Computation of solutions to dynamic models with occasionally binding constraints.

Tom D. Holden[1]*, School of Economics, University of Surrey*

Abstract: We construct the first algorithm for the perfect foresight solution of otherwise linear models with occasionally binding constraints, with fixed terminal conditions, that is guaranteed to return a solution in finite time, if one exists. We also provide a proof of the inescapability of the "curse of dimensionality" for this problem when nothing is known a priori about the model. We go on to extend our algorithm to deal with stochastic simulation, other non-linearities, and future uncertainty. We show that the resulting algorithm produces fast and accurate simulations of a range of models with occasionally binding constraints.

Keywords: *occasionally binding constraints, zero lower bound, computation, DSGE, linear complementarity problem*

JEL Classification: *C61, C63, E3, E4, E5*

This version: 30 July 2016

The latest version of this paper may be downloaded from:
https://github.com/tholden/dynareOBC/raw/master/ComputationalPaper.pdf

[1] Contact address: School of Economics, FASS, University of Surrey, Guildford, GU2 7XH, England
Contact email: thomas.holden@gmail.com Web-site: http://www.tholden.org/
The author would particularly like to thank Michael Paetz for joint work on an earlier algorithm for simulating models with occasionally binding constraints. Additionally, the author would like to thank the participants in numerous seminars at which versions of this paper were presented. Further special thanks are due to the following individuals for helpful discussions: Gary Anderson, Martin M. Andreasen, Paul Beaurdry, Saroj Bhattarai, Charles Brendon, Zsolt Csizmadia, Oliver de Groot, Michael B. Devereux, Bill Dupor, Roger Farmer, Michael Funke, William T. Gavin, Fabio Ghironi, Andy Glover, Luca Guerrieri, Pablo Guerrón-Quintana, Matteo Iacoviello, Tibor Illés, Michel Juillard, Hong Lan, Paul Levine, Jesper Lindé, Albert Marcet, Enrique Martinez-Garcia, Antonio Mele, Alexander Meyer-Gohde, Adrienn Nagy, Matthias Paustian, Jesse Perla, Frank Portier, Søren Ravn, Alexander Richter, Jonathan Swarbrick, Simon Wren-Lewis and Carlos Zarazaga. Financial support provided to the author by the ESRC and the EC is also greatly appreciated. Furthermore, the author gratefully acknowledges the use of the University of Surrey FASS cluster for the numerical computations performed in this paper, and the University of Washington for providing the office space in which the bulk of this paper was written.
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement "Integrated Macro-Financial Modelling for Robust Policy Design" (MACFINROBODS, grant no. 612796).

1. Introduction

Almost any macroeconomic model one might write down contains an occasionally binding constraint somewhere. These range from the positivity constraint on investment and the upper bound on hours, in simple real business cycle models, to the occasionally binding borrowing constraints on households, firms or banks in models of financial frictions. In recent times, one of the most prominent occasionally binding constraint (OBC) has been the zero lower bound (ZLB) on nominal interest rates.

While models with occasionally binding constraints may be simulated computationally with global approaches, these do not scale well to larger models, and they have no convergence guarantees for non-optimal economies such as New Keynesian models. Additionally, traditional approaches to simulating larger models, such as perturbation, cannot capture OBCs. While there have been recent advances in simulating models with OBCs (e.g. Holden (2010), Holden and Paetz (2012), Judd, Maliar, and Maliar (2012) or Guerrieri and Iacoviello (2015)), as yet there is still no algorithm that is capable, for any model, of determining in finite time if it has a solution, and finding it if it does. Furthermore, while the literature contains fast algorithms (such as that of Holden and Paetz (2012) and Guerrieri and Iacoviello (2015)) and accurate ones (such as that of Judd, Maliar, and Maliar (2012)), it arguably still lacks an algorithm in the sweet spot of reliable accuracy with sufficient speed to be used in the daily business of policy by policy makers accustomed to working with large models.

In this paper, we attack the problem of understanding the behaviour of models with occasionally binding constraints by developing new computational tools to handle them. These build upon the theoretical results on such models from a companion paper (Holden 2016). We prove that for otherwise linear models with a fixed terminal condition, under perfect foresight, the problem of finding if there is a solution, and if so what it is, may be represented as the solution to a mixed integer linear programming problem. This gives the first algorithm for such problems which is guaranteed to complete in finite time. We also prove results on the computational complexity of the problem which imply that finding a perfect foresight solution is computationally as difficult as mixed integer linear programming, implying that this representation is in a sense the best possible. We go on to exploit the particular properties of pruned perturbation approximations (Kim et al. 2008) to convert this robust perfect foresight solver into a general solution algorithm for non-linear dynamic stochastic models. To complement this paper, we provide a toolkit

("DynareOBC") implementing the described algorithm which provides robust, accurate and scalable simulation of models with OBCs.[2]

Our theoretical results on the computational complexity of finding a solution to models with occasionally binding constraints imply that were there an algorithm which could solve general non-stochastic models with OBCs in an amount of time that is polynomial in the number of states in the model, then such an algorithm could also be used for solving in polynomial time any problem the solution of which could be verified in polynomial time. This would imply, for example, efficient methods of breaking all standard forms of cryptography used to secure internet banking. Of course, neither we nor the computer science profession believe that such a polynomial time algorithm can exist, hence this provides a proof by contradiction of the "curse of dimensionality" for this class of models. It also raises doubts about the realism of assuming rational expectations in these situations.

