathbf{X}}^{obs}]'\mathbf{V}^{-1}[d\tilde{\mathbf{X}}^{obs}] = (L^{-1}d\tilde{\mathbf{X}}^{obs})'(L^{-1}d\tilde{\mathbf{X}}^{obs})$, where $L^{-1}d\tilde{\mathbf{X}}^{obs}$ is quick to evaluate. The bottleneck is that the cost of the Cholesky decomposition grows asymptotically with the cube of the matrix dimension $n_{obs}T_{obs}$, which can become costly if the number T_{obs} of time periods in the series or the number n_{obs} of variables observed in each period is high. But in practice, for the estimation exercises in table 7, this is not a bottleneck, except perhaps in the two-asset HANK case where we estimate the shock

[45] The padding with zeros to create $\hat{a}$ and $\hat{b}$ is necessary so that the wraparound $\hat{b}_{s+\ell}$ terms for large $s+\ell$ do not affect the sum.

[46] Since the inputs are real, the full transform is redundant and we can deal only with the first T entries; the final $T-2$ are complex conjugates of entries 1 through $T-1$. This economizes on the time for $\mathcal{F}$ and also for matrix multiplication.

[47] This is a longstanding approach in time series to explicitly evaluating the log-likelihood, first applied in the DSGE literature (to our knowledge) by Mankiw and Reis (2007).

process underlying 7 observable series, and calculating $\mathcal{L}$ takes (dominated by the Cholesky) 7 ms out of 11 ms total—i.e. a bottleneck only because other parts of the computation are so fast.

Potential alternatives for large cases. The Cholesky decomposition only requires symmetry, but $\mathbf{V}$ has additional structure: it is also *block Toeplitz*, since the blocks corresponding to each $t' - t$ are the same. An algorithm exploiting this structure to obtain a block Cholesky decomposition, descending from Levinson-Durbin recursion, is described in detail in Sowell (1989), and a form of it is applied to calculate the likelihood by Meyer-Gohde (2010). It has complexity $\mathcal{O}(T_{obs}^2 n_{obs}^3)$, rather than $\mathcal{O}(T_{obs}^3 n_{obs}^3)$ for the standard Cholesky. We have found, however, that the overhead in implementing this algorithm, versus the extremely efficient machine-level implementations of Cholesky, means that it only becomes superior in practice when T_{obs} is very high (over 400 periods).

Several other options remain. One is to use the Whittle (1953) approximation to the likelihood function. We did not use this, since we found that the difference between the approximation and the exact likelihood function was sizable in some instances. This may not matter in many applications, however, and the Whittle likelihood has an extensive history in time series estimation. Another option is to use iterative methods, at least for calculating the quadratic form $\mathbf{w}'\mathbf{V}_w^{-1}\mathbf{w}$; since $\mathbf{V}_w$ is Toeplitz, it is possible to calculate a matrix-vector product using $\mathbf{V}_w$ very quickly using the FFT. We have obtained early promising results using the preconditioned conjugate gradient method and a circulant preconditioner—see, e.g., Chan and Olkin (1994).

C Proofs

C.1 Proof of proposition 1

Consider the term in entry t, s of the fake news matrix for $t, s > 0$:

$$\mathcal{F}_{t,s} = \mathcal{P}_{t-1}'\mathcal{D}_s = \mathbf{y}_{ss}'(\Lambda_{ss}')^{t-1}(\Lambda_v(\mathbf{v}_\mathbf{v})^{s-1}\mathbf{v}_x)'\mathbf{D}_{ss} \tag{109}$$

Since Λ is always a row-stochastic Markov matrix, the entries of vector $(\Lambda_v(\mathbf{v}_\mathbf{v})^{s-1}\mathbf{v}_x)'\mathbf{D}_{ss}$ must always sum to zero (a perturbation to the transition matrix still conserves mass). Multiplying by Λ_{ss}', which is column-stochastic, preserves this property.

