

Saddle Cycles: Solving Rational Expectations Models Featuring Limit Cycles (or Chaos) Using Perturbation Methods

Dana Galizia^{*†}

January 2021

Final Version

Abstract

Unlike linear ones, non-linear business cycle models can generate sustained fluctuations even in the absence of shocks (e.g., via limit cycles/chaos). A popular approach to solving non-linear models is perturbation methods. I show that, as typically implemented, these methods are incapable of finding solutions featuring limit cycles or chaos. Fundamentally, solutions are only required not to explode, while standard perturbation algorithms seek solutions that meet the stronger requirement of convergence to the steady state. I propose a modification to standard algorithms that does not impose this overly strong requirement.

Key Words: Dynamic equilibrium economies; Computational methods; Non-linear solution methods; Limit cycles; Chaos; JEL Class.: C63; C68; E37

^{*}Department of Economics, Carleton University. Email: dana.galizia@gmail.com. Phone: (613) 520-2600 x7454.

[†]Thanks to Paul Beaudry, Josh Brault, Franck Portier, and anonymous referees for helpful comments. All errors are mine.

1 Introduction

Rational expectations models where non-linearities play an important role have become increasingly used in macroeconomics in recent years. Unlike their linear counterparts, non-linear models are in some cases capable of generating sustained fluctuations in economic aggregates even in the absence of stochastic forces, such as through limit cycle dynamics (i.e., endogenous fluctuations that repeat indefinitely in the absence of shocks).¹ A common criticism of many existing models of the business cycle is their perceived over-reliance on unobserved (and often empirically unjustified) exogenous stochastic forces in matching the persistence and volatility found in the data. As shown in Beaudry et al. (2016) and Beaudry et al. (2020),² models that feature limit cycles are capable of endogenously delivering that persistence and volatility, and need only be combined with relatively small stochastic forces in order to make fluctuations as unpredictable and irregular as those found in the data. This suggests that models featuring such “stochastic limit cycles” may be a promising avenue for business cycle research.

For a model featuring a limit cycle to be useful for quantitative applications, one must be able to (at least approximately) obtain a solution to it. There is a substantial body of literature devoted to the development of approximate solution methods for non-linear rational expectations models. These methods vary in complexity, computational burden, and accuracy. As several studies have noted (e.g., Gaspar & Judd (1997), Aruoba et al. (2006)), perturbation (i.e., Taylor approximation) methods are relatively simple, often produce a reasonable degree of accuracy, and are significantly faster than other popular alternatives, making them a good candidate for solving larger models and/or for use in estimation exercises. Perturbation methods would thus seem like an attractive choice for solving limit cycle models. However, standard perturbation algorithms³ are incapable of dealing with such models. This fact—

¹The solution method in this paper can also be applied without modification to non-repetitive endogenous fluctuations (i.e., chaos). However, the discussion will focus on limit cycles, which appear to be the more empirically plausible case.

²These papers were by no means the first to explore the relevance of limit cycles and/or chaos in macroeconomic fluctuations, and indeed this literature has a long history (see the references contained in Beaudry et al. (2020)). However, combining such forces with more conventional stochastic ones appears to have gone largely unexplored in that earlier literature.

³See, for example, Judd (1996), Judd & Guu (1997), Uhlig (1999), Schmitt-Grohé & Uribe (2004), Aruoba et al. (2006), Gomme & Klein (2011), Kollman et al. (2011), and Caldara et al. (2012). See Fernández-Villaverde et al. (2016) (section 4) for a recent comprehensive overview of these methods.

which, to my knowledge, this paper is the first to explicitly point out—has at least two important consequences. First, this potentially fruitful research avenue may have gone largely unexplored to date in part because of the unavailability of practicable solution methods. Second, there could in fact be existing models in the literature that can produce limit cycles for certain parameterizations, but for which (a) such parameterizations have been discarded because the researcher (or the software package used by the researcher) incorrectly believes that there is no solution to the model, or (b) in addition to solutions the researcher has found, there are additional solutions featuring limit cycles that have been inadvertently ignored.

The reason standard perturbation methods cannot solve models featuring attractive limit cycles—cycles to which the system would eventually converge in the absence of any shocks—is straightforward. A solution to a rational expectations model is typically defined as a function that maps current pre-determined variables into current jump variables/future pre-determined variables, and that (1) implicitly satisfies a particular difference equation (DE), and (2) results in trajectories that do not explode, i.e., that satisfy a transversality condition (TVC). Referring henceforth to a function that satisfies the DE (but not necessarily the TVC) as a “DE function”, a natural approach to finding such a solution is to search among DE functions—of which there is typically a finite multiplicity—for one that also satisfies the TVC.

As is well known,⁴ DE functions are generally indexed by their linear approximations about the steady state, and it is typically straightforward to obtain this indexing set directly from the DE using the implicit function theorem. Further, given a linear approximation to a DE function, it is straightforward to sequentially obtain the associated second-order Taylor approximation, then third-order, etc., up to any desired approximation order. Thus, when using a perturbation approach, selecting from among the set of DE functions boils down to selecting from among the set of linear approximations that index those functions.

The linear approximation selected by standard perturbation algorithms is the one that solves the *linearized model* (LM), i.e., the model obtained by linearizing the DE. Since the LM is by definition linear, its trajectories generically either converge to the steady state or explode.⁵ Thus, a trajectory satisfies the LM TVC if and only if it

⁴See, e.g., Judd (1998).

⁵When the endogenous part of the LM has eigenvalues lying exactly on the complex unit circle, there may also exist trajectories that neither converge nor explode. To keep the discussion simple, I ignore these cases since in practice they typically correspond to a measure-zero of the parameter

is convergent. However, if the solution to the original non-linear model features an attractive limit cycle then, by construction, its linear part will not be convergent, and therefore this linear part *cannot be a solution to the LM*. Put more succinctly, standard perturbation methods only look for convergent solutions, and therefore cannot find solutions involving limit cycles. In practice, when the unique solution involves a limit cycle, typical software packages (e.g., Dynare) would, for a single solution attempt, generally report an error indicating that the Blanchard-Kahn conditions are violated, while in an estimation exercise the associated parameterization would typically be discarded.

The present paper describes a relatively simple way to extend standard perturbation methods so as to be able to accommodate solutions that feature limit cycles. Essentially, rather than assuming that a DE function satisfies the TVC only if its linear approximation is a solution to the LM, the method described below proposes to actually check each DE function to see whether it satisfies the TVC. In practice, this involves obtaining higher-order perturbation approximations to each DE function, and then simulating numerically to check whether trajectories explode. Importantly, the method is a strict generalization of standard methods, and can therefore be applied without modification to models that do not feature limit cycles (and in many cases without a significant increase in computational burden). Thus, one need not determine in advance whether limit cycles are relevant in order to apply the method.

Other Methods

Perturbation methods are not the only possible way to solve models featuring limit cycles. Roughly speaking, solution methods for rational expectations models can be grouped into two classes. The first involves finding functions that (approximately) solve the DE. Within this class, there are two principal methods that have been used: local perturbation methods, as used in this paper, and the global projection methods formally introduced to economics by Judd (1992) and Gaspar & Judd (1997) (among others). Because of their global nature, projection methods are often capable of delivering a higher degree of accuracy than perturbation methods. However, even in the simplest models, solution times are typically on the order of tens of seconds or longer, and increase rapidly with the number of state variables (and the desired space. Nonetheless, the solution method described below will automatically deal with them.

degree of accuracy). As a result, estimation using projection methods is essentially infeasible in all but the very simplest of cases. In contrast, depending on the model and perturbation order used, solution times for perturbation methods are often on the order of fractions of a second. For example, in the application of Section 5, which features a model with four state variables (putting it at the high end among cases to which projection methods have been applied in the literature), a single solution using a third-order perturbation takes less than 0.01 seconds, while still delivering a reasonable degree of accuracy. Further, in addition to their heavier computational burden, projection methods are also typically more complicated to implement, and, in order to optimize computation times (a top priority with such methods), are frequently coded in a low-level programming language like Fortran or C++. In contrast, perturbation methods are relatively simple to implement, and are able to deliver fast solution times even when coded in a high-level language like MATLAB.

The second class of solution methods is based on solving for agents' value functions by iterating on the Bellman equations of the model. Similar to the projection methods discussed above, the principal drawback of these methods is their computational burden. This is especially acute in cases where the first welfare theorem fails to hold, since one must simultaneously solve not only for the private value and policy functions, but also for each function mapping the aggregate state into a relevant price (or other coordinating mechanism). Since limit cycles may be especially relevant when there are strategic complementarities,⁶ in which case the first welfare theorem would fail, this is precisely an application in which value function methods would be especially computationally expensive. Furthermore, complementarities may in fact cause the search problem to become non-convex, in which case numerical algorithms are not guaranteed to converge.

The remainder of the paper is organized as follows.⁷ Section 2 outlines the main elements of the solution algorithm as it applies to deterministic models. While modern business cycle analysis is conducted mainly using stochastic models, focusing first on the deterministic case allows the basic ideas underlying the method to be conveyed in the simplest possible setting. Section 3 then discusses how one can implement the

⁶See Beaudry et al. (2020).

⁷For considerations of length, it will generally be assumed that the reader has a basic familiarity with standard non-linear perturbation algorithms (both theory and implementation). See Fernández-Villaverde et al. (2016), section 4, for an overview of these standard methods.

method in practice. Section 4 extends the discussion to the more general stochastic case and discusses computation. Section 5 illustrates the method by applying it to a micro-founded model that features a stochastic limit cycle, and finally Section 6 concludes.

2 The Deterministic Case

To make exposition of the method as simple as possible, this Section addresses deterministic environments. Once results and intuition have been established for this case, it will be straightforward to extend them to the stochastic case in Section 4.

2.1 Set-Up

Suppose $x_t \in \mathbb{R}^n$ is a vector of endogenous variables at date t , and the economy evolves deterministically according to the potentially non-linear difference equation

$$\Gamma(x, x') = 0, \tag{1}$$

where x denotes current-period values, x' denotes values in the subsequent period, and $\Gamma: \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$ is a locally C^∞ function.⁸ We take Γ as primitive (and known to the researcher). Letting a function with a subscript denote its Jacobian with respect to the subscripted argument, we impose the following assumptions.

Assumption 1. $\Gamma(x, x) = 0$ has a unique solution, which we denote \bar{x} .

Assumption 2. For any x , there is a unique solution to (1) for x' , and $\Gamma_{x'}(\bar{x}, \bar{x})$ is invertible.

Assumption 1 ensures that the system has a unique steady state (SS).⁹ This SS will serve as the natural point around which to perturb the system. Assumption 2, meanwhile, requires both that, given x , a solution for x' exists (an unrestrictive assumption in most settings) and that this solution is both globally unique and unique to a first-order approximation. Uniqueness in turn requires as a first step that all

⁸In practice, we only require that Γ be C^k , where k is at least as big as the desired perturbation order.

⁹One can replace Assumption 1 with the less stringent assumption that there is only one *relevant* SS; that is, that there is only one SS of practical interest. This will allow the method to address, for example, models in which capital is a necessary input to production, which typically feature as an uninteresting SS the one associated with no capital.

“static” relationships be eliminated from the system. Assumption 2 can be easily relaxed to accommodate static relationships (e.g., as in King & Watson (1998), Klein (2000), or Sims (2002) for the case of linear models), though at the cost of significantly obscuring the intuition. We therefore maintain Assumption 2 for now, and relax it below in Section 4 when discussing the more general stochastic case.

Without loss of generality, henceforth make the normalization that the unique SS is given by $\bar{x} = 0$. Partition the state vector as $x = (y, z)$, where $y \in \mathbb{R}^{n_y}$ corresponds to the set of variables that are pre-determined at date t , $z \in \mathbb{R}^{n_z}$ corresponds to the set of variables free to change at date t (i.e., the jump variables), and $n_y + n_z = n$. A solution to the model takes the form of a pair of functions $\pi: \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_z}$ and $\phi: \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_z}$ giving, respectively, y' and z as functions of the current pre-determined variables y . Specifically, we have the following.

Definition 1. A *model solution* (π, ϕ) is a pair of functions such that, for some open subset Θ containing zero (the SS), we have

$$\Gamma(y, \phi(y), \pi(y), \phi(\pi(y))) = 0, \quad \forall y \in \Theta, \quad (2)$$

and for each $y \in \Theta$,

$$\limsup_{t \rightarrow \infty} \|x_t\| < \infty, \quad (3)$$

where $x_t = (y_t, z_t)$, $y_0 = y$, $y_t = \pi(y_{t-1})$, and $z_t = \phi(y_t)$.

In this definition, $\|\cdot\|$ denotes the Euclidean norm. Note that, since we will employ Taylor approximations, we are inherently restricted to considering only local dynamics, hence the focus on Θ . We also assume (as is always implicitly the case for perturbation methods) that Γ is such that, if a unique model solution exists, then it is analytic on Θ . As we will only be interested in actually obtaining a solution when it is unique,¹⁰ this assumption will allow us to focus our search within the set of analytic functions.