With the caveat of these theoretical results on maximum speed, we go on to present an algorithm for solving general models with occasionally binding constraints, as efficiently as is possible. In the otherwise linear, perfect foresight case, this is given by our representation of the solution as that of a mixed integer linear programming problem, a problem for which incredibly efficient solvers already exist. The key idea of the algorithm is that an OBC provides a source of endogenous news about the future. When a shock hits, driving the economy to the bound in some future periods, that tells us that in those future periods, the (lower) bounded variable will be higher that it would be otherwise.[3] The integer programming aspect of the problem comes because whether or not we are at the bound one period is a yes-no binary variable.

Our algorithm for finding the required news shocks to impose the zero lower bound is guaranteed to return a solution in finite time when one exists, and when there is no solution, the algorithm returns a certificate of this in finite time instead. This contrasts with approaches based on a fixed point iteration for which non-existence is not normally detectable in finite time, since one cannot rule out that the algorithm would converge if only it were left for another hour/day/year. Furthermore, where there are multiple solutions, our algorithm always returns one minimising an intuitive criterion, with a free parameter that enables the user to select the desired "type" of equilibrium. This algorithm may be applied to stochastic models using the idea of the extended path algorithm of Fair and Taylor (1983), and can take future uncertainty into account following the stochastic extended path algorithm of Adjemian and Juillard (2013), as discussed below.

[2] DynareOBC is available from: https://github.com/tholden/dynareOBC and is discussed further in section 3.4.

[3] The idea of imposing the zero lower bound by adding news shocks is also present in Holden (2010), Hebden et al. (2011), Holden and Paetz (2012) and Bodenstein et al. (2013). News shocks were introduced to the literature by Beaudry and Portier (2006).

For models that are non-linear even apart from the constraint, we exploit the convenient properties of pruned perturbation approximations (Kim et al. 2008), which enables the base algorithm to be applied to higher order approximations with only minimal modifications. Although this does not capture the risk of hitting the bound, it does at least help capture the fact that at the bound, the economy is a substantial distance from steady-state, and so the slopes of variables' responses will have changed.

To capture the effects of future uncertainty, such as precautionary motives to avoid the bound, we use a modified version of the stochastic extended path algorithm of Adjemian and Juillard (2013) that is designed to exploit both the special properties of our inner solution algorithm, and the special properties of pruned perturbation. Whereas in the original Adjemian and Juillard (2013) approach, integrating over S periods of future uncertainty required a number of solutions to the perfect foresight problem that was exponential in both S and the number of shocks, we are able to integrate over the same number of periods of future uncertainty with only polynomial in S solutions of the perfect foresight problem. In practice, this means that we can integrate over enough periods of future uncertainty to capture even one hundred periods of future uncertainty, and even in medium scale models.

Strictly, the original (2013) stochastic extended path approach is not fully consistent with rationality, since it is equivalent to assuming that agents act as if the uncertainty in all future periods would be resolved next period. However, we will see that in practice it performs well. The authors of the original stochastic path method now have a more complicated version that is fully consistent with rationality (Adjemian and Juillard 2016), however it appears to be too computationally expensive to be used in large models with even moderate numbers of periods of future uncertainty, so we continue to use the original method, which nonetheless provides a good approximation.

Our paper is structured as follows. In the following section, we present our solution algorithm for otherwise linear perfect foresight models, and discuss the computational complexity of the problem. We then extend this to non-linear, non-perfect foresight models in section 3. In section 4 we go on to assess the algorithm's numerical accuracy and speed, and discuss its relationship to other algorithms in the literature. Section 5 concludes. All files needed for the replication of this paper's numerical results are included in the "Examples" directory of the author's DynareOBC toolkit.[4]

[4] These files may be viewed online at https://github.com/tholden/dynareOBC/tree/master/Examples.

2. Computation of solutions in the otherwise linear case

In this section, we present the key theoretical results that establish representations of the perfect foresight solution of an otherwise linear model with occasionally binding constraints. We start by defining the problem to be solved, and examining its relationship both to the problem without OBCs, and to a related problem with news shocks to the bounded variable. Using the news shock representation, we demonstrate that solving the model with OBCs is equivalent to solving a linear complementarity problem. This material closely follows this paper's companion theoretical paper (Holden 2016), and the reader is referred to that paper for proofs and further details.

Given the linear complementarity representation, we then examine the computational complexity of the problem. Finally, we show that this linear complementarity representation is in turn representable as the solution to a mixed integer linear programming problem.

2.1. Problem set-ups

We start by describing the problem set-up without bounds. Suppose that for $t \in \mathbb{N}^+$, (i.e. $t \in \mathbb{N}, t > 0$), the first order conditions of some model may be represented as:

$$\left(\hat{A} + \hat{B} + \hat{C}\right)\hat{\mu} = \hat{A}\hat{x}_{t-1} + B\hat{x}_t + \hat{C}\mathbb{E}_t\hat{x}_{t+1} + \hat{D}\varepsilon_t,$$

where $\hat{\mu} \in \mathbb{R}^{\hat{n}}$ and $\hat{x}_t \in \mathbb{R}^{\hat{n}}$, $\varepsilon_t \in \mathbb{R}^m$, $\mathbb{E}_{t-1}\varepsilon_t = 0$ for all $t \in \mathbb{N}^+$, and suppose that $\hat{x}_0$ is given as an initial condition. Throughout this paper, we will refer to first order conditions such as these as "the model", conflating them with the optimisation problem(s) which gave rise to them.