If we let $B = B' = I - \mathbf{1}\mathbf{1}'/n_g$ be the *demeaning* matrix, which does nothing to vectors whose entries sum to zero, it follows that (109) equals

$$\mathbf{y}_{ss}'(B'\Lambda_{ss}')^{t-1}(\Lambda_v(\mathbf{v}_\mathbf{v})^{s-1}\mathbf{v}_x)'\mathbf{D}_{ss} \tag{110}$$

Now, $\mathcal{D}_s = (\Lambda_v(\mathbf{v}_\mathbf{v})^{s-1}\mathbf{v}_x)'\mathbf{D}_{ss}$ converges uniformly to zero as $s \to \infty$ at an exponential rate bounded by the largest magnitude eigenvalue of $\mathbf{v}_\mathbf{v}$, which the proposition assumes is inside the unit circle. Similarly, $\tilde{\mathcal{P}}_t \equiv (\Lambda_{ss}B)^{t-1}\mathbf{y}_{ss}$ converges to zero at a rate bounded by the largest eigenvalue of Λ_{ss} that is not the eigenvalue associated with the right eigenvector $\mathbf{1}$, which is annihilated

by the demeaning matrix B. This eigenvalue is strictly inside the unit circle precisely when there is a unique ergodic distribution, like the proposition assumes, and hence $\tilde{\mathcal{P}}_t$ converges uniformly to zero as $t \to \infty$ at an exponential rate bounded by this eigenvalue.

Since both $\mathcal{D}_s$ and $\tilde{\mathcal{P}}_t$ converge to zero as $s \to \infty$ and $t \to \infty$ bounded by an exponential rate (less than 1), the product $\mathcal{F}_{t,s} = \mathcal{P}'_{t-1}\mathcal{D}_s$ converges to zero as $s + t \to \infty$ at an exponential rate bounded by the maximum of these two rates.

C.2 Proof of proposition 2

Here, we derive the rules for multiplication of the operator $Q_{i,m} = \begin{cases} S_i Z_m & i > 0 \\ Z_m S_i & i < 0 \end{cases}$, where S_i is the shift operator on sequences by i and Z_m zeros out the first m elements of sequences, by doing case-by-case analysis on the product

$$Q_{i,m}Q_{j,n}$$

In our derivation, we will exploit the following fact about multiplication of S_i:

$$S_i S_j = \begin{cases} S_{i+j}Z_{-j} & i > 0,\ j < 0,\ i + j > 0 \\ Z_i S_{i+j} & i > 0,\ j < 0,\ i + j < 0 \\ S_{i+j} & \text{otherwise} \end{cases}$$

and the rules $S_{-i}Z_j = Z_{\max(j-i,0)}S_{-i}$ and $Z_j S_i = S_i Z_{\max(j-i,0)}$ for multiplication of S and Z.

Case 1: positive i, positive j. Here we have

$$\begin{aligned} Q_{i,m}Q_{j,n} &= S_i Z_m S_j Z_n \\ &= S_i S_j Z_{\max(m-j,0)} Z_n \\ &= S_{i+j} Z_{\max(m-j,n)} \\ &= Q_{i+j,\max(m-j,n)} \end{aligned} \tag{111}$$

Case 2: positive i, negative j. Here we have

$$\begin{aligned} Q_{i,m}Q_{j,n} &= S_i Z_m Z_n S_j \\ &= S_i Z_{\max(m,n)} S_j \end{aligned}$$

If $i + j > 0$, then we write

$$Z_{\max(m,n)}S_j = S_j Z_{\max(m,n)-j}$$

and then

$$\begin{aligned} S_i Z_{\max(m,n)} S_j &= S_i S_j Z_{\max(m,n)-j} \\ &= S_{i+j} Z_{-j} Z_{\max(m,n)-j} \\ &= S_{i+j} Z_{\max(m,n)-j} \\ &= Q_{i+j,\max(m,n)-j} \end{aligned}$$

If $i + j < 0$, then we write

$$S_i Z_{\max(m,n)} = Z_{\max(m,n)+i} S_i$$

and then

$$\begin{aligned} S_i Z_{\max(m,n)} S_j &= Z_{\max(m,n)+i} S_i S_j \\ &= Z_{\max(m,n)+i} Z_i S_{i+j} \\ &= Z_{\max(m,n)+i} S_{i+j} \\ &= Q_{i+j,\max(m,n)+i} \end{aligned}$$