Also of interest will be the linearized system. Let

$$A \equiv -\Gamma_{x'}^{-1} \Gamma_x,$$

where a Jacobian without an argument gives its value evaluated at the SS. The linearized system is thus given by

$$x' = Ax.$$

¹⁰Issues related to indeterminacy are addressed in Section 3.

Let Λ denote the set of eigenvalues of A . For an eigenvalue $\lambda \in \Lambda$, let $e(\lambda)$ denote the corresponding generalized eigenspace, and let

$$r(\lambda) \equiv \text{Re}[e(\lambda)] \oplus \text{Im}[e(\lambda)] ,$$

where $\text{Re}[a]$ and $\text{Im}[a]$ denote the real and imaginary parts of a , respectively, and \oplus denotes the direct sum.¹¹ We will refer to $r(\lambda)$ as the real generalized eigenspace (RGE) associated with λ . Note that $r(\lambda) = r(\bar{\lambda})$ (where $\bar{\lambda}$ denotes the complex conjugate), $r(\lambda)$ is an A -invariant¹² linear subspace, and if λ is real then the dimension of $r(\lambda)$ is the same as $e(\lambda)$, while if it is complex the dimension is twice that of $e(\lambda)$.

Finally, let \mathcal{W} denote the set of all A -invariant linear subspaces of \mathbb{R}^n (excluding the trivial A -invariant subspace $\{0\}$), and let \mathcal{S} denote the set of analytic function pairs (π, ϕ) satisfying (2) (regardless of whether they also satisfy (3)).

2.2 Some Useful Properties

For ease of exposition, in the remainder of Section 2, we assume that all eigenvalues of A have algebraic multiplicity one. While this assumption is usually not restrictive (cases with repeated eigenvalues typically represent a measure zero of the parameter space), we will nonetheless relax it in Section 3 below when discussing how to implement the method in practice. Note that this assumption implies that RGEs corresponding to real eigenvalues always have dimension one, and those corresponding to complex eigenvalues always have dimension two. It also implies that $w \in \mathcal{W}$ if and only if it is either an RGE itself, or the direct sum of two or more RGEs.

Since they are analytic by assumption, the functions $(\pi, \phi) \in \mathcal{S}$ can be equivalently and uniquely represented by their sets of partial derivatives (of every order) evaluated at the SS. The first-order SS derivatives (π_y, ϕ_y) clearly must satisfy the linear approximation of (2), i.e.,

$$\Gamma_y + \Gamma_z \phi_y + \Gamma_{y'} \pi_y + \Gamma_{z'} \phi_y \pi_y = 0 \quad \forall y . \tag{4}$$

As is well known, while there are typically multiple solutions (π_y, ϕ_y) to (4), given any one such solution, using (2) we can sequentially and (typically) uniquely obtain all

¹¹The direct sum of two sets X and Y is given by $X \oplus Y \equiv \{x + y : x \in X, y \in Y\}$.

¹²A set M is invariant with respect to a function F if $x \in M$ implies $F(x) \in M$. By A -invariant we mean invariant with respect to the function $x \mapsto Ax$.

higher-order SS derivatives of π and ϕ .¹³ In this sense, the elements of \mathcal{S} are indexed by the solutions of (4).

Given a pair of functions $(\pi, \phi) \in \mathcal{S}$, let

$$M(\phi) \equiv \{(y, z) \in \mathbb{R}^n : y \in \Theta, z = \phi(y)\}$$

be the n_y -dimensional surface in \mathbb{R}^n mapped out by $z = \phi(y)$, $y \in \Theta$, and

$$W(\phi) \equiv \{(y, z) \in \mathbb{R}^n : y \in \mathbb{R}^{n_y}, z = \phi_y y\}$$

be the n_y -dimensional linear subspace tangent to $M(\phi)$ at $x = 0$. The following proposition establishes that $W(\phi)$ is A -invariant.¹⁴

Proposition 1. *If $(\pi, \phi) \in \mathcal{S}$, then $W(\phi) \in \mathcal{W}$.*

All proofs are in Appendix B.

The next proposition establishes a similar result in the reverse direction: if w is an A -invariant linear subspace that is mapped out by $z = \psi y$, $y \in \mathbb{R}^{n_y}$, for some matrix ψ , then there is a pair $(\pi, \phi) \in \mathcal{S}$ with $M(\phi)$ tangent to w at $x = 0$ (or, equivalently, there is a pair (π_y, ϕ_y) satisfying (4) with $\phi_y = \psi$).

Proposition 2. *Define \mathcal{W}^* as follows: $w \in \mathcal{W}^*$ if and only if (i) $w \in \mathcal{W}$, and (ii) there is a matrix ψ such that we can write*

$$w = \{(y, z) \in \mathbb{R}^n : y \in \mathbb{R}^{n_y}, z = \psi y\} . \tag{5}$$

Then for any $w \in \mathcal{W}^$, there exists a pair $(\pi, \phi) \in \mathcal{S}$ such that $W(\phi) = w$. Further, the eigenvalues of π_y are given by the eigenvalues associated with the RGEs that make up w .*

Together, Propositions 1 and 2 establish an equivalence between first-order approximations of functions satisfying (2), and A -invariant linear subspaces. The last

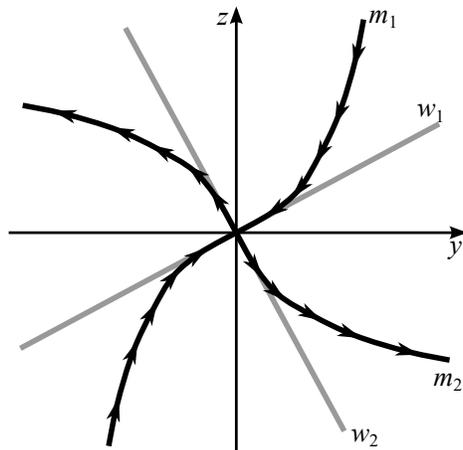
¹³To obtain SS derivatives of order $k \geq 2$ given SS derivatives up to and including order $k - 1$, first obtain all k -th order derivatives of (2), and then evaluate the results at the SS. The resulting system of equations will be linear in the k -th order SS derivatives (the only unknowns), and a unique solution for them will typically exist. See Fernández-Villaverde et al. (2016) (section 4), and the example of Section 2.5 below.

¹⁴The properties highlighted by Propositions 1 and 2 in this section are not novel. However, I am not aware of any existing explicit statement of these results in the literature. Given that they form the basis for the proposed solution method, it is worth explicitly stating (and formally proving) them here.

part of Proposition 2 further establishes that the dynamics associated with any $(\pi, \phi) \in \mathcal{S}$ —which are driven by the “projected” system $y' = \pi(y)$ —are, to first order, determined to by the eigenvalues associated with $W(\phi)$.

The content of these Propositions is illustrated by example in Figure 1, which shows a phase diagram for a simple case with $n_y = n_z = 1$ of the type often encountered in macro. In particular, A is taken to have two distinct real positive eigenvalues, with one stable and one unstable. The first element of all eigenvectors is taken to be non-zero, and thus there are two w 's in \mathcal{W}^* , given by the two (one-dimensional) RGEs. These are shown as the gray lines w_1 and w_2 in the Figure, where w_1 is associated with the stable eigenvalue and w_2 the unstable one. From Proposition 2, associated with each w_j is a pair of functions $(\pi, \phi) \in \mathcal{S}$ for which $M(\phi)$ is tangent to w_j . These latter surfaces are shown as the thick black curves m_1 and m_2 in the Figure. Since w_1 is associated with the stable eigenvalue, the dynamics along m_1 converge (at least locally) to the SS. Conversely, since w_2 is associated with the unstable eigenvalue, the dynamics along m_2 diverge from the SS. These convergence/divergence properties are indicated by the arrows placed along m_1 and m_2 .

Figure 1: Illustrating Propositions 1 and 2



Notes: Figure is for the case where A has two distinct real positive eigenvalues, one stable and one unstable. The w_j 's are the elements of \mathcal{W}^ , and the m_j 's are the associated elements of \mathcal{S} .*

2.3 The Solution Method

The results in Section 2.2 suggest a way to find (analytic) functions (π, ϕ) satisfying (2). First, find an element $w \in \mathcal{W}^*$ using any of the well established numerical methods for doing so. Then, for such a w , one can obtain the associated (π_y, ϕ_y) as suggested by Proposition 2, and then use it to obtain the sequence of higher-order derivatives of π and ϕ in the usual way.

While there are generally multiple elements of \mathcal{S} , typically only a subset (in many cases of interest, only one) of them will also satisfy (3) and therefore be model solutions as per Definition 1. To determine which of the elements of \mathcal{S} actually constitute solutions, standard perturbation algorithms (e.g., those described in Fernández-Villaverde et al. (2016)) implicitly make the following additional assumption.¹⁵

Assumption 3. Any trajectory that does not converge to the SS violates (3).

Under Assumption 3—which the method proposed in this paper does *not* make—in order for a pair $(\pi, \phi) \in \mathcal{S}$ to satisfy (3), the dynamics implied by the first-order approximation to π (i.e., $y' = \pi_y y$) must involve convergence to the SS. That is, the eigenvalues of π_y must all be stable. In contexts where it is justified, Assumption 3 is useful since, from the last part of Proposition 2, we need only consider elements $w \in \mathcal{W}^*$ whose constituent RGEs are each associated with stable eigenvalues, and if there is a unique such w , it is straightforward to find it using any of several known algorithms.¹⁶

Unfortunately, as noted in the Introduction, Assumption 3 is not universally valid. In particular, consider a $(\pi, \phi) \in \mathcal{S}$ for which the system $y' = \pi(y)$ features a globally attractive limit cycle. Since all trajectories generated by this system necessarily satisfy (3), (π, ϕ) is in fact a model solution. However, because its first-order approximation is necessarily locally unstable—i.e., π_y will have unstable eigenvalues—naïvely imposing Assumption 3 would cause this solution to be inappropriately discarded.

Figure 2 presents an example of such a situation for a case with two pre-determined variables ($n_y = 2$) and one jump variable ($n_z = 1$). Each panel of the Figure shows the same phase diagram from a slightly different angle. In this example, A has a pair of complex unstable eigenvalues, with the third eigenvalue real, positive, and

¹⁵Recall that, to simplify the discussion, we have abstracted from the existence of unit-modulus eigenvalues. It is nonetheless straightforward to adapt the following discussion to the general case.

¹⁶See, e.g., Blanchard & Kahn (1980), King & Watson (1998), Klein (2000), Sims (2002).

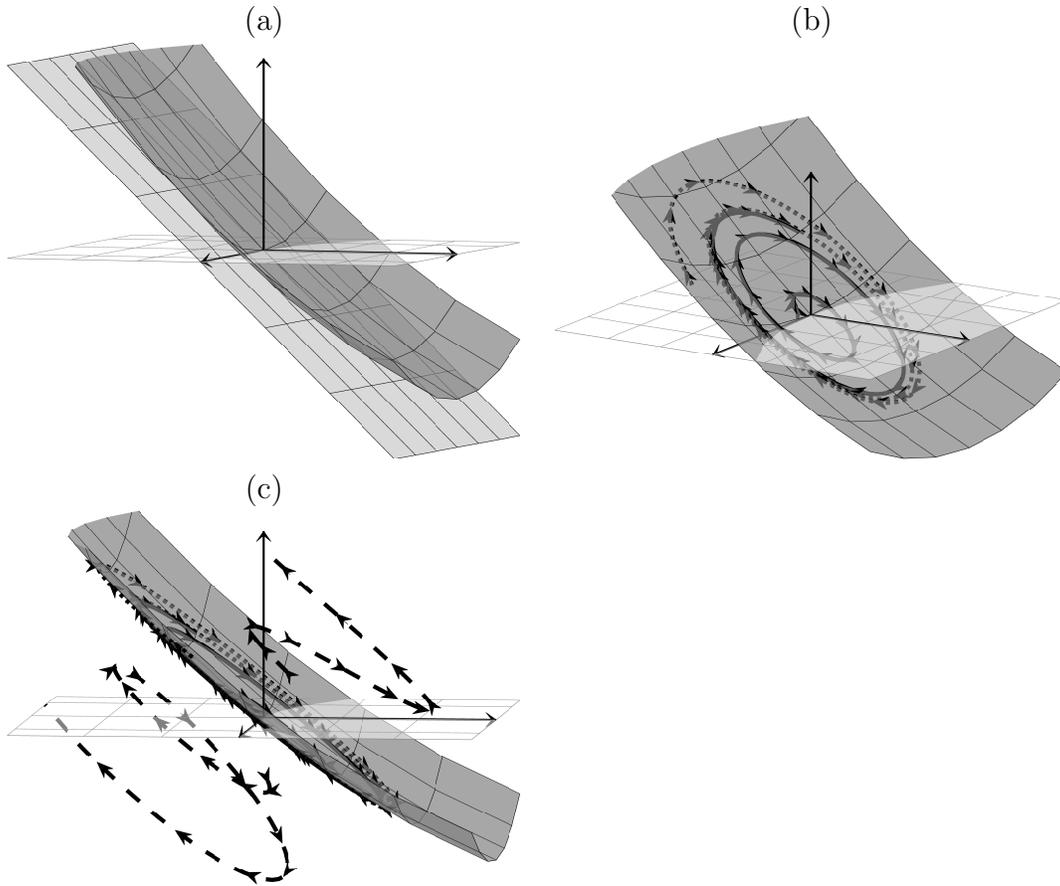
unstable. \mathcal{W}^* has one element w , which is the two-dimensional RGE associated with the pair of complex eigenvalues. This is shown as the light gray plane in panel (a) of the Figure. Associated with w is a $(\pi, \phi) \in \mathcal{S}$. $M(\phi)$, which is tangent to w at zero, is shown as the dark gray non-linear surface in each panel. The system $y' = \pi(y)$ features a globally attractive limit cycle: except for those beginning exactly at the SS (zero), all trajectories on $M(\phi)$ converge to the limit cycle over time (and therefore satisfy (3)). Examples of such trajectories are given by the solid and dotted curves in panel (b) of the Figure. Trajectories beginning at any point in \mathbb{R}^n other than on $M(\phi)$ diverge from the SS and become unbounded, violating (3). Examples of such trajectories are shown as the dashed curves in panel (c) of the Figure. We see, then, that there is a unique model solution (π, ϕ) . However, since its local dynamics are driven by the pair of complex unstable eigenvalues, standard perturbation algorithms would be unable to discover it. Instead, since all eigenvalues of A are unstable, they would incorrectly conclude that there are no model solutions.