Furthermore, suppose that $\varepsilon_t = 0$ for $t > 1$, as in an impulse response or perfect foresight simulation exercise. Additionally, we assume the existence of a terminal condition of the form $\hat{x}_t \to \hat{\mu}$ as $t \to \infty$, coming, for example, from the source model's transversality constraints. In a New Keynesian model, this terminal condition might also capture a belief in the credibility of the central bank's long-run inflation target.

For $t \in \mathbb{N}^+$, define $x_t := \begin{bmatrix} \hat{x}_t \\ \varepsilon_{t+1} \end{bmatrix}$, $\mu := \begin{bmatrix} \hat{\mu} \\ 0 \end{bmatrix}$, $A := \begin{bmatrix} \hat{A} & \hat{D} \\ 0 & 0 \end{bmatrix}$, $B := \begin{bmatrix} \hat{B} & 0 \\ 0 & I \end{bmatrix}$, $C := \begin{bmatrix} \hat{C} & 0 \\ 0 & 0 \end{bmatrix}$, then, for $t \in \mathbb{N}^+$:

$$(A + B + C)\mu = Ax_{t-1} + Bx_t + Cx_{t+1}, \tag{1}$$

and we have the extended initial condition $x_0 = \begin{bmatrix} \hat{x}_0 \\ \varepsilon_1 \end{bmatrix}$, and the extended terminal condition $x_t \to \mu$ as $t \to \infty$. Expectations have disappeared since there is no uncertainty after period 0. Thus, the problem of solving the original model has the same form as that given in:

Problem 1 Suppose that $x_0 \in \mathbb{R}^n$ is given. Find $x_t \in \mathbb{R}^n$ for $t \in \mathbb{N}^+$ such that $x_t \to \mu$ as $t \to \infty$, and such that for all $t \in \mathbb{N}^+$, equation (1) holds.

We make the following assumption in all of the following:

Assumption 1 For any given $x_0 \in \mathbb{R}^n$, Problem 1 has a unique solution, which takes the form $x_t = (I - F)\mu + Fx_{t-1}$, for $t \in \mathbb{N}^+$, where $F = -(B + CF)^{-1}A$, and where all of the eigenvalues of F are weakly inside the unit circle.

Sims's (2002) generalisation of the standard Blanchard-Kahn (1980) conditions is necessary and sufficient for this. Further, to avoid dealing specially with the knife-edge case of exact unit eigenvalues (even if they are constrained to the part of the model that is solved forward), in the following we rule it out with the subsequent assumption, which is, in any case, a necessary condition for perturbation to produce a consistent approximation to a source non-linear model, and which is also necessary for the linear model to have a unique steady-state:

Assumption 2 $\det(A + B + C) \neq 0$.

The combination of Assumption 1 and Assumption 2 imply that all of the eigenvalues of F are strictly inside the unit circle.

We are interested in models featuring occasionally binding constraints. We will concentrate on models featuring a single zero lower bound type constraint in their first equation, which we treat as defining the first element of x_t. Generalising from this special case is straightforward, and is discussed in section 2.3. First, let us write $x_{1,t}$, $I_{1,\cdot}$, $A_{1,\cdot}$, $B_{1,\cdot}$, $C_{1,\cdot}$ for the first row of x_t, I, A, B, C (respectively) and $x_{-1,t}$, $I_{-1,\cdot}$, $A_{-1,\cdot}$, $B_{-1,\cdot}$, $C_{-1,\cdot}$ for the remainders. Likewise, we write $I_{\cdot,1}$ for the first column of I, and so on. Then we are interested in the solution to:

Problem 2 Suppose that $x_0 \in \mathbb{R}^n$ is given. Find $T \in \mathbb{N}$ and $x_t \in \mathbb{R}^n$ for $t \in \mathbb{N}^+$ such that $x_t \to \mu$ as $t \to \infty$, and such that for all $t \in \mathbb{N}^+$:

$$x_{1,t} = \max\{0, I_{1,\cdot}\mu + A_{1,\cdot}(x_{t-1} - \mu) + (B_{1,\cdot} + I_{1,\cdot})(x_t - \mu) + C_{1,\cdot}(x_{t+1} - \mu)\},$$
$$(A_{-1,\cdot} + B_{-1,\cdot} + C_{-1,\cdot})\mu = A_{-1,\cdot}x_{t-1} + B_{-1,\cdot}x_t + C_{-1,\cdot}x_{t+1},$$

and such that $x_{1,t} > 0$ for $t > T$.

Note that in this problem we are implicitly ruling out any solutions which get permanently stuck at an alternative steady-state, by assuming that the terminal condition remains as before. In the monetary policy context, this amounts to assuming that the central banks' (positive) long-run inflation target is credible, as the alternative steady-state features deflation.

Since $x_{1,t} \to \mu_1$ as $t \to \infty$, it is without loss of generality to assume the existence of a $T \in \mathbb{N}$ such that $x_{1,t} > 0$ for $t > T$, but this T will play an important role in the below, so we introduce it now. We continue to assume that there is no uncertainty after period 0, so, in this non-linear model, the path of the endogenous variables will not

necessarily match up with the path of their expectation in a richer model in which there was uncertainty after period 0.