Both these cases boil down to the simpler form

$$Q_{i,m} Q_{j,n} = Q_{i+j,\max(m,n)+\min(i,-j)} \tag{112}$$

Case 3: negative i, positive j. Then we have

$$\begin{aligned} Q_{i,m} Q_{j,n} &= Z_m S_i S_j Z_n \\ &= Z_m S_{i+j} Z_n \end{aligned}$$

If $i + j > 0$, then we write $Z_m S_{i+j} = S_{i+j} Z_{\max(m-i-j,0)}$ and get

$$Q_{i,m} Q_{j,n} = Q_{i+j,\max(m-i-j,n)} \tag{113}$$

If $i + j < 0$, then we write $S_{i+j} Z_n = Z_{\max(n+i+j,0)} S_{i+j}$ and get

$$Q_{i,m} Q_{j,n} = Q_{i+j,\max(n+i+j,m)} \tag{114}$$

Case 4: negative i, negative j. Then we have

$$\begin{aligned} Q_{i,m} Q_{j,n} &= Z_m S_i Z_n S_j \\ &= Z_m Z_{\max(n+i,0)} S_i S_j \\ &= Z_{\max(m,n+i)} S_{i+j} \\ &= Q_{i+j,\max(m,n+i)} \end{aligned} \tag{115}$$

Combined, (111)-(115) give (39) and (40) in proposition 2.

C.3 Equivalence between SHADE impulse responses and $MA(\infty)$

Throughout, use a filtration $\{\mathcal{F}_t\}$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We denote all stochastic variables and processes with a tilde. We begin by defining stochastic heterogeneous-agent blocks.

Definition 6. A *stochastic heterogeneous-agent block* maps stochastic processes $\tilde{\mathbf{X}}^i$, for $i = 1, \ldots, n_x$, to stochastic processes $\tilde{\mathbf{Y}}^o$, for $o = 1, \ldots, n_y$. Stacking the $\tilde{\mathbf{X}}^i$s into a vector $\tilde{\mathbf{X}}$, and the $\tilde{\mathbf{Y}}^o$s into a vector $\tilde{\mathbf{Y}}$, we represent this map with a function

$$\tilde{\mathbf{Y}} = h(\tilde{\mathbf{X}}) \tag{116}$$

The time-t output of this function can be written as

$$\tilde{\mathbf{Y}}_t = y(\mathbb{E}_t \tilde{\mathbf{v}}_{t+1}, \tilde{\mathbf{X}}_t)' \tilde{\mathbf{D}}_t$$

where $\{\tilde{\mathbf{D}}_t, \tilde{\mathbf{v}}_t\}$ satisfy

$$\begin{aligned} \tilde{\mathbf{D}}_{t+1} &= \Lambda(\mathbb{E}_t \tilde{\mathbf{v}}_{t+1}, \tilde{\mathbf{X}}_t)' \tilde{\mathbf{D}}_t \\ \tilde{\mathbf{v}}_t &= v(\mathbb{E}_t \tilde{\mathbf{v}}_{t+1}, \tilde{\mathbf{X}}_t) \end{aligned}$$

Stochastic heterogeneous-agent blocks are analogous to deterministic blocks, except that all sequences are replaced by stochastic processes, and functions are defined over expected values of those processes. Note that $\tilde{\mathbf{D}}_{t+1}$ is predetermined as it only depends on variables that are measurable at time t. We also define stochastic simple blocks.

Definition 7. A *stochastic simple block* is a function between stochastic processes $\tilde{\mathbf{y}} = h(\tilde{\mathbf{x}})$ such that the time-t realization $\tilde{\mathbf{y}}_t$ is a time-invariant function $\tilde{h}$ of $\tilde{\mathbf{x}}$ in periods around t,

$$\tilde{\mathbf{y}}_t = \tilde{h}(\tilde{\mathbf{x}}_{t-k}, \ldots, \tilde{\mathbf{x}}_t, \mathbb{E}_t \tilde{\mathbf{x}}_{t+1}, \ldots, \mathbb{E}_t \tilde{\mathbf{x}}_{t+l}), \tag{117}$$

where $k, l \in \mathbb{N}$.

We are then ready to define a stochastic SHADE model, exactly as before.

Definition 8. A *stochastic SHADE model* is a defined like a regular SHADE model, except that blocks are either stochastic heterogeneous-agent blocks or stochastic simple blocks. A *steady state* is defined as before (in particular, it is deterministic).