The method proposed in this paper is, conceptually, a straightforward modification of standard perturbation algorithms that is designed to address the possibility of such a scenario. In particular, rather than relying on Assumption 3 to automatically select solutions (π, ϕ) from \mathcal{S} exclusively on the basis of the eigenvalues of π_y , we will instead individually check (in practice, using numerical simulations), for each $(\pi, \phi) \in \mathcal{S}$, whether the system $y' = \pi(y)$ satisfies (3).

2.4 Applicability

As with all perturbation methods, the method developed here suffers from the limitation that, for any given approximation order, the quality of the approximation generally deteriorates as one moves further away from the expansion point. When the model solution admits a limit cycle, the system tends to spend much of its time at points near (or exactly on) the cycle, which is, by definition, away from the SS. This raises an important question: In such cases, is the quality of a Taylor approximation necessarily too poor to be useful? The answer to this is, no, not necessarily, for the same reason that it is not necessarily inappropriate to apply perturbation methods to locally stable *stochastic* models. For a desired error tolerance, there is generally some maximal neighborhood Υ (in y -space) around the expansion point on which the approximation error is always within the specified tolerance. In the case of a locally stable stochastic model, as long as shocks are not large enough to push the system

Figure 2: A Saddle Limit Cycle



Notes: Illustration of a unique solution (shown as the dark gray surface in each panel) featuring a globally attractive limit cycle. Tangent to it is the associated element of \mathcal{W}^ (light gray plane in panel (a)). Panel (b) shows trajectories on the solution surface, which always converge to the limit cycle. Panel (c) shows trajectories off the solution surface, which always diverge.*

outside of Υ , then on the range in which the system spends most of its time, the approximation error is within the specified tolerance. Similarly, for a model featuring a limit cycle, as long as the cycle is located within Υ , the approximation error will be within the specified tolerance, and thus a perturbation method would also be acceptable.

Of course, since the true solution is typically unknown, so is Υ (though there are diagnostic tools available to evaluate the accuracy of any approximation method). Generally speaking, the stronger are any expected non-linearities in the true solution functions (e.g., if the solution functions are discontinuous or kinked), the poorer would be the quality of the approximation. As is always the case, judgment must be exercised in ascertaining whether any expected increase in accuracy from using global methods (such as projection methods) is large enough to offset their more difficult implementation and slower computation times. Importantly, though, there is nothing intrinsic to limit cycles that necessarily balances the scales in favor of global methods.

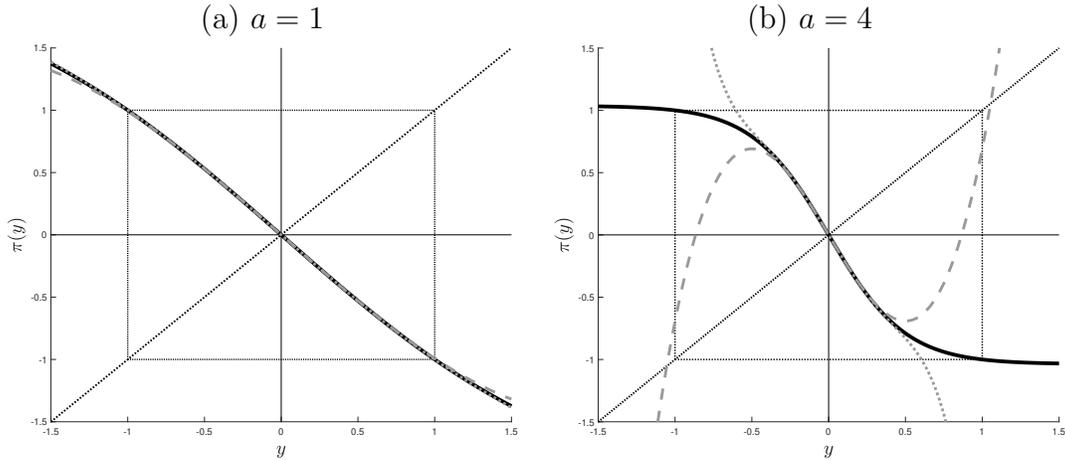
To illustrate, consider the case where $n_y = 1$, and suppose the true solution for π is given by the decreasing sigmoid function

$$\pi(y) = b(a) \frac{1 - e^{ay}}{1 + e^{ay}}, \quad (6)$$

where $a > 0$ is a parameter that controls the degree of non-linearity, and $b(a) \equiv -(1 + e^a)/(1 - e^a)$ is a normalization factor that ensures there is always a 2-period cycle at $y = \pm 1$. Figure 3 illustrates the case for two different values of a . In each panel, the solid black curve plots $\pi(y)$, the gray dashed curve a third-order Taylor approximation to π , and the gray dotted curve a fifth-order approximation. Also plotted as black dotted lines are the 45° line as well as a box that highlights the 2-cycle at ± 1 . In panel (a), we show the case where $a = 1$, for which π exhibits a mild degree of non-linearity. In this case, over a range extending well beyond the cycle, both Taylor approximations lie almost directly on top of the true π , suggesting that even a third-order approximation may be sufficiently accurate in practice. In panel (b), on the other hand, which shows the case of $a = 4$, the true π is much more non-linear, and both Taylor approximations are unacceptably poor when the system is out near the 2-cycle.¹⁷

¹⁷Indeed, it can be verified that the Taylor series about zero does not converge for $y = \pm 1$.

Figure 3: Limit Cycle Taylor Approximations



Notes: Each panel plots $\pi(y)$ from equation (6) for a different value of a as the black curve, and the third- and fifth-order Taylor approximations as the gray dashed and dotted curves, respectively.

2.5 A Worked Example Featuring a Limit Cycle

This section presents a simple model for which a rational-expectations solution exists and can be found by the proposed method, but not by standard methods. Consider the simple forward-looking model

$$\begin{aligned} K' &= (1 - \delta)K + I, \\ G(I) &= I' - \alpha K, \end{aligned} \tag{7}$$

where K (“capital”) is pre-determined at date t , I (“investment”) is free to jump, $\delta \in (0, 1)$ and $\alpha > 0$ are parameters, and G is a strictly increasing function satisfying $G(0) = 0$ and $G_3 > 0$, where a function with a subscript i denotes the i -th derivative at the SS. To rule out the possibility of multiple SSs we assume further that $\alpha > \delta$. Finally, to keep things as simple as possible, we also assume that $G_2 = 0$.

We can write the evolution of this economy as

$$\begin{pmatrix} K' \\ I' \end{pmatrix} = \begin{pmatrix} (1 - \delta)K + I \\ \alpha K + G(I) \end{pmatrix}.$$

We then have that

$$A = \begin{pmatrix} 1 - \delta & 1 \\ \alpha & G_1 \end{pmatrix},$$

which has eigenvalues

$$\lambda_1 = \frac{1 - \delta + G_1 - \sqrt{(1 - \delta - G_1)^2 + 4\alpha}}{2},$$

$$\lambda_2 = \frac{1 - \delta + G_1 + \sqrt{(1 - \delta - G_1)^2 + 4\alpha}}{2}.$$

Suppose $G_1 < \alpha + \delta - 2$. It can then be verified that $\lambda_1 < -1$ and $\lambda_2 > 1$. Since the eigenvalues are both real, there are two elements of \mathcal{W}^* , given by the two one-dimensional RGEs of A . There are thus also two elements of \mathcal{S} . Let $(\pi^{(j)}, \phi^{(j)})$ denote the element of \mathcal{S} associated with λ_j , $j = 1, 2$, and note that $\pi_1^{(j)} = \lambda_j$. Since $|\pi_1^{(j)}| > 1$ for $j = 1, 2$, the linear approximation to $\pi^{(j)}$ is always unstable. Thus, if one were to apply standard perturbations methods (i.e., impose Assumption 3), one would conclude that there are no solutions. As we show now, this conclusion is erroneous here: there exists (at least to a third-order approximation) a unique solution to the non-linear model, which features a limit cycle.

Solutions $(\pi, \phi) \in \mathcal{S}$ by definition satisfy (2), which for this example is given by

$$\begin{pmatrix} \pi(K) \\ \phi(\pi(K)) \end{pmatrix} = \begin{pmatrix} (1 - \delta)K + \phi(K) \\ \alpha K + G(\phi(K)) \end{pmatrix}.$$

Note for future reference that the first of these equations implies that $\pi_1 = (1 - \delta) + \phi_1$, and $\pi_i = \phi_i$ for $i \geq 2$. Substituting the first equation into the second for $\pi(K)$ yields

$$\phi((1 - \delta)K + \phi(K)) = \alpha K + G(\phi(K)). \quad (8)$$

Since condition (8) must hold for all K , sequentially taking derivatives three times of this equation with respect to K , and then evaluating each at $K = 0$, we may obtain the conditions

$$\phi_1^2 + (1 - \delta - G_1)\phi_1 - \alpha = 0, \quad (9)$$

$$[(1 - \delta + \phi_1)^2 + \phi_1 - G_1]\phi_2 = 0, \quad (10)$$

$$[(1 - \delta + \phi_1)^3 + \phi_1 - G_1]\phi_3 = G_3\phi_1^3 - 3(1 - \delta + \phi_1)\phi_2^2. \quad (11)$$

Equation (9) is a quadratic in ϕ_1 , which has two solutions. Each of these solutions indexes a different $(\pi, \phi) \in \mathcal{S}$. It can be verified that these solutions are given by $\phi_1^{(j)} = \lambda_j - (1 - \delta)$, $j = 1, 2$. Substituting this into equation (10) for ϕ_1 , we find that $\phi_2^{(j)}$ satisfies

$$[\lambda_j^2 + \lambda_j - (1 - \delta + G_1)]\phi_2^{(j)} = 0.$$

Except in knife-edge circumstances, the term in square brackets will be non-zero, and thus $\phi_2^{(j)} = 0$.¹⁸ Substituting this and the expression for $\phi_1^{(j)}$ into (11) yields

$$\phi_3^{(j)} = \frac{[\lambda_j - (1 - \delta)]^3}{\lambda_j^3 - \lambda_{-j}} G_3,$$

where the notation λ_{-j} denotes the eigenvalue that is not λ_j . Since $\lambda_1 < -1$ and $\lambda_2 > 1$, the numerator and denominator of the fraction on the right-hand side of this expression are of the same sign (negative when $j = 1$, positive when $j = 2$), and thus, since $G_3 > 0$, we also have $\phi_3^{(j)} > 0$. It can then be verified that, to a third-order approximation, the evolution of K is given by

$$K' = \lambda_j K + \frac{1}{6} \phi_3^{(j)} K^3. \quad (12)$$

Since $\lambda_2 > 1$ and $\phi_3^{(2)} > 0$, for $j = 2$ this system is globally unstable: for any initial value of K (except $K = 0$), the system will explode, violating (3). This is illustrated in panel (b) of Figure 4. The solid black curve plots the third-order approximation to $\pi^{(2)}$, while the gray dotted lines illustrate a typical path for K . Thus, $(\pi^{(2)}, \phi^{(2)})$ cannot be a solution.

Consider instead the $j = 1$ case. Since $\lambda_1 < -1$, system (12) will not converge to the SS. However, since $\phi_3^{(1)} > 0$, as long as the initial value of $|K|$ is not too large,¹⁹ the implied trajectory will remain bounded, and therefore satisfy (3). These dynamics will generically feature limit cycles or chaos, so that the system will typically neither explode nor converge to a single point. This is illustrated in panel (a) of Figure 4, where the solid black curve plots the third-order approximation to $\pi^{(1)}$, and the gray dotted lines illustrate a typical path for K . In this particular case, the model features a limit cycle: beginning from a point near the SS, $|K|$ grows each period, with K alternating signs. However, as the system moves away from the SS, the growth in $|K|$ eventually tapers off, and the system settles into a 2-cycle given by $K_t = (-1)^t \bar{K}$ for some $\bar{K} > 0$.