In many models, the occasionally binding constraint comes from the KKT conditions of an optimisation problem. We provide several examples of translating KKT conditions into the max operator form in section 4.3 and 4.4. The intuition is that one can use the model's equations to find the value the (lower) constrained variable would take were there no constraint and were the Lagrange multiplier on the constraint equal to zero today. This gives a "shadow" value of the constrained variable, and the actual value it takes will be the maximum of the bound and this shadow value. An alternative approach is to replace complementary slackness conditions of the form $f(x)\lambda = 0$, where $0 \leq f(x)$ and $0 \leq \lambda$, with equations of the form $0 = \max\{-f(x), -\lambda\}$.

We will analyse Problem 2 with the help of solutions to the auxiliary problem:

Problem 3 Suppose that $T \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$ and $y_0 \in \mathbb{R}^T$ is given. Find $x_t \in \mathbb{R}^n, y_t \in \mathbb{R}^T$ for $t \in \mathbb{N}^+$ such that $x_t \to \mu$, $y_t \to 0$, as $t \to \infty$, and such that for all $t \in \mathbb{N}^+$:

$$(A + B + C)\mu = Ax_{t-1} + Bx_t + Cx_{t+1} + I_{\cdot,1}y_{1,t-1},$$
$$y_{T,t} = 0, \qquad \forall i \in \{1, \dots, T-1\}, \qquad y_{i,t} = y_{i+1,t-1}.$$

This may be thought of as a version of Problem 1 with news shocks up to horizon T added to the first equation. The value of $y_{t,0}$ gives the news shock that hits in period t, i.e. $y_{1,t-1} = y_{t,0}$ for $t \leq T$, and $y_{1,t-1} = 0$ for $t > T$.

2.2. The linear complementarity representation

For future reference, let $x_t^{(3,k)}$ be the solution to Problem 3 when $x_0 = \mu$, $y_0 = I_{\cdot,k}$ (i.e. a vector which is all zeros apart from a 1 in position k). Then, we show in Holden (2016) that by linearity, for arbitrary y_0 the solution to Problem 3 when $x_0 = \mu$ is given by:

$$x_t - \mu = \sum_{k=1}^{T} y_{k,0}\left(x_t^{(3,k)} - \mu\right).$$

Furthermore, if $M \in \mathbb{R}^{T \times T}$ satisfies:

$$M_{t,k} = x_{1,t}^{(3,k)} - \mu_1, \qquad \forall t, k \in \{1, .., T\}, \tag{2}$$

i.e. M horizontally stacks the (column-vector) relative impulse responses to the news shocks, then this result implies that for arbitrary y_0, the path of the first variable in the solution to Problem 3 when $x_0 = \mu$ is given by: $(x_{1,1:T})' = \mu_1 + My_0$, where $x_{1,1:T}$ is the row vector of the first T values of the first component of x_t. Furthermore, for both arbitrary x_0 and y_0, the path of the first variable in the solution to Problem 3 is given by: $(x_{1,1:T})' = q + My_0$, where $q := \left(x_{1,1:T}^{(1)}\right)'$ and $x_t^{(1)}$ is the unique solution to Problem 1, for the given x_0.[5] This ease in solving Problem 3 given y_0 will be crucial to the efficiency of our solution algorithm for Problem 2.

[5] This representation was also exploited by Holden (2010) and Holden and Paetz (2012).

The two problems are linked by the following proposition which we prove in Holden (2016):

Proposition 1 The following hold:

1) Let $x_t^{(3)}$ be the unique solution to Problem 3 when initialized with some x_0, y_0. Then $x_t^{(3)}$ is a solution to Problem 2 when initialized with x_0 if and only if $y_0 \geq 0$, $y_0 \circ (q + My_0) = 0$, $q + My_0 \geq 0$ and $x_{1,t}^{(3)} \geq 0$ for all $t \in \mathbb{N}$ with $t > T$.
2) Let $x_t^{(2)}$ be any solution to Problem 2 when initialized with x_0. Then there exists a $y_0 \in \mathbb{R}^T$ such that $y_0 \geq 0$, $y_0 \circ (q + My_0) = 0$, $q + My_0 \geq 0$, such that $x_t^{(2)}$ is the unique solution to Problem 3 when initialized with x_0, y_0.

(Holden 2016)

Proposition 1 establishes that providing we initially choose T sufficiently high, to find a solution to Problem 2, it is sufficient to solve the following problem instead:

Problem 4 Suppose $q \in \mathbb{R}^T$ and $M \in \mathbb{R}^{T \times T}$ are given. Find $y \in \mathbb{R}^T$ such that $y \geq 0$, $y \circ (q + My) = 0$ and $q + My \geq 0$. We call this the **linear complementarity problem (LCP)** (q, M). (Cottle 2009)

Since the computational complexity of this problem will be determined by the properties of the M matrix, we would like to determine if the particular structure of our M matrix implies any such properties. Unfortunately, however, it turns out that we can infer nothing about M just from knowing that it stacks impulse responses to news shocks in some model. In particular, we prove the following proposition in online appendix A:

Proposition 2 For any matrix $\mathcal{M} \in \mathbb{R}^{T \times T}$, there exists a model in the form of Problem 2 with a number of state variables given by a quadratic in T, such that $M = \mathcal{M}$ for that model, where M is defined as in equation (2), and such that for all $q \in \mathbb{R}^T$, there exists an initial state for which $q = q$, where q is the path of the bounded variable when constraints are ignored.