Finally, we link stochastic SHADE models and deterministic SHADE models.

Definition 9. The *sequence-space-equivalent* SHADE model of a stochastic SHADE model is a SHADE model whose blocks have the same inputs and outputs; each of whose simple blocks is defined by

the same function $\tilde{h}$ (e.g. as in definition 7); each of whose stochastic blocks is defined by the same y, Λ, v (as in definition 6).

We are now ready to state and prove the main result of this section: to first order, equilibria of stochastic SHADE models and equilibria of deterministic SHADE models are equivalent.

Proposition 4. *Fix a (deterministic) steady state $\{\mathbf{x}^{*i}, i \in \mathcal{N}\}$ of a stochastic SHADE model. Let $\{\mathbf{x}^{*i} + \epsilon\tilde{\mathbf{y}}^i, i \in \mathcal{N}\}$ be an equilibrium of the stochastic SHADE model. Define*

$$\mathbf{y}_t^i = \mathbb{E}_s\tilde{\mathbf{y}}_{s+t}^i - \mathbb{E}_{s-1}\tilde{\mathbf{y}}_{s+t}^i \tag{118}$$

Then, $\{\mathbf{x}^{*i} + \epsilon\mathbf{y}^i, i \in \mathcal{N}\}$, satisfies the equilibrium conditions of the sequence-space-equivalent SHADE model to first order in ϵ, for any $s \geq 0$.

Proof. We prove the result that if

$$\mathbf{x}^{*i} + \epsilon\tilde{\mathbf{y}}_t^i = \tilde{h}(\mathbf{x}^* + \epsilon\tilde{\mathbf{y}}_{t-k}, \ldots, \mathbf{x}^* + \epsilon\tilde{\mathbf{y}}_t, \mathbf{x}^* + \epsilon\mathbb{E}_t\tilde{\mathbf{y}}_{t+1}, \ldots, \mathbf{x}^* + \epsilon\mathbb{E}_t\tilde{\mathbf{y}}_{t+l}) \tag{119}$$

then, to first order in ϵ it holds that

$$\mathbf{y}_t^i = \tilde{h}(\mathbf{x}^* + \epsilon\mathbf{y}_{t-k}, \ldots, \mathbf{x}^* + \epsilon\mathbf{y}_t, \mathbf{x}^* + \epsilon\mathbf{y}_{t+1}, \ldots, \mathbf{x}^* + \epsilon\mathbf{y}_{t+l}) \tag{120}$$

where $\mathbf{y}_t^i$ is defined as in (118). Denote partial derivatives of $\tilde{h}$ by $\tilde{\mathbf{h}}_d \equiv \frac{\partial\tilde{h}}{\partial\mathbf{x}_d}(\mathbf{x}^*, \ldots, \mathbf{x}^*)$ for $d = -k, \ldots, l$. A Taylor expansion of (119) in ϵ then reads

$$\tilde{\mathbf{y}}_t^i = \sum_{d=-k}^{l} \tilde{\mathbf{h}}_d\tilde{\mathbf{y}}_{t+d} + \mathcal{O}(\epsilon) \tag{121}$$

Evaluating (119) at time $t+s$ and applying the expectations operator $\mathbb{E}_s$ we find

$$\mathbb{E}_s\tilde{\mathbf{y}}_{t+s}^i = \sum_{d=-k}^{l} \tilde{\mathbf{h}}_d\mathbb{E}_s\tilde{\mathbf{y}}_{t+s+d} + \mathcal{O}(\epsilon) \tag{122}$$

Subtracting $\mathbb{E}_{s-1}$ applied to (121) from (122) we obtain

$$\mathbf{y}_t^i = \sum_{d=-k}^{l} \tilde{\mathbf{h}}_d\mathbf{y}_{t+d} + \mathcal{O}(\epsilon)$$

implying that (120) holds to first order in ϵ.