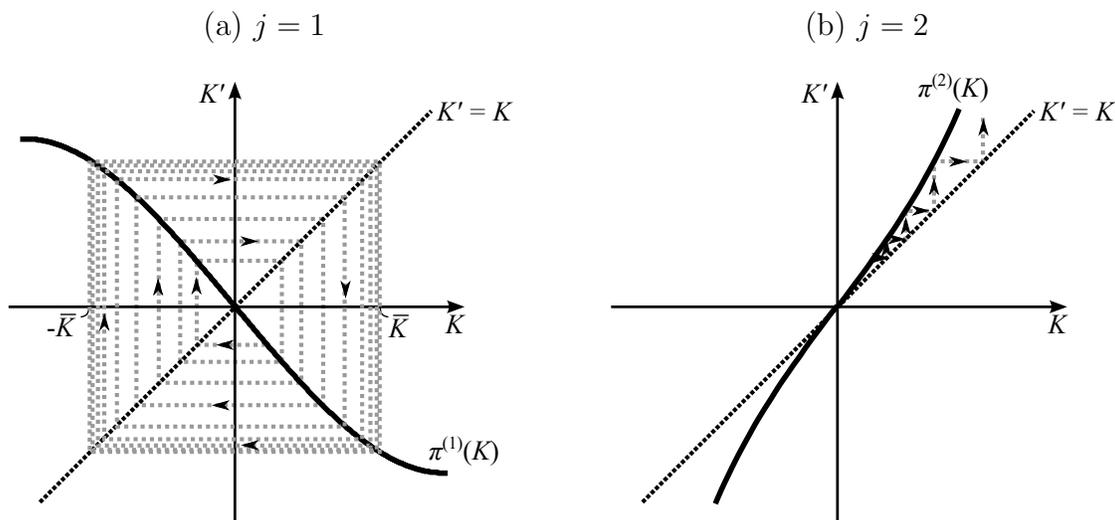
Since the system explodes for $j = 2$, but remains bounded for $j = 1$, we have verified that, despite the conclusions one would draw by maintaining Assumption 3,

¹⁸This property is inherited directly from our earlier simplifying assumption that $G_2 = 0$.

¹⁹For an initial value of $|K|$ large enough, the accuracy of the third-order approximation to $\pi^{(1)}$ will have degraded to the point that the approximated system no longer captures even the qualitative dynamics of the true system $K' = \pi^{(1)}(K)$. This is clear from the fact that, unlike the true system given by (7), the approximated system (12) possesses two unstable non-zero SSs $\pm K^*$ for some $K^* > 0$.

a model solution does in fact exist (and is unique), and is given by $(\pi^{(1)}, \phi^{(1)})$, with the resulting dynamics of the system governed (to a third-order approximation) by (12) for $j = 1$.

Figure 4: Evolution of K , Third-Order Approximation to $\pi^{(j)}$



Notes: Figure shows, for the deterministic example, the evolution of K according to the third-order approximation to (12) for each $j = 1, 2$. For $j = 2$, K becomes unbounded, while for $j = 1$ it approaches a 2-cycle at $\pm \bar{K}$.

3 Implementation

This section discusses how the solution method may be implemented in practice, including how to apply it to a more general environment than described above, how one can speed up the computation in practice, and how to deal with issues related to indeterminacy.

3.1 Generalizing the Problem

Before proceeding with implementation details, we first generalize the problem. First, we will no longer require that all eigenvalues of A have algebraic multiplicities equal to one.

Second, up to now we have largely ignored the issue of indeterminacy. In particular, we have only sought solution functions (π, ϕ) that map y into y' and z . This is overly restrictive, however. In reality, the only fundamental restriction placed on solution functions is that, since they are pre-determined, elements of the current y may not be outputs of any of these functions. However, it is in principle allowed for one or more elements of the current z to be *inputs* of the functions determining y' and the remaining values of z . The structure of Definition 1 does not allow for this, so we now generalize it.

In particular, given a set Q of indices chosen from the set $\{1, \dots, n_z\}$, let u be the subvector of z corresponding to these indices, let $\hat{y} \equiv (y, u)$, let \hat{z} be the remaining jump variables, and if necessary re-order the elements of x so that $x = (\hat{y}, \hat{z})$. The discussion in Section 2—and in particular Definition 1—corresponds to the special case where Q is empty. We thus generalize Definition 1 as follows.

Definition 2. A *model solution* is a quadruple (Q, π, ϕ, χ) that satisfies Definition 1 after replacing y and z with, respectively, \hat{y} and \hat{z} , and, for a given initial $y = y_0$, setting the initial value of \hat{y} to $\hat{y}_0 = (y_0, u_0)$, where $u_0 = \chi(y_0)$.

We will henceforth use Definition 2. Note that if we can find a model solution with Q non-empty and where there is more than one possible choice for the function χ pinning down the initial value of u , then there typically exists a multiplicity of solutions; that is, we would have indeterminacy.

As should be clear, for any given Q , after making the same replacements detailed in Definition 2, all results and analysis of Section 2 go through without any further modification. In particular, each Q is associated with a set \mathcal{W}_Q^* of A -invariant linear subspaces from Proposition 2, and a set \mathcal{S}_Q of pairs of analytic functions (π, ϕ) satisfying (2), each of which is associated with a $W(\phi) \in \mathcal{W}_Q^*$. Thus, finding analytic model solutions for any Q can be done exactly as detailed above for the special case where Q is empty.

3.2 Finding Elements of \mathcal{S}

In this subsection we discuss how to find elements of $\mathcal{S} = \mathcal{S}_\emptyset$, i.e., \mathcal{S}_Q for the special case where Q is empty. As noted above, it is straightforward to extend this procedure to any Q . The algorithm uses real Schur decompositions of the matrix A to generate

elements of \mathcal{W}_\emptyset^* . Such a decomposition is given by $A = UTU^\top$,²⁰ where B^\top indicates the transpose of B , U is a real orthogonal matrix (and in particular, $UU^\top = I_n$), and T is a real block-upper-triangular matrix with the following properties: (a) the diagonal blocks are all either 1×1 or 2×2 ; (b) the real eigenvalues of A appear in the 1×1 blocks; and (c) the eigenvalues of each 2×2 block are a complex pair, and equal to such a pair from A . Note that this decomposition is not unique.

Given such a Schur decomposition of A , consider the first $q > 0$ columns of U , denoted U_1 , and the upper-left $q \times q$ block of T , denoted T_{11} (where we require only that this partition does not split one of the 2×2 diagonal blocks of T described in property (c) above). As is well known, the columns of U_1 form a basis for a q -dimensional A -invariant linear subspace; that is, the column space w of U_1 is an element of \mathcal{W} . Further, the convergence properties of the system $x' = Ax$ restricted to w are governed by the eigenvalues of T_{11} . Finally, for any element $w \in \mathcal{W}$ with dimension q , we can always find a Schur decomposition of A such that the column space of U_1 is precisely w . Thus, in principle we can generate any element of \mathcal{W} through an appropriate choice (and partition) of the Schur decomposition.

While the particular ordering of a Schur decomposition (which determines which element of \mathcal{W} is spanned by the columns of U_1) is unpredictable for standard numerical implementations, a given Schur decomposition can be “reordered” so as to obtain an alternative Schur decomposition for which the first q columns of U are a basis for a different element of \mathcal{W} .²¹ It would be useful if, by considering every possible reordering of a given Schur decomposition, we could so generate every possible element of \mathcal{W} . Unfortunately, if we wish to allow for the possibility of eigenvalues with algebraic multiplicities greater than one (as we now do), this will not generally be the case. In particular, when A has at least one eigenvalue with a *geometric* multiplicity greater than one, \mathcal{W} will have an uncountable number of elements. For example, for two linearly independent eigenvectors associated with the same real eigenvalue, every linear combination of these eigenvectors is also an eigenvector, and the space w spanned by that new eigenvector is an element of \mathcal{W} . We can clearly generate an

²⁰See, for example, Golub & Van Loan (1996), chapter 7. Numerical routines to compute a real Schur decomposition are widely available. In MATLAB and Julia it can be done using the function `schur` which, for a real matrix, computes the real Schur decomposition by default. For implementation in Fortran or C++, various routines are available, including the LAPACK routine `dgees`.

²¹Re-ordering of the real Schur decomposition can be done using the function `ordschur` in MATLAB and Julia, or the LAPACK routine `dtrsen` in Fortran or C++.

uncountably infinite number of w 's in this way. Since there are only a finite number of possible reorderings of the Schur decomposition, it is clearly not possible to find every element of \mathcal{W} in this way, and indeed it will not be feasible in practice to find every one of the infinite number of elements of \mathcal{W} using *any* algorithm. By necessity, we thus make the following additional assumption.

Assumption 4. Suppose there exists a unique model solution (π, ϕ) .²² Then for any eigenvalue λ , if $W(\phi)$ and $r(\lambda)$ have any non-zero element in common, then $r(\lambda) \subseteq W(\phi)$.

Assumption 4 states that, aside from the zero vector (which is always an element of both $W(\phi)$ and $r(\lambda)$), $W(\phi)$ cannot contain only part of a given RGE: either the whole RGE is contained in $W(\phi)$, or none of it is.²³

Let \mathcal{R} be the set of RGE's of A . Under Assumption 4, $W(\phi)$ is the direct sum of a subset of \mathcal{R} whose dimensions sum to n_y . Since there are at most n RGE's (fewer if A has any complex or repeated eigenvalues), it is feasible to compute direct sums of each possible combination of the elements of \mathcal{R} whose dimensions sum to n_y (though in general, as we show in Section 3.3, it will not be necessary to do so).

This suggests a simple way to find a solution if it exists. First, obtain any Schur decomposition $A = UTU^\top$. Next, let $r(\lambda_1), \dots, r(\lambda_k)$ be k distinct elements of \mathcal{R} whose dimensions sum to n_y . Re-order the Schur decomposition so that the upper-left $n_y \times n_y$ block of T has eigenvalues $\lambda_1, \dots, \lambda_k$ (with the same algebraic multiplicities as they appear in A) along with the corresponding conjugates of any complex λ_j 's. The first n_y columns of U are then a basis for $w \equiv r(\lambda_1) \oplus \dots \oplus r(\lambda_k) \in \mathcal{W}_\emptyset^*$. By Proposition 2,²⁴ associated with this w is a $(\pi, \phi) \in \mathcal{S}_\emptyset$ such that $W(\phi) = w$. Partitioning U and T conformably with $x = (y, z)$ as

$$U = \begin{pmatrix} U_{yy} & U_{yz} \\ U_{zy} & U_{zz} \end{pmatrix}, \quad T = \begin{pmatrix} T_{yy} & T_{yz} \\ 0 & T_{zz} \end{pmatrix},$$

as noted in the proof of Proposition 2, the first-order Taylor coefficients for these

²²Recall that uniqueness implies that the only model solution has Q empty.

²³Assumption 3—which standard algorithms use, but the present method does not—implies Assumption 4. Thus, we have essentially replaced Assumption 3 with the much weaker Assumption 4.

²⁴It can be verified that Propositions 1-2 do not rely on our earlier simplifying assumption that there are no repeated eigenvalues, and thus also apply here where we have relaxed this assumption.

functions are given by

$$\begin{aligned}\pi_y &= U_{yy} T_{yy} U_{yy}^{-1}, \\ \phi_y &= -U_{zz}^{\top -1} U_{yz}^{\top}.\end{aligned}$$

These coefficients can then be used in the usual way to sequentially obtain higher-order coefficients up to the desired order. One can then simulate to check whether the resulting approximate system satisfies (3). If it does, we have found a model solution. We can then repeat this procedure for each different combination of RGEs whose dimensions sum to n_y in order to find every model solution for Q empty, and then repeat the procedure for each possible Q (of which there are a finite number).

3.3 Narrowing Down The Search

The algorithm above involves checking every possible combination of RGEs whose dimensions sum to at least n_y to see whether the approximated system satisfies (3). However, we can often avoid checking many (even most) of these combinations. In many applications, one is only interested in the actual model solution if it is unique, so if at any point a second model solution is found one can stop searching for more. We can also use theory to rule out many of the combinations, as established in the following proposition.