2.3. Generalisations to richer otherwise linear models

It is straightforward to generalise these results to less restrictive otherwise linear models with occasionally binding constraints.

Firstly, if the constraint is on a variable other than $x_{1,t}$, or in another equation than the first, then it is immediately clear that all of the results must go through as before (just by relabelling and rearranging). Furthermore, if the constraint takes the form of $z_{1,t} = \max\{z_{2,t}, z_{3,t}\}$, where $z_{1,t}$, $z_{2,t}$ and $z_{3,t}$ are linear expressions in the contemporaneous values, lags and leads of x_t , then, assuming without loss of generality that $z_{3,\cdot} > z_{2,\cdot}$ in steady-state, we have that $z_{1,t} - z_{2,t} = \max\{0, z_{3,t} - z_{2,t}\}$. Hence, adding a new auxiliary variable $x_{n+1,t}$, with the associated equation $x_{n+1,t} =$

$z_{1,t} - z_{2,t}$, and replacing the constrained equation with $x_{n+1,t} = \max\{0, z_{3,t} - z_{2,t}\}$, we have a new equation in the form covered by our original results. Moreover, if rather than a max we have a min, we just use the fact that if $z_{1,t} = \min\{z_{2,t}, z_{3,t}\}$, then $-z_{1,t} = \max\{-z_{2,t}, -z_{3,t}\}$, which is in the form covered by the generalisation just established.

We may also readily deal with multiple occasionally binding constraints, following the representation used in Holden and Paetz (2012). Suppose there are c constrained variables in the model. For $a \in \{1, \dots, c\}$, let $q^{(a)}$ be the path of the a^{th} constrained variable in the absence of all constraints. For $a, b \in \{1, \dots, c\}$, let $M^{(a,b)}$ be the matrix created by horizontally stacking the column vector relative impulse responses of the a^{th} constrained variable to magnitude 1 news shocks at horizon $0, \dots, T-1$ to the equation defining the b^{th} constrained variables. For example, if $c = 1$ so there is a single constraint, then we would have that $M^{(1,1)} = M$ as defined in equation (2). Finally, let:

$$q := \begin{bmatrix} q^{(1)} \\ \vdots \\ q^{(c)} \end{bmatrix}, \qquad M := \begin{bmatrix} M^{(1,1)} & \dots & M^{(1,c)} \\ \vdots & \ddots & \vdots \\ M^{(c,1)} & \dots & M^{(c,c)} \end{bmatrix},$$

and let y be a solution to the LCP (q, M). Then the vertically stacked paths of the constrained variables in a solution which satisfies these constraints is given by $q + My$, and again any solution satisfying the constraints corresponds to a solution to the LCP.

2.4. On the difficulty of the problem

Before presenting the computational algorithm for solving the linear complementarity problem, we give a note of caution. If no properties of the matrix M are known a priori, then Problem 4 is provably a computationally difficult problem; more formally, it may be shown to be "strongly-NP complete" (Chung 1989), and this remains true even if M is restricted to be a "P_0-matrix" (Kojima et al. 1991). The class of P_0-matrices is defined in Holden (2016), but for here it suffices to note that the best behaved LCPs (i.e. those with a unique equilibrium for all q) are those in which M is a P-matrix, where the definition of a P_0-matrix just replaces a strict inequality in the definition of a P-matrix with a weak one. Thus, an arbitrarily small step away from the well behaved P-matrix class is sufficient to produce a strongly-NP complete problem.

Strongly-NP completeness means that even if the inputs q and M have descriptions which are of a polynomial length in T, then if we could solve Problem 4 in an amount of time that was polynomial in T, we could also solve in polynomial time any problem for which the solution could be verified in polynomial time. In the language of computer science, this would mean that "P=NP", something almost all computer scientists believe to be false. The strength of computer scientists' conviction that this is not true is best exemplified by the fact that were P=NP, all commonly used forms of cryptography (such as those used to secure internet banking) could potentially be defeated.

Since there is a bijection between solutions for Problem 4 and solutions for Problem 2, this means that while forming expectations in linear models without occasionally binding constraints is computationally easy (polynomial algorithms exist for it), in models with OBCs, forming expectations may be incredibly difficult. It also means that we should be sceptical of claims of computational efficiency from other algorithms for solving models with OBCs. A proof that such algorithms actually ran in time polynomial in the number of state variables of the model, on all models, would again function as a proof that "P=NP", since we showed in Proposition 2 that there is a model corresponding to any $T \times T \;\; M$ matrix, featuring polynomial in T state variables.