Applying this logic to any stochastic simple block, as well as to any stochastic heterogeneous-agent block, this proves the proposition. □

The equivalence result in proposition 4 implies that if $\tilde{\mathbf{y}}_t$ has $MA(\infty)$ representation $\tilde{\mathbf{y}}_t = \sum_{s=0}^{\infty} m_s\epsilon_{t-s}$, with $\{\epsilon_t\}$ iid standard normal, and is part of an equilibrium of a stochastic SHADE

model, then its MA coefficients

$$m_t \epsilon_s = \mathbb{E}_s \tilde{\mathbf{y}}^i_{s+t} - \mathbb{E}_{s-1} \tilde{\mathbf{y}}^i_{s+t}$$

are part of an equilibrium of the sequence-space-equivalent (deterministic) SHADE model.

C.4 Discretized SHADE models

We next consider discretized SHADE models. We define them in two steps, beginning with discretized heterogeneous-agent blocks.

Definition 10. A *discretized heterogeneous-agent block* maps inputs $\mathbf{X}^i$, for $i = 1, \ldots, n_x$, to outputs $\mathbf{Y}^o$, for $o = 1, \ldots, n_y$. Stacking the $\mathbf{X}^i$s into a vector $\mathbf{X}$, and the $\mathbf{Y}^o$s into a vector $\mathbf{Y}$, we represent this map with the function

$$\mathbf{Y} = h(\mathbf{X}) \tag{123}$$

The time-t output of this function takes the form (16), where $\mathbf{v}_{t+1}$ is determined in (14) and $\mathbf{D}_t$ in (15).

This leads us to the following definition of discretized SHADE models.

Definition 11. A *discretized SHADE model* is a SHADE model, except that its blocks are either simple blocks or discretized heterogeneous-agent blocks.

We also slightly generalize simple blocks.

Definition 12. An *extended simple block* maps input sequences $\mathbf{X}^i$, for $i = 1, \ldots, n_x$ to output sequences $\mathbf{Y}^o$ for $o = 1, \ldots, n_y$. Stacking the $\mathbf{X}^i_t$'s into a vector $\mathbf{X}_t$, and the $\mathbf{Y}^o_t$'s into a vector $\mathbf{Y}_t$, there must exist $k, l \in \mathbb{N} \cup \{\infty\}$ and a time-invariant function h such that $\mathbf{Y}_t$ is only a function of neighboring $\mathbf{X}_t$'s, that is,

$$\mathbf{Y}_t = h(\mathbf{X}_{t-k}, \ldots, \mathbf{X}_{t+l})$$

Crucially, an extended simple block can depend on infinitely many leads or lags.

C.5 Proof of lemma 1

We first prove an intermediate result.

Lemma 2. *The Jacobians $\frac{\partial h^i}{\partial \mathbf{x}^t}$ of a simple block are asymptotically time invariant.*

Proof. Denote by $\tilde{\mathbf{h}}_d \equiv \frac{\partial \tilde{h}}{\partial \mathbf{x}_d}(\mathbf{x}^*, \ldots, \mathbf{x}^*)$, $d = -k, \ldots, l$ the partial derivatives of the function $\tilde{h}$ of the simple block (as in proposition 4). Then,

$$\frac{\partial h_{s+d}}{\partial \mathbf{x}_s} = \tilde{\mathbf{h}}_d$$

which hence trivially converges as $s \to \infty$. Each $\tilde{\mathbf{h}}_d$ is a $n_y \times n_x$ block matrix. □

This means that the Jacobians of both heterogeneous-agent blocks and simple blocks are asymptotically time invariant. Since the entire model's Jacobians are built up from individual blocks' Jacobians as in (37), the former Jacobians must be asymptotically time invariant, too, in light of the following simple result.

Lemma 3. *If two block matrices A, B are asymptotically time invariant, with $A_{s+j,s} \to \mathcal{A}_j \in \mathbb{R}^{n_1 \times n_2}$ and $B_{s+j,s} \to \mathcal{B}_j \in \mathbb{R}^{n_3 \times n_4}$, then so are $A + B$ (if $n_1 = n_3$ and $n_2 = n_4$) and $A \cdot B$ (if $n_2 = n_3$).*

Proof. $A + B$ is obviously asymptotically time invariant. For the matrix product, we have

$$(A \cdot B)_{s+j,s} = \sum_{t=-s}^{\infty} A_{s+j,s+t} B_{s+t,s}$$

For each s, j, this is a well-defined infinite sum as $A_{s+j,s+t} \to 0$ and $B_{s+t,s} \to 0$ for $t \to \infty$, exponentially fast. Since $A_{s+j,s+t} \to \mathcal{A}_{j-t}$ and $B_{s+t,s} \to \mathcal{B}_t$ uniformly in t, it follows that

$$(A \cdot B)_{s+j,s} \to \sum_{t=-\infty}^{\infty} \mathcal{A}_{j-t} \mathcal{B}_t$$