Proposition 3. *Let λ be an eigenvalue of A .*

- (i) *Suppose λ is real with $\lambda > 1$. Then for any Q and any $(\pi, \phi) \in \mathcal{S}_Q$ for which $r(\lambda) \subset W(\phi)$, (π, ϕ) violates (3).*
- (ii) *Suppose (a) there exists a unique model solution (π, ϕ) , (b) $|\lambda| < 1$, and (c) we can write $r(\lambda) = \{Vy : y \in \mathbb{R}^{n_y}\}$ for some $n \times n_y$ matrix V . Then $r(\lambda) \subset W(\phi)$.*

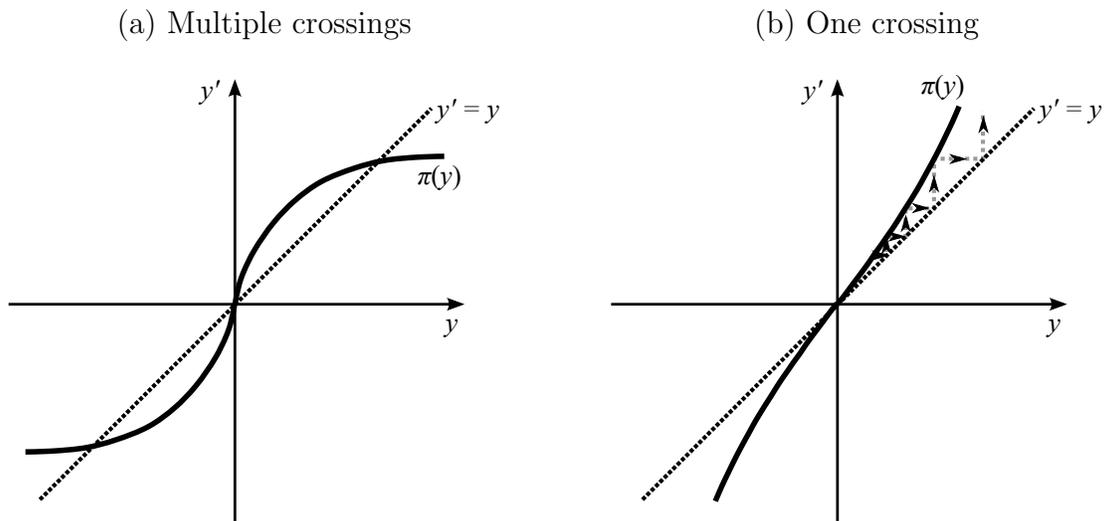
Part (i) of Proposition 3 implies that we need only consider Schur re-orderings for which T_{yy} has no real, positive, unstable eigenvalues. Part (ii) implies that, if we are willing to assume that a model solution is unique whenever it exists, then we also need only consider re-orderings for which every stable eigenvalue appears in T_{yy} .²⁵

The intuition for Proposition 3 is easiest to see in the simple case where $n_y = n_z = 1$ and A has two distinct real eigenvalues. To see part (i), suppose by hypothesis that

²⁵Note that (c) in part (ii) restricts attention to RGEs that can be written as a function of the pre-determined variables. This condition is likely to be met for all RGEs in most practical applications (and can be easily checked).

$\lambda > 1$, and $(\pi, \phi) \in \mathcal{S}_\emptyset$ is such that $r(\lambda) \subset W(\phi)$ (and thus $r(\lambda) = W(\phi)$). Consider the evolution of y for this solution, given by $y' = \pi(y)$. This system has an SS at $y = 0$, and the dynamics near that SS are given to a first-order approximation by $y' = \lambda y$. Since $\lambda > 1$, there are two possible configurations for the function π , displayed separately in panels (a) and (b) of Figure 5. In the first configuration, illustrated in panel (a), π crosses the 45-degree line at one or more non-zero points (in addition to the crossing at zero). Each additional crossing, however, would represent an additional SS of $y' = \lambda y$, and therefore an additional SS of (1), a possibility that is ruled out by Assumption 1. Thus, this configuration for π cannot occur under the maintained assumptions. The remaining possible configuration, wherein π crosses the 45-degree line at only one point (zero), is illustrated in panel (b). In this case, as long as the initial $y \neq 0$, $|y|$ will grow over time without bound (as illustrated by the gray line in the figure). Since y becomes unbounded, so does x , and thus (3) becomes violated, which in turn implies that (π, ϕ) cannot be a model solution.

Figure 5: Configurations For $\pi(y)$ When $\lambda > 1$



Notes: Figure illustrates the content of part (i) of Proposition 3 for a system with $n_y = n_z = 1$ and for which A has two distinct real eigenvalues. The dynamics of a solution (π, ϕ) associated with an eigenvalue $\lambda > 1$ are determined by the mapping $y' = \pi(y)$, which may either cross the 45-degree line multiple times (panel (a)), or one time (panel (b)).

To see part (ii) of Proposition 3, suppose instead that $|\lambda| < 1$, and there exists a unique model solution (π, ϕ) . Let (π^*, ϕ^*) be the element of \mathcal{S}_\emptyset associated with λ ,

whose dynamics are driven by $y' = \pi^*(y)$. For y sufficiently close to zero, the system is dominated by its linear part, $\pi_y^* = \lambda$. Since $|\lambda| < 1$, these trajectories converge to the SS, and thus (π^*, ϕ^*) is a model solution. Since part (ii) supposes that any model solution is unique, $\phi^* = \phi$, and since $r(\lambda) = W(\phi^*)$, it follows immediately that $r(\lambda) \subset W(\phi)$.

Proposition 3 can significantly reduce the number of Schur re-orderings that need to be checked. For example, in a modestly sized system with $n_y = n_z = 5$ and n distinct real eigenvalues, there are $10\text{-choose-}5 = 252$ possible re-orderings to check. However, if, say, 3 of the eigenvalues of this system are stable, and a further 3 are real and greater than one, then using Proposition 3 one could narrow down the set of re-orderings to check to $4\text{-choose-}2 = 6$. For larger systems, the potential computational savings are even greater.

3.4 Indeterminacy

Thus far we have largely proceeded under the assumption that if a solution exists then it is unique. In general, of course, this may not be the case. In this section we discuss several ways to test for indeterminacy.²⁶ It is worth emphasizing, however, that as with any solution method, without imposing more structure on the problem we generally cannot in practice ever entirely rule out the possibility of indeterminacy.

Under Assumption 3, there is a simple diagnostic test for indeterminacy (originally stated in Blanchard & Kahn (1980)) that most macroeconomists are familiar with: the model features indeterminacy if and only if the number of stable eigenvalues of A , which we denote by n_s , exceeds the number of pre-determined variables, n_y . When replacing Assumption 3 with the less stringent Assumption 4 (as we have), this result carries over, but in a weaker form. In particular, we have the following indeterminacy test (IT):

Indeterminacy Test 1. If $n_s > n_y$, then the model features indeterminacy.

Notice that $n_s > n_y$ is a sufficient condition for indeterminacy here, but is no longer a necessary one as it was under Assumption 3. Sufficiency of $n_s > n_y$ for

²⁶If it turns out that a particular model features indeterminacy, the usefulness of the present paper comes to an end. In particular, it does not contain any results about how one might characterize the set of indeterminate solution paths in such cases (as, for example, was done in Lubik & Schorfheide (2003) for the case of linear models). While such results could be useful, developing them is beyond the scope of this paper.

indeterminacy can be seen directly from Proposition 3(ii), which implies that, if there is a unique solution (π, ϕ) , then all “stable” RGEs (i.e., those associated with stable eigenvalues) must be contained in $W(\phi)$. Since $W(\phi)$ necessarily has dimension n_y (since $(\pi, \phi) \in \mathcal{S}$ in the case of a unique solution), if $W(\phi)$ is to contain all the stable RGEs then the sum of the dimensions of those stable RGEs—which equals n_s —cannot exceed n_y . Thus, IT 1 gives a simple way to test for indeterminacy that, in most applications, requires a negligible amount of computing time regardless of the size of the system. Of course, since $n_s \leq n_y$ does not necessarily rule out indeterminacy, further investigation may be merited.

A second test for indeterminacy is given as follows:

Indeterminacy Test 2. If there are multiple elements of $\mathcal{S} = \mathcal{S}_\emptyset$ satisfying (3), then the model features indeterminacy.

IT 2 follows obviously from the fact that, by definition, if (π, ϕ) is both an element of \mathcal{S} and satisfies (3), then it is a solution, and therefore if there are multiple elements of \mathcal{S} satisfying (3), we have multiple solutions. Thus, IT 2 suggests that another way to test for indeterminacy is simply to check all elements of \mathcal{S} for conformity with (3). In modestly sized systems, performing this check will not be very costly computationally, especially if one first uses Proposition 3(i) to rule out elements of \mathcal{S} wherever possible. Furthermore, the extra computer code required to implement IT 2 is negligible, making it quite simple to apply this test in practice.

ITs 1 and 2 both provide sufficient conditions for establishing the presence of indeterminacy, but in neither case are those conditions necessary. In order to conclusively test for indeterminacy, one must instead use the following test:

Indeterminacy Test 3. If there are multiple elements of $\mathcal{S}_{\text{all}} \equiv \bigcup_Q \mathcal{S}_Q$ satisfying (3), then the model features indeterminacy.

IT 3 involves testing, for every Q , each pair of functions $(\pi, \phi) \in \mathcal{S}_Q$ (see Definition 2) for conformity with (3). The set \mathcal{S}_{all} containing all of these pairs of functions is in turn indexed by the set $\mathcal{W}_{\text{all}}^*$ made up of direct sums of RGEs whose dimensions sum to *at least* n_y (rather than *exactly* n_y , as is the case with \mathcal{W}_\emptyset^*). Depending on the size and particular structure of the system, $\mathcal{W}_{\text{all}}^*$ could have a very large number of elements. Some of these may be eliminated from consideration by applying Proposition 3(i). The following proposition gives a way to further—and potentially greatly—reduce

the number of elements that one must check for conformity with (3) when applying IT 3.

Proposition 4. *Suppose $w \in \mathcal{W}_{\text{all}}^*$ is associated with some model solution (Q, π, ϕ, χ) , and that we can find $w_1, w_2 \in \mathcal{W}_{\text{all}}^*$ which are both distinct proper subsets of w .²⁷ Then associated with each w_j is a model solution $(Q_j, \pi^j, \phi^j, \chi^j)$, with $(Q_1, \pi^1, \phi^1) \neq (Q_2, \pi^2, \phi^2)$.*

Proposition 4 says that, if $w \in \mathcal{W}_{\text{all}}^*$ is associated with a model solution and contains two or more distinct elements of $\mathcal{W}_{\text{all}}^*$ as proper subsets, then there are at least two model solutions of a lower “dimension”. This implies that we do not need to check any elements of $\mathcal{W}_{\text{all}}^*$ that contain at least two such proper subsets.²⁸

As an example, suppose the eigenvalues of A are all real and distinct, so that the RGEs all have dimension one. If $w \in \mathcal{W}_{\text{all}}^*$ has dimension $b > n_y$, it can be written as the direct sum of b RGEs. We can then generate an n_y -dimensional proper subset of w by choosing n_y of those b RGEs and taking their direct sum. We can generate multiple distinct proper subsets of w in this way, each of which is also an element of $\mathcal{W}_{\text{all}}^*$. Proposition 4 implies that if w is associated with a model solution, then so is each of those subsets. It is therefore sufficient to look only for n_y -dimensional model solutions in this example, since if we cannot find more than one of those then there cannot be any model solutions of a higher dimension.

While Propositions 3(i) and 4 can greatly reduce the computational burden associated with IT 3, the burden may still be significant, especially in larger systems. Further, these tools cannot address the other main drawback of applying IT 3, which is the added complexity of its implementation. For example, one must write code to identify the subset of $\mathcal{W}_{\text{all}}^*$ that does not have at least two other elements as proper subsets. Also, since the elements of this subset will in general have different dimensions, the code to obtain the derivatives of (2) must be adapted for different partitions $x = (\hat{y}, \hat{z})$ (and possibly re-orderings as well). While these are by no means insurmountable difficulties, they are not trivial either.

²⁷i.e., $w_j \subset w$, $\exists x \in w$ with $x \notin w_j$, and $w_1 \neq w_2$.

²⁸Note that if $w \in \mathcal{W}_{\text{all}}^*$ contains exactly one proper subset $w_1 \in \mathcal{W}_{\text{all}}^*$, then it is not generally sufficient to check only the (π, ϕ) associated with w_1 .

4 Extension to the Stochastic Case

Extending the solution method to the stochastic case is straightforward. Letting ϵ be an n_ϵ -vector of mean-zero i.i.d. innovations, assume that the evolution of the n_θ -vector of exogenous stochastic variables θ evolves according to

$$\theta' = \Pi(\theta, \zeta\epsilon'),$$

for some function $\Pi: \mathbb{R}^{n_\theta+n_\epsilon} \rightarrow \mathbb{R}^{n_\theta}$. Here, $\zeta \in \mathbb{R}_+$ is a “perturbation parameter”, which will be equal to one for the desired solution (see Fernández-Villaverde et al. (2016)).

We assume that the n model equations governing the endogenous variables, expressed in deviations from the non-stochastic SS (i.e., the SS when $\zeta = 0$), can be written

$$\mathbb{E}[\Gamma(x, x', \theta, \theta')] = 0, \quad (13)$$

where \mathbb{E} denotes expectation conditional on current information. To keep the presentation simple, assume that the endogenous state y is known one period in advance, and in particular y' is not affected by the realized value of ϵ' .²⁹ In addition to extending the model to the stochastic case, we also now relax Assumption 2 to allow for the inclusion of “static” relationships in (13), and in particular we will no longer require that $\Gamma_{x'}$ be invertible.

A solution is defined analogously to Definition 1³⁰ as follows.