A natural response to this is that in macroeconomics, we are only concerned with approximate solutions, whereas the previous computational complexity results were for exact solutions. In fact, allowing approximation error will not change these results. Note that a sufficiently accurate approximation to the solution would tell us when the constraint binds in the exact solution. However, the difficulty of the exact LCP comes from the fact that there are 2^T possible combinations of periods in which the constraint might bind, so no solution procedure can "quickly" tell us the periods constraint binds. Hence, the approximate problem cannot be easier than the original problem. More formally, we establish the following proposition in online appendix B:

Proposition 3 For any problem in the form of Problem 2, let $\mathcal{D}^* \subseteq \mathbb{R}^n$, and $p^*: \mathcal{D}^* \to \mathcal{D}^*$ be an exact policy function for Problem 2, by which we mean that:

1. For all $x \in \mathcal{D}^*$:
$$x_1 = \max\{0, I_{1,\cdot}\mu + A_{1,\cdot}(x-\mu) + (B_{1,\cdot} + I_{1,\cdot})(p^*(x) - \mu) + C_{1,\cdot}(p^*(p^*(x)) - \mu)\},$$
$$(A_{-1,\cdot} + B_{-1,\cdot} + C_{-1,\cdot})\mu = A_{-1,\cdot}x + B_{-1,\cdot}p^*(x) + C_{-1,\cdot}p^*(p^*(x)).$$
2. For all $x_0 \in \mathcal{D}^*$, if $x_t = p^*(x_{t-1})$ for all $t \in \mathbb{N}^+$, then $x_t \to \mu$ as $t \to \infty$.
3. For all $x_0 \notin \mathcal{D}^*$, there is no $(x_t)_{t=1}^{\infty} \subseteq \mathbb{R}^n$ solving the given instance of Problem 2.

Suppose that for all $\kappa, \epsilon > 0$, and for any problem in the form of Problem 2, we can calculate in time polynomial in n a set $\mathcal{D}_{\kappa,\epsilon} \subseteq \mathbb{R}^n$ with $\mathcal{D}^* \subseteq \mathcal{D}_{\kappa,\epsilon}$ and an approximate policy function $p_{\kappa,\epsilon}: \mathcal{D}_{\kappa,\epsilon} \to \mathbb{R}^n$, where membership of $\mathcal{D}_{\kappa,\epsilon}$ may be tested in time polynomial in n, where $p_{\kappa,\epsilon}$ may be evaluated in time polynomial in n, and where for all $x \in \mathcal{D}^*$ with $\|x - \mu\|_\infty < \kappa$, $\left\|p^*(x) - p_{\kappa,\epsilon}(x)\right\|_\infty < \epsilon$, then P=NP, i.e. all problems verifiable in polynomial time may be solved in polynomial time.

Providing one believes (along with almost all of the computer science profession) that P≠NP, this provides a reductio ad absurdum of our assumption that there was a general procedure capable of providing the policy function $p_{\kappa,\epsilon}$ in polynomial time. Thus, for example, global methods will never escape the curse of dimensionality in general models with OBCs, even using methods explicitly designed to do this such as

that of Judd, Maliar, and Maliar (2012). In fact, even proving the finiteness of algorithms for solving these problems is non-trivial (see e.g. Csizmadia and Illés 2006), and, for example, there is no reason to believe that the iterations in Guerrieri and Iacoviello (2015) will converge in finite time on all models.

Admittedly, for some special classes further discussed in online appendix C, it has been shown that the problem is solvable in polynomial time in T. However, as discussed further in that appendix, it turns out that checking whether M is in one of the relevant special cases is itself not possible in polynomial time, so this is of little use. The combination of the results of that appendix and the results of Holden (2016) imply that were there a general algorithm running in time polynomial in the number of state variables for testing if a particular model always had a unique solution or always had a unique solution when away from the bound, then that algorithm would also serve as a proof that P=NP, since Proposition 2 implies that a model could be constructed producing any given $T \times T$ M matrix, with polynomial in T state variables.

2.5. The mixed integer linear programming representation

Given that there is no reason to believe that there is a polynomial time algorithm to solve the LCPs we encounter, it is important that we choose an algorithm, which, although it may not complete in polynomial time in the worst case, is nonetheless as computationally efficient as possible, particularly on average. One way to do this is to reduce the problem of finding an LCP to the solution of a problem for which highly efficient algorithms and computational libraries are available. One such problem is mixed integer linear programming (MILP), for which algorithms are included in most major optimisation suites (e.g. CPLEX, Gurobi, XPress MP, MOSEK, etc.). Conveniently, we can reduce the LCP problem to the MILP one in a way that gives not only a solution when one exists, but also a definite answer on whether or not there is a solution. This is an improvement over more naïve approaches, such as those of Holden (2010), Holden and Paetz (2012) or Guerrieri and Iacoviello (2015), for which a failure of convergence may just mean that the optimiser got stuck at some local minimum.

To motivate the MILP representation, suppose that y solves the LCP (q, M). Then $y \geq 0$, $0 \leq q + My$ and if $y_j > 0$ then $(q + My)_j = 0$. Now let $\tilde{\omega} > 0$ be an arbitrary constant, let $\alpha := \min\{\|y\|_\infty^{-1}, \tilde{\omega}\|q + My\|_\infty^{-1}\} > 0$, (where $\|\cdot\|_\infty$ is the usual sup norm), let $\hat{y} := \alpha y$, let $1_{T\times 1}$ be a $T \times 1$ vector of ones, and let $z \in \{0,1\}^T$ be such that for all $j \in \{1, \dots, T\}$, $z_j = 1$ if and only if $y_j > 0$ (i.e. z is an indicator for being away from the bound). Then $0 \leq \hat{y} \leq 1_{T\times 1}$ and $0 \leq \alpha q + M\hat{y} \leq \alpha\|q + My\|_\infty \leq \tilde{\omega} 1_{T\times 1}$. Now, if $z_j = 0$ for some $j \in \{1, \dots, T\}$, then $y_j = 0$. Hence, in fact, $0 \leq \hat{y} \leq z$. Likewise, if $z_j = 1$, then

$y_j > 0$, so since y solves the LCP, $0 = \alpha(q + My)_j = (\alpha q + M\hat{y})_j$. Hence, similarly, $0 \leq \alpha q + M\hat{y} \leq \tilde{\omega}(1_{T\times 1} - z)$.