Note that this means the asymptotic shape of $A \cdot B$ around the diagonal is determined by the convolution of A and B. □

C.6 Proof of proposition 3

Fix a discretized SHADE model. First, note that every heterogeneous-agent block can be split up into three extended simple blocks (definition 12): the first one computes $\mathbf{v}_t$ as in (14), then $\mathbf{D}_t$ as in (15), and finally $\mathbf{y}_t$ as in (16). We hence arrive at a version of the SHADE model in which all heterogeneous-agent blocks are replaced by three extended simple blocks.

Then, without loss, denote by $\mathcal{U} = \{1, \dots, U\}$ the set of unknowns and by $\mathcal{O} = \{U+1, \dots, U+O\}$ the set of outputs, such that the order of $\mathcal{O}$ corresponds to the order in which outputs can be computed without any cyclicalities. Finally, denote the set of targets by $\mathcal{T} = \{\tau_1, \dots, \tau_U\}$. We can write the first-order change in each output o in response to a shock as

$$dx_t^o = \sum_{i=1}^{o-1} \sum_{j=-\infty}^{t} a_j^{o,i} dx_{t-j}^i + dz_t^o \tag{124}$$

where dz_t^o is a linear combination of past and future exogenous shocks whose form is irrelevant for determinacy. $a_j^{o,i}$ are real numbers, for numbers $o \in \mathcal{O}$ and $i \in \mathcal{U} \cup \mathcal{O}$ with $i < o$ and $j \leq t$. They correspond to the partial derivatives of individual blocks. For instance, if o is determined by a simple block with function $\tilde{h}$, $a_j^{o,i} = \frac{\partial \tilde{h}^o}{\partial x_{-j}^i}(\mathbf{x}_{-k}, \dots, \mathbf{x}_l)$, with notation as in lemma 2.

Stacking (124), this yields

$$\sum_{j=-\infty}^{t} A_j d\overline{\mathbf{X}}_{t-j} = d\overline{\mathbf{Z}}_t \tag{125}$$

where

$$A_j \equiv \begin{bmatrix} a_j^{U+1,1} & \cdots & a_j^{U+1,U} & -1 & & & \\ a_j^{U+2,1} & \cdots & \cdots & a_j^{U+2,U+1} & -1 & & \\ \vdots & & & & \ddots & \ddots & \\ a_j^{U+O,1} & \cdots & & & \cdots & a_j^{U+O,U+O-1} & -1 \\ \delta_{j0}\delta_{\tau_1,1} & \cdots & & & & \cdots & \delta_{j0}\delta_{\tau_1,U+O} \\ \vdots & & & & & & \vdots \\ \delta_{j0}\delta_{\tau_U,1} & \cdots & & & & \cdots & \delta_{j0}\delta_{\tau_U,U+O} \end{bmatrix}$$

and

$$d\overline{\mathbf{X}}_t \equiv \begin{pmatrix} dx_t^1 \\ \vdots \\ \\ \\ \\ \vdots \\ dx_t^{U+O} \end{pmatrix}, \qquad d\overline{\mathbf{Z}}_t = \begin{pmatrix} dz_t^{U+1} \\ \vdots \\ dz_t^{U+O} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Here, δ_{mn} is the Kronecker delta, equal to 1 if $m = n$ and 0 otherwise. The top half of A_j consists of the $a_j^{o,i}$ in (124) as well as a -1 at the position of the output dx_t^o that is determined in each row. The bottom half of A_j consists of Kronecker deltas which in each row τ_k equal 1 in column $i = \tau_k$. Thus, (125) describes the first-order behavior of the model in response to an exogenous shock. We define $A(\lambda) \equiv \sum_{j=-\infty}^{\infty} A_j e^{ij\lambda}$ and $a^{o,i}(\lambda) \equiv \sum_{j=-\infty}^{\infty} a_j^{o,i}(\lambda) e^{ij\lambda}$, so that