Definition 3. A *model solution* (π, ϕ) is a pair of functions such that, for some open subset Θ containing zero, we have

$$\mathbb{E} \left[\Gamma \left((y, \phi(y, \theta, \zeta)), (\pi(y, \theta, \zeta), \phi(\pi(y, \theta, \zeta), \Pi(\theta, \zeta\epsilon'), \zeta)), \theta, \Pi(\theta, \zeta\epsilon') \right) \right] = 0, \quad \forall (y, \theta, \zeta) \in \Theta; \quad (14)$$

and for each $(y, \theta, \zeta) \in \Theta$,

$$\limsup_{t \rightarrow \infty} \mathbb{E}[\|x_t\|] < \infty, \quad (15)$$

where $x_t = (y_t, z_t)$, $(y_0, \theta_0, \zeta) = (y, \theta, \zeta)$, $\theta_t = \Pi(\theta_{t-1}, \zeta\epsilon_t)$, $y_t = \pi(y_{t-1}, \theta_{t-1}, \zeta)$, and $z_t = \phi(y_t, \theta_t, \zeta)$.

²⁹One can typically recast the system in this way if need be. In any case, it is tedious but not difficult to adapt the discussion to the more general case.

³⁰It is straightforward to adapt Definition 2 to the stochastic case as well.

We will take our perturbation approximations around the non-stochastic SS. Since ζ equals one, not zero, for our desired solution, ζ itself will be one of the variables in the Taylor expansion. For example, the first-order approximation to π will in general be $\pi_\theta \cdot \theta + \pi_y \cdot y + \pi_\zeta \cdot \zeta$, and in practice we will always feed $\zeta = 1$ into this approximation.

Of interest will be the linearized system, which we can write as

$$\begin{aligned}\Gamma_x x + \Gamma_{x'} \mathbb{E}[x'] + \Gamma_\theta \theta + \Gamma_{\theta'} \mathbb{E}[\theta'] &= 0, \\ \mathbb{E}[\theta'] &= \Pi_\theta \theta,\end{aligned}$$

or, combining, and letting $A \equiv -\Gamma_x$, $B \equiv \Gamma_{x'}$, and $C \equiv -(\Gamma_\theta + \Gamma_{\theta'} \Pi_\theta)$,

$$B \mathbb{E}[x'] = Ax + C\theta. \tag{16}$$

From this point, the solution method is a straightforward extension of the non-stochastic case. In particular, the set of functions (π, ϕ) satisfying (14) is denoted by \mathcal{S} , and as before this set is indexed by a set \mathcal{W}^* of invariant linear subspaces w of (16) that can be written as in (5) for some $n_z \times n_y$ matrix ψ , with $\phi_y = \psi$ the first-order approximation to the associated solution, and the eigenvalues of π_y are those associated with the RGEs making up w . The key differences are that, since we now allow for the possibility that $\Gamma_{\theta'} = B$ is singular, (i) the eigenvalues in question are *generalized* eigenvalues of (A, B) ,³¹ (ii) an invariant linear subspace of (16) is a linear subspace w such that for any $x \in w$ there is a unique $x' \in w$ satisfying $Bx' = Ax$, and (iii) the RGEs making up $w \in \mathcal{W}^*$ are spanned by a set of generalized eigenvectors from the generalized eigenvalue problem.

In practice, the process for finding elements of \mathcal{W}^* is now based on the *generalized* real Schur—or, more compactly, the (real) QZ—decomposition,³² i.e., $A = RTU^\top$ and $B = RSU^\top$, which has the following properties: (i) R and U are each real orthogonal matrices, (ii) T is a block-upper-triangular matrix, and (iii) S is an upper-triangular matrix of the same rank as B with exactly $n_\infty \equiv n - \text{rank}(B)$ zeros on the main diagonal.³³ Further, for $q > 0$, if the upper-left $q \times q$ block S_{11} of S is non-singular,

³¹ λ is a generalized eigenvalue of (A, B) with associated eigenvector v if $Av = \lambda Bv$.

³²See Klein (2000) and Golub & Van Loan (1996) (chapter 7).

³³In general, a QZ decomposition can have an S with more than n_∞ zeros on the main diagonal in the rare case where zero is a defective eigenvalue of B with an associated eigenvector v that is also an eigenvector of A . In our application, however, assuming $\Gamma_{\theta'}$ is rank-deficient due only to the inclusion of static relationships in (13) (and in particular, the mathematical relationships

and if the upper-left $q \times q$ block T_{11} of T does not split a diagonal block of T , then (i) the first q columns of U form a basis for a q -dimensional invariant subspace w , and (ii) the eigenvalues of the associated RGEs are given by the eigenvalues of $S_{11}^{-1}T_{11}$. Just like with an ordinary real Schur decomposition, real QZ decompositions can be easily re-ordered so as to generate the invariant subspace associated with any desired collection of (generalized) eigenvalues.³⁴

4.1 Computation

Details of how to compute the first-order approximation associated with a given re-ordering of the QZ decomposition are presented in Appendix A. MATLAB code is also provided to conduct this step.³⁵ Given such a first-order approximation, one can then compute a higher-order approximation in the usual fashion. The general process for doing this is well known (see, e.g., Fernández-Villaverde et al. (2016)), and typically involves using software capable of analytically obtaining derivatives (such as Mathematica or the symbolic toolbox in MATLAB) to sequentially differentiate the appropriate model equations and then evaluate them at the SS. From these, one can then create functions that take as inputs the quantitative solutions for all k -th-order and lower SS derivatives and output the SS derivatives of order $k + 1$.³⁶

5 Application

This section applies the methodology to a version of the micro-founded New Keynesian model of Beaudry et al. (2020). In addition to some minor changes to the model itself, the application here differs from Beaudry et al. (2020) in two key ways. First, Beaudry et al. (2020) solve a restricted version of their model, whereas a general solution is obtained here for the full unrestricted model. Second, and more importantly,

fundamentally allow for n_∞ of the current jump variables to always be solved for uniquely in terms of the remaining current endogenous and exogenous variables), it can be verified that this will never happen.

³⁴The real QZ decomposition can be computed in MATLAB and Julia using the function `qz` with the `'real'` flag, and in Fortran or C++ using the LAPACK routine `dhgeqz`. Re-ordering can be done in MATLAB and Julia via `ordqz`, and in Fortran or C++ using the LAPACK routine `dtgsen`.

³⁵See the MATLAB function `InvSubGen.m`, which is available for download at <http://www.danagalizia.com/research>.

³⁶In the following section, we implement the proposed solution method for a particular example. The MATLAB code to do this is also provided, including code to obtain and use the relevant higher-order derivatives for that example.

Beaudry et al. (2020) use a third-order approximation for only one of the model’s equations (the risk premium equation), while the remaining equations are only linearly approximated. In contrast, here we non-linearly approximate all of the model’s equations.

The key features of the model are (i) equilibrium unemployment, (ii) limited enforcement of debt contracts, and (iii) sticky prices. The combination of these features causes households to face an endogenous risk premium on their borrowing that increases with the default rate, which is in turn increasing in the unemployment rate. Because prices are sticky, output is partially demand-determined, creating an equilibrium feedback mechanism: when unemployment increases, so does the risk premium, which raises the interest rate faced by households. This lowers household demand for consumption goods, causing output to fall, which lowers firms’ demand for labor, thus increasing the unemployment rate even further. This complementarity causes the steady state of the model to be unstable, and a limit cycle to appear.

We briefly present the main components of the model here. Further details can be found in Beaudry et al. (2020). The model features a continuum of households of mass one, each of which is made up of a continuum of mass one of workers. The head of the household maximizes

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \xi_{t-1} [U(C_t - \gamma C_{t-1}^*) + \nu(1 - e_t)] ,$$

where $\beta, \gamma \in (0, 1)$ are parameters, \mathbb{E}_t denotes expectation at date t , C is consumption services enjoyed by the household, e is the fraction of household members employed, U and ν are strictly increasing, concave utility functions, C^* denotes average consumption of all households (so that γC^* is an external habit term), and ξ is an exogenous discount factor shock.

Labor is assumed to be indivisible, with each employed worker supplying one unit of labor. Each morning, the household head dictates a reservation wage to the worker members, and instructs them to accept any job that meets that reservation wage. Workers then depart the household and proceed first to the consumption market, where they place orders for consumption to be received later that day. These orders must be paid for in advance, which requires the workers to borrow from the bank at endogenous nominal interest rate r_t , with all debts to be repaid at the beginning of the next period. Importantly, consumption orders must be placed and paid for—and therefore debts incurred—before individual workers know whether they will actually

be employed. If a particular worker ends up unemployed, they will be unable to pay back their debts. In that case, with some probability $\phi < 1$ the bank can recover the owed amount from the worker's household by paying a cost $\Phi_t < 1$ per unit of the loan, while with probability $1 - \phi$ there is no recourse for the bank, who must absorb the full amount of the default. As a result of this, the total expected loan repayment is $[e + (1 - e)\phi](1 + r)$ per unit of the loan. One can show that this leads to a consumption Euler equation for the household of the form

$$\mu_t U'(C_t - \gamma C_{t-1}) = \beta(1 + r_t) [e_t + (1 - e_t)\phi] \mathbb{E}_t \left[\frac{U'(C_{t+1} - \gamma C_t)}{1 + \pi_{t+1}} \right],$$

where π_{t+1} is the inflation rate and $\mu_t \equiv \xi_{t-1}/\xi_t$.

In order to make loans and finance any loan recovery costs, banks borrow at the risk-free nominal rate i_t . The total expected amount they must repay at the beginning of the next period is therefore $[1 + (1 - e)\phi\Phi](1 + i)$. Since banks must earn zero expected profits in equilibrium, this implies

$$1 + r_t = \frac{1 + (1 - e_t)\phi\Phi_t}{e_t + (1 - e_t)\phi} (1 + i_t).$$

The fraction on the right-hand side of this expression, which is greater than one whenever $e_t < 1$ (i.e., when there is unemployment), is an endogenous risk premium that compensates banks for the possibility of default.

Consumption services are produced additively using an existing stock of durable goods X_t and newly produced consumption goods $z_t F(e_t)$, where z_t is exogenous productivity and F is a strictly increasing, concave production function. Thus, $C_t = X_t + z_t F(e_t)$. A fraction ψ of newly produced goods are assumed to be durable, with a fraction δ of existing durables depreciating each period, so that

$$X_{t+1} = (1 - \delta) X_t + \psi z_t F(e_t).$$

Finally, i_t is set according to a Taylor rule.

In practice, we use power production $F(e) = e^\alpha$ and CRRA utility $U(x) = (x^{1-\omega} - 1)/(1-\omega)$. We assume $\Phi_t = \Phi(e_t)$, where $\Phi(\cdot)$ is a function satisfying $\Phi'(\bar{e}) = 0$, where \bar{e} is the steady state value of e . In particular, we adopt the flexible log-polynomial specification $\Phi(e) = \bar{\Phi} \exp\{\bar{\Phi}_2[100(e - \bar{e})]^2 + \bar{\Phi}_3[100(e - \bar{e})]^3\}$. As in Beaudry et al. (2020), we use a Taylor rule of the form

$$1 + i_t = \Theta \mathbb{E}_t \left[e_{t+1}^{\varphi_e} \frac{U'(C_{t+1} - \gamma C_t)}{\mathbb{E}_t \left[\frac{U'(C_{t+1} - \gamma C_t)}{1 + \pi_{t+1}} \right]} \right] \approx \Theta E_t [e_{t+1}^{\varphi_e} (1 + \pi_{t+1})],$$

where $\Theta, \varphi_e > 0$ are parameters that control the steady state nominal interest rate and elasticity of the interest rate with respect to employment, respectively. While this Taylor rule is somewhat non-standard, it simplifies the model by allowing us to ignore firms' price-setting behavior. Letting $Q(e) \equiv [1 + (1 - e)\phi\Phi(e)]$ and $Y_{t+1} \equiv z_t e_t^\alpha$, we can combine the above to obtain the equilibrium system

$$\begin{aligned}\mu_t \lambda_t &= Q(e_t) \mathbb{E}_t [e_{t+1}^{\varphi_e} \lambda_{t+1}] , \\ \lambda_t &= \left(Y_{t+1} + \frac{1 - \delta - \gamma}{1 - \delta} X_t - \frac{1 - \delta - \psi}{1 - \delta} \gamma Y_t \right)^{-\omega} , \\ X_{t+1} &= (1 - \delta) X_t + \psi Y_{t+1} , \\ Y_{t+1} &= z_t e_t^\alpha .\end{aligned}\tag{17}$$

Finally, we assume that μ_t and z_t both follow AR(1) processes in logs:

$$\begin{aligned}\log(\mu_t) &= \rho_\mu \log(\mu_{t-1}) + \sigma_\mu \varepsilon_{\mu,t} \\ \log(z_t) &= \rho_z \log(z_{t-1}) + \sigma_z \varepsilon_{z,t} .\end{aligned}$$

5.1 Solution and Estimation

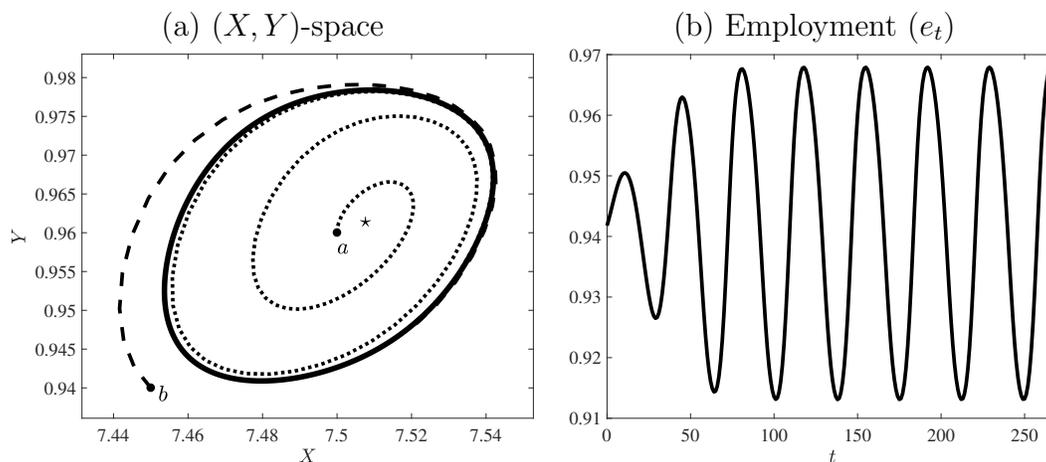
The model is solved to a third-order approximation (see Appendix C for details). The depreciation rate δ and the labor share α are calibrated to 0.05 and 0.67, respectively. For given values of the other parameters, β and Θ together determine the steady state employment rate \bar{e} , and are not separately identified. Thus, rather than setting β and Θ , we instead treat \bar{e} as a parameter directly. The remaining parameters are estimated as follows. First, we shut down the technology shock z , and then estimate the parameters (including \bar{e}) to match as closely as possible (i) the (ergodic) mean unemployment rate in the data, (ii) the spectrum of log-hours worked³⁷ at periodicities between 2 and 50 quarters, and (iii) the skewness of de-trended log-hours.³⁸ We then repeat the process with μ_t shut down instead of z_t , giving us two different parameterizations of the model: a μ -shock version, and a z -shock version.