Moreover, for any $\tilde{\alpha} > \alpha$, we claim that there is no $z \in \{0,1\}^T$ such that $0 \leq \tilde{\alpha} y \leq z$ and $0 \leq \tilde{\alpha} q + M(\tilde{\alpha} y) \leq \tilde{\omega}(1_{T\times 1} - z)$. To see this, suppose for a contradiction that there were. Then $\tilde{\alpha}\|y\|_\infty \leq 1$, so $\alpha < \tilde{\alpha} \leq \|y\|_\infty^{-1}$. Hence, $\alpha = \tilde{\omega}\|q + My\|_\infty^{-1}$. But, by assumption $\tilde{\alpha}\|q + My\|_\infty \leq \tilde{\omega}$, hence $\tilde{\omega}\|q + My\|_\infty^{-1} = \alpha < \tilde{\alpha} \leq \tilde{\omega}\|q + My\|_\infty^{-1}$, which gives the required contradiction.

Therefore, α, $\hat{y}$, z are feasible for the following MILP problem (though they may not necessarily be the solution):

Problem 5 Suppose $\tilde{\omega} > 0$, $q \in \mathbb{R}^T$ and $M \in \mathbb{R}^{T\times T}$ are given. Find $\alpha \in \mathbb{R}$, $\hat{y} \in \mathbb{R}^T$, $z \in \{0,1\}^T$ to maximise α subject to the following constraints: $\alpha \geq 0$, $0 \leq \hat{y} \leq z$, $0 \leq \alpha q + M\hat{y} \leq \tilde{\omega}(1_{T\times 1} - z)$. We call this the **mixed integer linear programming (MILP) representation of the LCP** (q, M).

A version of this representation with $\tilde{\omega} = 1$ was first given by Pardalos and Rosen (1988), and its properties in that special case were proven by Rosen (1990).

We now establish that solutions of the MILP representation are solutions of the LCP. Suppose that α, $\hat{y}$, z solve Problem 5. If $\alpha = 0$, then there is no $\alpha > 0$ such that $0 \leq \hat{y} \leq z$, $0 \leq \alpha q + M\hat{y} \leq \tilde{\omega}(1_{T\times 1} - z)$. Now, we showed above that if the LCP (q, M) had a solution, then there would be an $\alpha > 0$, $\hat{y}$ and z which were feasible for Problem 5, hence, this $\alpha > 0$ provides a lower bound on the solution to Problem 5. Thus, if $\alpha = 0$, the LCP cannot have a solution.

Alternatively, suppose $\alpha > 0$. Then if for some $j \in \{1, \dots, T\}$, $z_j = 1$, then $0 = (\alpha q + M\hat{y})_j$, and if for some $j \in \{1, \dots, T\}$, $z_j = 0$, then $\hat{y}_j = 0$. Thus, $\hat{y} \circ (\alpha q + M\hat{y}) = 0$. Finally, define $y := \frac{\hat{y}}{\alpha} \geq 0$, hence $\hat{y} = \alpha y$, $0 \leq q + My$ and $y \circ (q + My) = 0$, i.e. y solves the LCP (q, M).

This establishes the following result:

Proposition 4 Suppose $\tilde{\omega} > 0$, $q \in \mathbb{R}^T$ and $M \in \mathbb{R}^{T\times T}$ are given. If y solves Problem 4, then the solution to Problem 5 has $\alpha \geq \min\{\|y\|_\infty^{-1}, \tilde{\omega}\|q + My\|_\infty^{-1}\}$. If the solution to Problem 4 is unique, then this last inequality holds with equality, and $\hat{y} = \alpha y$, $z = \begin{cases} 1 & \text{if } y_j > 0 \\ 0 & \text{if } y_j = 0 \end{cases}$ in the solution to Problem 5. Conversely, if α, $\hat{y}$, z solve Problem 5, then if $\alpha = 0$, Problem 4 has no solution, and if $\alpha > 0$, then $y := \frac{\hat{y}}{\alpha}$ solves Problem 4.

This result establishes that we can use the MILP representation both to find out if the LCP problem has a solution, and to find a solution when one exists. Furthermore, by varying $\tilde{\omega}$ we can determine which solution is returned, when there are multiple. In the limit as $\tilde{\omega} \to 0$, the MILP solver will return the solution which minimises $\|q + My\|_\infty$, and in the limit as $\tilde{\omega} \to \infty$, the MILP solver will return the solution to the

LCP which minimises $\|y\|_\infty$. The former objective would ensure that the returned solution does not generate large fluctuations in the path of the constrained variable, and the latter would reduce the fluctuations in other variables. Intermediate values of $\tilde{\omega}$ result in a solution being returned that features balanced concern for these two extremes. In practice, we suggest choosing $\tilde{\omega} = \omega\|q\|_\infty$ where $\omega > 0$ is another constant, to ensure that the solution returned scales appropriately with q. We suggest $\omega = 1000$, which places heavy weight on minimising $\|y\|_\infty$, without inducing numerical instabilities. It would also be possible to make ω stochastic in order to capture sunspot solutions to the model.