$$A(\lambda) \equiv \begin{bmatrix} a^{U+1,1}(\lambda) & \cdots & a^{U+1,U}(\lambda) & -1 & & & \\ a^{U+2,1}(\lambda) & \cdots & \cdots & a^{U+2,U+1}(\lambda) & -1 & & \\ \vdots & & & & \ddots & \ddots & \\ a^{U+O,1}(\lambda) & \cdots & & & \cdots & a^{U+O,U+O-1}(\lambda) & -1 \\ \delta_{\tau_1,1} & \cdots & & & & \cdots & \delta_{\tau_1,U+O} \\ \vdots & & & & & & \vdots \\ \delta_{\tau_U,1} & \cdots & & & & \cdots & \delta_{\tau_U,U+O} \end{bmatrix} \tag{126}$$

We can apply the Onatski (2006) determinacy criterion to (125). It implies that the model is determinate iff the winding number of det $A(\lambda)$ is equal to zero. What is det $A(\lambda)$? The following lemma characterizes the determinant.

Lemma 4. *Let $U \in \mathbb{N}, O \in \mathbb{N}_0$ and let $\tau_1, \dots, \tau_U$ be U elements of $\{1, \dots, U+O\}$. Further, let*

$$A \equiv \begin{bmatrix} a^{U+1,1} & \dots & a^{U+1,U} & -1 & & & \\ a^{U+2,1} & \dots & \dots & a^{U+2,U+1} & -1 & & \\ \vdots & & & & \ddots & \ddots & \\ a^{U+O,1} & \dots & & & \dots & a^{U+O,U+O-1} & -1 \\ \delta_{\tau_1,1} & \dots & & & & \dots & \delta_{\tau_1,U+O} \\ \vdots & & & & & & \vdots \\ \delta_{\tau_U,1} & \dots & & & & \dots & \delta_{\tau_U,U+O} \end{bmatrix} \in \mathbb{R}^{(U+O)\times(U+O)} \tag{127}$$

Then,

$$\det A = (-1)^{O(U-1)} \det J$$

where $J = (J^{\tau_k,u})_{1\le k,u\le U,} \in \mathbb{R}^{U\times U}$ is defined iteratively as $J^{u,v} = \delta_{uv}$ for $u, v \in \{1, \dots, U\}$ and for all $o \in \{U+1, \dots, U+O\}$ as

$$J^{o,u} = \sum_{i=1}^{o-1} a^{o,i} J^{i,u} \tag{128}$$

To apply lemma 4 to $\det A(\lambda)$, define the matrix-valued function $J(\lambda) \equiv (J(\lambda)^{\tau_k,u})_{1\le k,u\le U} \in \mathbb{R}^{U\times U}, \lambda \in [0, 2\pi)$ recursively as $J(\lambda)^{u,v} = \delta_{uv}$ for $u, v \in \{1, \dots, U\}$

$$J(\lambda)^{o,u} = \sum_{i=1}^{o-1} a^{o,i}(\lambda) J^{i,u}(\lambda) \tag{129}$$

Then, lemma 4 implies that

$$\det A(\lambda) = (-1)^{O(U-1)} \det J(\lambda)$$

where the constant $(-1)^{O(U-1)}$ is irrelevant for the winding number. What is $J(\lambda)$? We conclude our proof by showing that $J(\lambda)$ is indeed equal to $\mathcal{H}(\lambda) = \sum_{j=-\infty}^{\infty} \mathcal{H}_j e^{ij\lambda}$.

To show this, we define the mapping $\mathcal{F}(M)$ on the linear space of asymptotically translation invariant matrices, such that each image $\mathcal{F}(M)$ is a function itself from $\lambda \in [0, 2\pi)$ to $\mathcal{F}(M)(\lambda) = \sum_{j=-\infty}^{\infty} \lim_{s\to\infty} M_{s+j,s} e^{ij\lambda}$. $\mathcal{F}$ preserves matrix multiplication, that is,

$$\mathcal{F}(M \cdot M') = \mathcal{F}(M) \cdot \mathcal{F}(M') \tag{130}$$

for two asymptotically time invariant matrices M, M'. Furthermore,

$$\mathcal{F}\left(\frac{\partial h^o}{\partial \mathbf{x}^i}\right)(\lambda) = a^{o,i}(\lambda) \tag{131}$$