³⁷The unemployment rate series is the rate for ages 16 years and over. The hours series is non-farm business hours worked, divided by population. The labor market series are both from the BLS, while population is taken from the FRED database. The sample period is 1948I to 2015II. The spectrum is computed by smoothing the raw periodogram using a Hamming kernel.

³⁸Log-hours was de-trended using a band-pass filter that removes fluctuations longer than 50 quarters.

Parameter estimates and details of the models' fit can be found in Appendix D. Both estimated versions of the model feature a locally unstable steady state and an attractive limit cycle. This is illustrated in Figure 6 for the μ -shock version, and Figure 7 for the z -shock version. Panel (a) in each Figure illustrates the evolution of the system in $y = (X, Y)$ -space when we feed $\log(\mu_t) = \log(z_t) = 0$ into the model in every period,³⁹ beginning from two different initial points: $y_0 = a$ (dotted trajectory), and $y_0 = b$ (dashed trajectory). The system clearly always converges to the limit cycle, which is shown as the solid black closed curve. The stars in the Figure, meanwhile, show the locations of the non-stochastic SS's.

Figure 6: Limit Cycle: μ -Shock Model



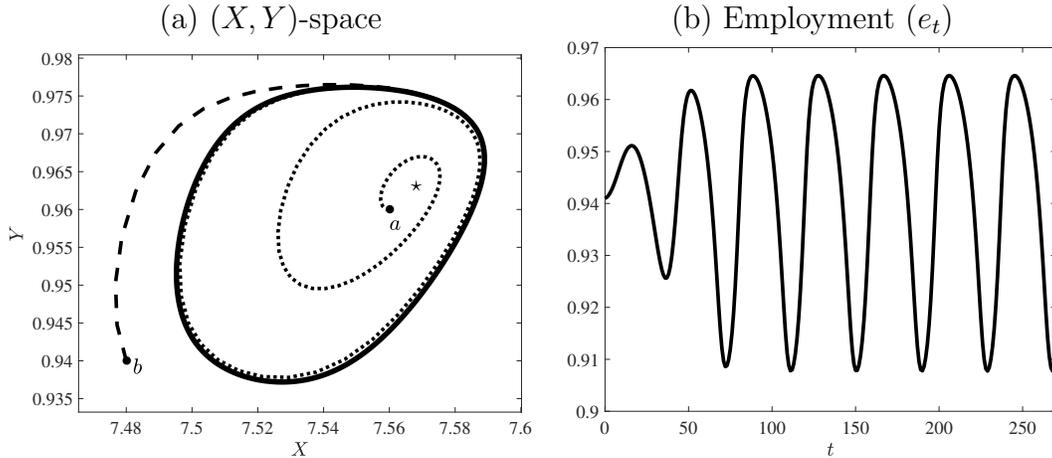
Notes: The dotted and dashed curves in panel (a) show the evolution of the system beginning from $a = (7.5, 0.96)$ and $b = (7.45, 0.94)$, respectively, when we feed in $\log(\mu_t) = 0$ every period. Panel (b) shows the path for e_t over time beginning from a when we feed in $\log(\mu_t) = 0$ every period.

Panel (b) in Figures 6-7 shows the evolution of employment e_t over time associated with the dotted trajectory from panel (a). As one can see, by about $t = 80$ the system has essentially converged to the cycle in both models, after which it repeats itself every 40 quarters or so.

Finally, Figure 8 shows, for the two different models, typical paths for e_t when we feed in a random draw for the shock processes. Comparing panel (b) in Figures 6-7 with Figure 8, one can see that the shock processes not only cause the amplitudes of

³⁹Note that this corresponds only to the case where the *realizations* of the shocks are all zero. In particular, agents in the economy nonetheless continue to believe that shocks will arrive in the future, and this affects their current choices.

Figure 7: Limit Cycle: z -Shock Model



Notes: The dotted and dashed curves in panel (a) show the evolution of the system beginning from $a = (7.56, 0.96)$ and $b = (7.48, 0.94)$, respectively, when we feed in $\log(z_t) = 0$ every period. Panel (b) shows the path for e_t over time beginning from a when we feed in $\log(z_t) = 0$ every period.

the cycles to fluctuate, but also makes the cycles quite irregular: unlike in Figures 6-7 where there is a recession every 40 quarters, in Figure 8 the interval between successive recessions varies, ranging from approximately 26 to 42 quarters in the μ -shock model, and from 17 to 42 quarters in the z -shock model.

5.2 Accuracy

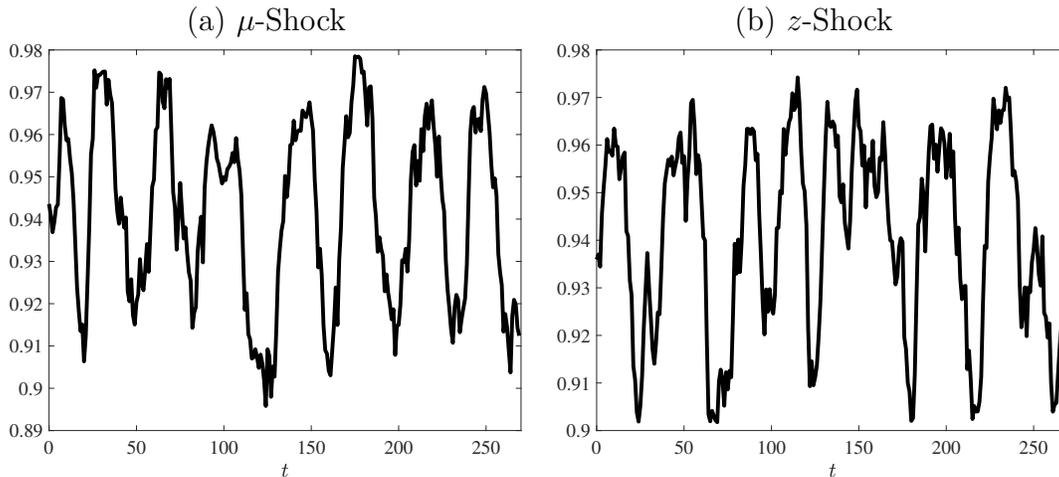
Relative to other popular non-linear approaches (e.g., projection methods), standard perturbation methods are typically easy to implement and have fast running times, while often still producing a reasonable degree of accuracy. As we show now for the two estimated versions of the model, this remains true for the method introduced in this paper.

A common way to measure the accuracy of a solution to a DSGE model, first proposed by Judd (1992), is to consider normalized Euler equation errors. Letting

$$\mathcal{E}_t \equiv \frac{(\mu_t^{-1} Q(e_t) \mathbb{E}_t [e_{t+1}^{\varphi_e} \lambda_{t+1}])^{-\frac{1}{\omega}} + \gamma C_{t-1}}{C_t} - 1, \quad (18)$$

we can express the Euler equation for this model simply as $\mathcal{E}_t \equiv 0$. This Euler equation will generally only hold exactly for the true solution, however. Thus, one way to evaluate the quality of an approximate solution is to compute $|\mathcal{E}_t|$ for a given

Figure 8: Stochastic Simulations



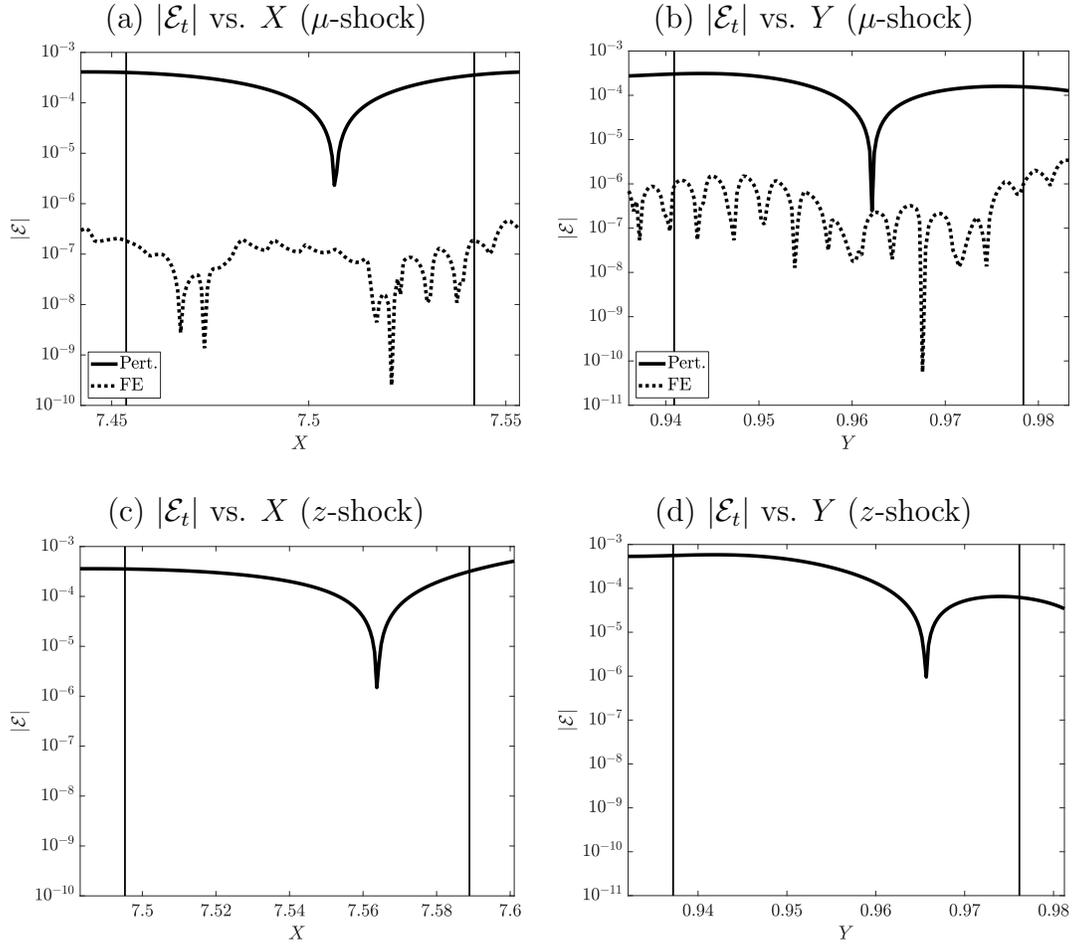
value of the state vector using the approximated policy functions, and see how close it is to zero. Note that, for a particular approximate solution, the numerator of the fraction in (18) is what C_t should be equal to, while the denominator is what C_t actually is. Thus, roughly speaking, one can interpret $|\mathcal{E}_t|$ as the size of the optimization error an agent would make by using the approximate solution instead of the true one, expressed as a percentage of consumption (see Judd & Guu (1997)).

The solid curve in panel (a) of Figure 9 plots the magnitudes of the errors for the perturbation approximation to the μ -shock model over a range of X , holding the other state variables (i.e., Y and the shock) at their non-stochastic SS values.⁴⁰ The errors are reported in base-10 logarithms, where, for example, $|\mathcal{E}_t| = 10^{-4}$ indicates an optimization error worth 0.01% of consumption. Panel (b) plots the errors for a cross-section of Y instead, holding X and the shocks at their non-stochastic SS values. Panels (c) and (d) show the same as panels (a) and (b), but for the z -shock model. For reference, the vertical lines in each panel show the minimum and maximum values the relevant state variable achieves along the deterministic limit cycle.