This implies that, by iteratively applying (130) and (131) to (37),

$$\mathcal{F}(J^{o,u})(\lambda) = J(\lambda)^{o,u}$$

and therefore

$$J(\lambda) = (J(\lambda)^{\tau_k,u})_{1\le k,u\le U} = \sum_{j=-\infty}^{\infty} \left(\lim_{s\to\infty} J^{\tau_k,u}_{s+j,s}\right)_{1\le k,u\le U} e^{ij\lambda} = \mathcal{H}(\lambda)$$

This concludes our proof: the model is determinate if and only if the winding number of $\det A(\lambda) = \det \mathcal{H}(\lambda)$ is equal to zero.

Proof of Lemma 4. We prove the result by induction over O. For simplicity, we call O the "number of outputs" in A. Begin with $O = 0$. In that case,

$$A = \begin{bmatrix} \delta_{\tau_1,1} & \cdots & \delta_{\tau_1,U} \\ \vdots & & \vdots \\ \delta_{\tau_U,1} & \cdots & \delta_{\tau_U,U} \end{bmatrix} = J$$

and therefore $\det A = \det J$, for any $U \in \mathbb{N}$. Assume now the result holds for matrices of the form (127) with a given number of outputs $O-1$ (and for any $U \in \mathbb{N}$). We wish to prove it Distinguish two cases.

Case I: $U + O \notin \{\tau_1, \ldots, \tau_U\}$. In that case, applying Laplace's formula to the last column of A, we find

$$\det A = (-1)^{U-1} \det A_{-O,-(U+O)}$$

where $A_{-O,-(U+O)}$ is matrix A without the Oth row and $(U+O)$th column. Since the submatrix $A_{-O,-(U+O)}$ is of the same shape as in (127), just with $O-1$ outputs, we can apply the induction hypothesis to arrive at $\det A = (-1)^{O(U-1)} \det J$.

Case II: $U + O \in \{\tau_1, \ldots, \tau_U\}$. Without loss (by reordering the τ's) assume $\tau_U = U + O$. As before, let J be defined as in (128). In that case, applying Laplace's formula to the last row of A, and reordering the rows, we find

$$\det A = (-1)^{U-1} \det \begin{bmatrix} a^{U+1,1} & \ldots & a^{U+1,U} & -1 & & \\ \vdots & & & \ddots & \ddots & \\ a^{U+O-1,1} & \ldots & & \ldots & a^{U+O-1,U+O-2} & -1 \\ \delta_{\tau_1,1} & \cdots & & & \cdots & \delta_{\tau_1,U+O-1} \\ \vdots & & & & & \vdots \\ \delta_{\tau_{U-1},1} & \cdots & & & \cdots & \delta_{\tau_{U-1},U+O-1} \\ a^{U+O,1} & \ldots & & & \ldots & a^{U+O,U+O-1} \end{bmatrix}$$

Observe that, if the last row were different—with just zeros except for a single 1 in some column j (call this matrix $A^{(j)}$)—the induction hypothesis would apply so that the determinant is equal to $\det J^{(j)}$ where the last row in $J^{(j)}$ is given by $\sum_{i=1}^{O+U-1} \delta_{ij} J^{i,u}$ for $u = 1, \ldots, U$. Using the linearity of

the determinant in the last row, we find

$$\det A = (-1)^{U-1} \sum_{j=1}^{U+O-1} a^{U+O,j} \det A^{(j)} = \sum_{j=1}^{U+O-1} a^{U+O,j} \det J^{(j)}$$

where we note that $\det A^{(j)} = (-1)^{(O-1)(U-1)} \det J^{(j)}$ as $A^{(j)}$ has U unknowns and $O-1$ outputs and $(-1)^{U-1}(-1)^{(O-1)(U-1)} = (-1)^{O(U-1)}$. Again, by the linearity of the determinant in the last row

$$\sum_{j=1}^{U+O-1} a^{U+O,j} \det J^{(j)} = \det J$$

since

$$\sum_{j=1}^{U+O-1} a^{U+O,j} \sum_{i=1}^{O+U-1} \delta_{ij} J^{i,u} = J^{U+O,u}$$

This proves that $\det A = (-1)^{O(U-1)} \det J$ and therefore the induction step. □