Figure 9 shows that, at least for these cuts of X and Y , the quality of the perturbation approximation is generally quite good in both versions of the model. Within the limit-cycle ranges, the errors are always less than 0.058% of consumption regardless of the state variable or model, and in many cases are one or two orders of magnitude

⁴⁰See Appendix F for details of how the Euler errors were computed.

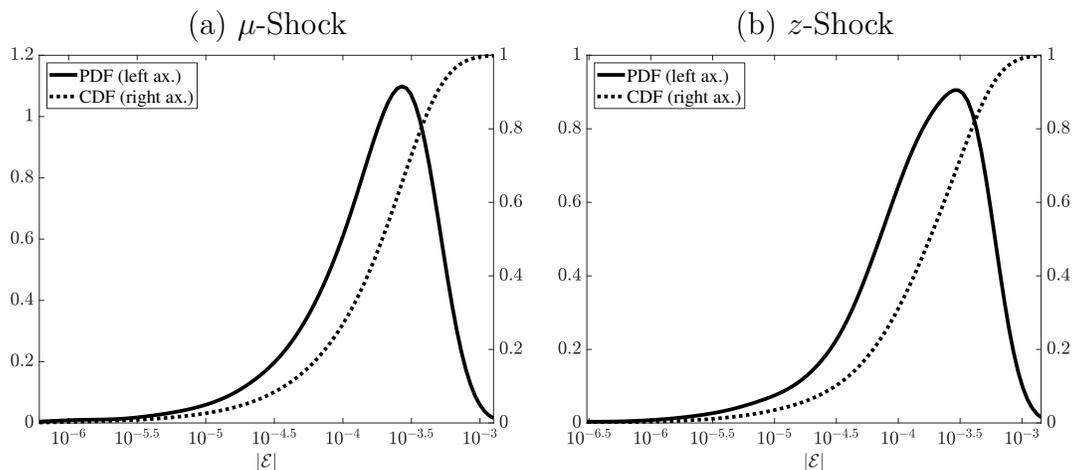
Figure 9: Euler Errors vs. X, Y



smaller than that.

Figure 9 gives cuts of the Euler errors allowing one state variable to vary, holding the others at their SS values. Since the state variables tend to move together in particular ways, however, these may give a poor representation of what a *typical* error is. An alternative is to consider the ergodic distribution of the errors. Panels (a) and (b) of Figure 10 show these ergodic distributions for the μ - and z -shock models, respectively. Specifically, they show non-parametric estimates of the PDF (left axis) and CDF (right axis) of the distributions of Euler errors associated with the ergodic distributions of the state variables.⁴¹ As the plots indicate, for both models the bulk of the errors are in the range of 0.01-0.03% of consumption, and are less than 0.1% over 99% of the time.

Figure 10: Euler Errors, Ergodic Distribution



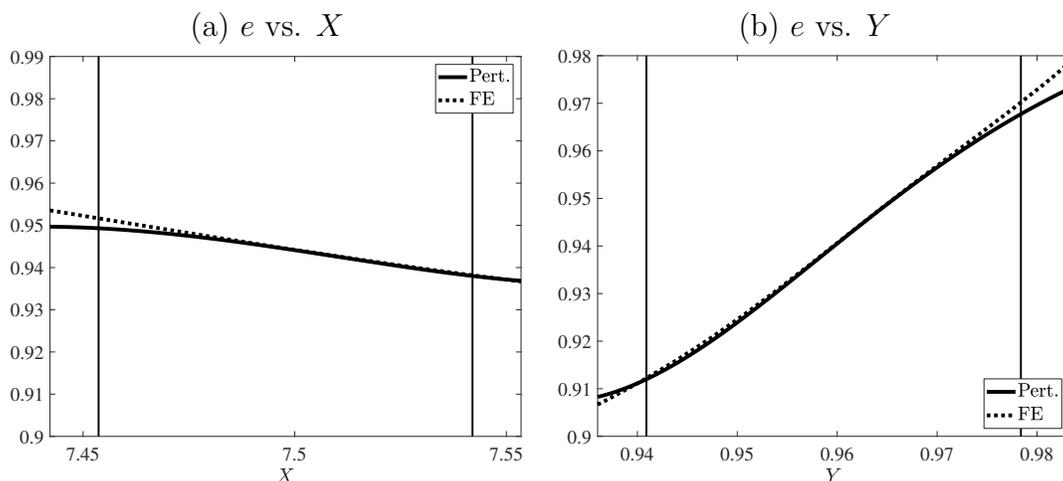
Another way to measure the accuracy of the perturbation solution is to compare the approximated policy functions to functions approximated using a global method that is known to be capable of delivering highly accurate solutions. To that end, I re-solve the μ -shock model using a finite elements (FE) method (see Appendix E for details). As shown by the dotted curves in panels (a) and (b) of Figure 9, this

⁴¹Specifically, we first simulate $T = 20,000$ periods of data from the model, and then drop the first 10,000 periods as a burn-in sample. For each of the remaining 10,000 periods of simulated data, we compute the Euler error for the corresponding state (X_t, Y_t, μ_t, z_t) . The solid curves in each panel plot kernel density estimates for the resulting empirical distributions of Euler errors, while the dotted curves show the associated CDFs.

FE approximation delivers Euler errors that are quite small (almost always less than 0.0001% of consumption within the limit cycle range), suggesting that the FE solution is highly accurate. Thus, when it comes to the policy functions, any significant difference between the perturbation and FE approximations can be taken to be an indication of error in the perturbation solution.

Figure 11 shows transversal cuts of the perturbation (solid curve) and FE (dotted curve) policy functions for employment for the μ -shock model. Panel (a) shows the function for the same range of X as in the Euler error plot in panels (a) and (c) of Figure 9, with Y and μ set to their respective non-stochastic SS levels. Panel (b) plots the same thing, except over the range of Y , with X now set to its non-stochastic SS level. The perturbation policy functions track the FE ones quite well throughout most of the limit cycle range, though in each case they are starting to deviate somewhat towards one of the ends of these ranges, which suggests the accuracy of the perturbation solution is beginning to deteriorate in those regions. This highlights the standard drawback of perturbation methods, which is that they are inherently local, and their accuracy tends to suffer as the system moves further away from the non-stochastic SS. For applications where one is interested in the behavior of the economy in regions far away from this SS (e.g., outside of the limit cycle bounds in this example), a global method such as FE would be called for.

Figure 11: Employment Policy Function (μ -Shock Model)



5.3 Computation Times

While the Euler error plots in Figure 9 make it clear that the global FE approximation dominates the local perturbation approximation in terms of accuracy, where the perturbation approximation shines is in terms of computation time. A solution run of the model for a single parameterization—and including both the μ and z stochastic processes as state variables—takes around 0.008 seconds on a 3.8GHz AMD Ryzen 9 3900X computer running Windows 10, and implemented entirely in MATLAB. This includes all steps of the solution algorithm, including finding the first-order solution coefficients associated with the elements of \mathcal{W}^* , using these to obtain the higher-order coefficients, and then simulating 1,000 periods of data using the resulting approximation in order to verify that the system does not explode. As a result of these fast running times, the estimations undertaken above were easily feasible, each taking less than 20 minutes.

For the FE approximation, the time it takes to solve the model depends on a variety of factors, including the desired accuracy of the solution, the quality of the initial guess, and how well-suited the particular non-linear solver used is to the problem. In this case, for a single FE solution to the μ -shock model—which features one less state variable (i.e., z) than the perturbation approximation—computation times ranged from around 10 minutes (i.e., $\sim 80,000$ times longer than the perturbation method) to obtain a solution that was roughly as accurate as the perturbation solution, up to several hours to obtain the highly accurate FE approximation used in Section 5.2.⁴² While solution times of this magnitude may be acceptable for applications involving only a handful of parameterizations, they make estimation exercises—such as the ones conducted above using the perturbation approximation—largely infeasible using FE methods.

⁴²For comparison purposes, both the perturbation and FE solutions were implemented entirely in MATLAB. While low-level languages such as Fortran or C++ can in some cases offer significant speed advantages over MATLAB, the expected gain is generally small for tasks that can be highly vectorized/parallelized, as FE methods can. Thus, while there may be some modest speed gains from using Fortran/C++, they would almost certainly be insufficient to make FE methods competitive with perturbation methods.

6 Conclusion

Non-linear rational expectations models are becoming increasingly important in the macroeconomic literature. While there are a variety of algorithms that can be used to solve these models, perturbation methods are relatively simple to implement, fast to execute, and can produce a reasonable degree of accuracy. Indeed, outside of the simplest economic models, the perturbation approach is the only one that is sufficiently fast that it can feasibly be used for most estimation purposes.

Standard perturbation algorithms fundamentally require the linear approximation of a solution to be linearly stable, so that all trajectories converge to the steady state. As a result, these methods cannot be applied to models that feature attractive limit cycles or chaos—which generally have linearly unstable solutions—effectively excluding a large class of interesting and potentially empirically relevant economic models. As argued in this paper, however, this limitation is not inherent to perturbation methods in general, only to the precise way in which existing algorithms implement them. In particular, I generalize existing perturbation algorithms and show that the approach can be used to solve rational expectations models that may (or may not) feature limit cycles or chaos.

I then apply the method to a fully microfounded New Keynesian model in Section 5 and show that the method delivers reasonably accurate solutions, and in a small fraction of the time that it takes to solve the model using a more accurate global solution method (finite elements).

References

- Aruoba, S. B., Fernández-Villaverde, J. & Rubio-Ramírez, J. F. (2006), ‘Comparing solution methods for dynamic equilibrium economies’, *Journal of Economic Dynamics and Control* **30**(12), 2477–2508.
- Beaudry, P., Galizia, D. & Portier, F. (2016), Is the Macroeconomy Locally Unstable and Why Should We Care?, in J. A. Parker & M. Woodford, eds, ‘NBER Macroeconomics Annual’, Vol. 31, The University of Chicago Press, pp. 479–530.
- Beaudry, P., Galizia, D. & Portier, F. (2020), ‘Putting the Cycle Back into Business Cycle Analysis’, *American Economic Review* **110**(1), 1–47.
- Blanchard, O. J. & Kahn, C. M. (1980), ‘The Solution of Linear Difference Models under Rational Expectations’, *Econometrica* **48**(5), 1305–1311.
- Caldara, D., Fernández-Villaverde, J., Rubio-Ramírez, J. F. & Yao, W. (2012), ‘Computing DSGE models with recursive preferences and stochastic volatility’, *Review of Economic Dynamics* **15**, 188–206.
- Fernández-Villaverde, J., Rubio-Ramírez, J. F. & Schorfheide, F. (2016), ‘Solution and estimation methods for DSGE models’, *Handbook of Macroeconomics* **2**, 527–724.
- Gaspar, J. & Judd, K. L. (1997), ‘Solving Large-Scale Rational-Expectations Models’, *Macroeconomic Dynamics* **1**(1), 45–75.
- Golub, G. H. & Van Loan, C. F. (1996), *Matrix Computations*, 3rd edn, The Johns Hopkins University Press, Baltimore, Maryland.
- Gomme, P. & Klein, P. (2011), ‘Second-order approximation of dynamic models without the use of tensors’, *Journal of Economic Dynamics and Control* **35**, 604–615.
- Judd, K. L. (1992), ‘Projection Methods for Solving Aggregate Growth Models’, *Journal of Economic Theory* **58**(2), 410–452.
- Judd, K. L. (1996), ‘Approximation, Perturbation, and Projection Methods in Economic Analysis’, *Handbook of Macroeconomics* **1**, 509–585.
- Judd, K. L. (1998), *Numerical Methods in Economics*, MIT press.
- Judd, K. L. & Guu, S.-M. (1997), ‘Asymptotic methods for aggregate growth models’, *Journal of Economic Dynamics and Control* **21**(6), 1025–1042.

- King, R. G. & Watson, M. W. (1998), ‘The Solution of Singular Linear Difference Systems Under Rational Expectations’, *International Economic Review* **39**(4), 1015–1026.
- Klein, P. (2000), ‘Using the generalized Schur form to solve a multivariate linear rational expectations model’, *Journal of Economic Dynamics & Control* **24**(10), 1405–1423.
- Kollman, R., Maliar, S., Malin, B. A. & Pichler, P. (2011), ‘Comparison of solutions to the multi-country Real Business Cycle model’, *Journal of Economic Dynamics and Control* **35**, 186–202.
- Lubik, T. A. & Schorfheide, F. (2003), ‘Computing sunspot equilibria in linear rational expectations models’, *Journal of Economic Dynamics and Control* **28**(2).
- Schmitt-Grohé, S. & Uribe, M. (2004), ‘Solving dynamic general equilibrium models using a second-order approximation to the policy function’, *Journal of Economic Dynamics and Control* **28**, 755–775.
- Sims, C. A. (2002), ‘Solving Linear Rational Expectations Models’, *Computational Economics* **20**(1-2), 1–20.
- Uhlig, H. (1999), A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily, *in* R. Marimon & A. Scott, eds, ‘Computational Methods for the Study of Dynamic Economies’, Oxford University Press, pp. 30–61.