We can further constrain the solution returned in the presence of multiplicity by solving Problem 5 first with $T = 0$ (i.e. testing if $q \geq 0$), then with $T = 1$, and so on. Doing this ensures that the time to finally escape the bound is minimised, and this is the approach that is followed by default in the author's DynareOBC toolkit. While this means that we may have to solve multiple LCPs in the worst case in which the required T is large, by ensuring that on average we only have to solve small LCPs, it may still increase average performance.

3. Algorithms for general non-linear models, without perfect foresight

Up to now, we have solely been concerned with the perfect foresight solution of models which were linear apart from the occasionally binding constraint. In this section, we will apply these insights to the solution of general non-linear models, allowing for future uncertainty, i.e. we attempt to solve the following general problem:

Problem 6 Suppose that $x_0 \in \mathbb{R}^n$ is given and that $f\colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^c \times \mathbb{R}^m \to \mathbb{R}^n$, $g, h\colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^c \times \mathbb{R}^m \to \mathbb{R}^c$ are given continuously $d \in \mathbb{N}^+$ times differentiable functions. Find $x_t \in \mathbb{R}^n$ and $r_t \in \mathbb{R}^c$ for $t \in \mathbb{N}^+$ such that for all $t \in \mathbb{N}^+$:

$$0 = \mathbb{E}_t f(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t),$$

$$r_t = \mathbb{E}_t \max\{h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t), g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\}$$

where $\varepsilon_t \sim \text{NIID}(0, \Sigma)$, where the max operator acts elementwise on vectors, and where the information set is such that for all $t \in \mathbb{N}^+$, $\mathbb{E}_{t-1}\varepsilon_t = 0$ and $\mathbb{E}_t\varepsilon_t = \varepsilon_t$.

We construct our algorithm in two steps. First, we present an algorithm which accounts for the effects of uncertainty that would be present even without the bound, but which treats hitting the bound as a probability zero event. Then, we extend the aforementioned algorithm to capture the risk of hitting the bound in the future. The algorithms are implemented in the author's open source "DynareOBC" toolkit,[6]

[6] Available from http://github.org/tholden/dynareOBC.

which extends Dynare (Adjemian et al. 2011) with the ability to deal with OBCs, and which is further discussed in section 3.4.

3.1. Dealing with non-linearity other than the bounds

Setup and assumptions Given a non-linear, but $d \in \mathbb{N}$ times continuously differentiable model, a common practice in macroeconomics is to take a perturbation approximation to the model around its deterministic steady-state. Given that high order perturbation approximations are often unstable, the use of a "pruned" approximation (Kim et al. 2008) is usually advisable. We will proceed along similar lines, taking a perturbation approximation to the model ignoring the bound, and then imposing the bound on the approximated model. The advantage for our purposes of the pruned approximation is that the result is linear in an augmented state space, which will assist us transferring results from the linear case to the non-linear one.

We start by making a further assumption that is necessary for us to be able to construct a perturbation approximation to the model without the bound.

Assumption 3 In the setup of Problem 6, there exists $\mu_x \in \mathbb{R}^n$ and $\mu_r \in \mathbb{R}^c$ such that:

$$0 = f(\mu_x, \mu_x, \mu_x, \mu_r, 0),$$
$$\mu_r = \max\{h(\mu_x, \mu_x, \mu_x, \mu_r, 0), g(\mu_x, \mu_x, \mu_x, \mu_r, 0)\},$$

and such that for all $a \in \{1, \dots, c\}$, $\big(h(\mu_x, \mu_x, \mu_x, \mu_r, 0)\big)_a \neq \big(g(\mu_x, \mu_x, \mu_x, \mu_r, 0)\big)_a$.

This is necessary because if any of the constraints just bind in steady-state, then the equation defining the corresponding element of μ_r is not differentiable at $(\mu_x, \mu_x, \mu_x, \mu_r, 0)$, preventing us from taking a perturbation approximation.

Henceforth, we suppose without loss of generality that $h(\mu_x, \mu_x, \mu_x, \mu_r, 0) \ll g(\mu_x, \mu_x, \mu_x, \mu_r, 0)$. We claim that we may further assume without loss of generality that $h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) \equiv 0$. First, note we can rewrite the equation defining r_t as:

$$r_t = h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) + \mathbb{E}_t \max\{0, g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) - h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\}.$$

Then if we define $\hat{x}_t := \begin{bmatrix} x_t \\ r_t \end{bmatrix}$ for all $t \in \mathbb{N}$, and:

$$\hat{g}(\hat{x}_{t-1}, \hat{x}_t, \hat{x}_{t+1}, \hat{r}_t, \varepsilon_t) := g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) - h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t),$$
$$\hat{f}(\hat{x}_{t-1}, \hat{x}_t, \hat{x}_{t+1}, \hat{r}_t, \varepsilon_t) := f(x_{t-1}, x_t, x_{t+1}, g_1(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) + \hat{r}_t, \varepsilon_t),$$

then for all $t \in \mathbb{N}^+$:

$$0 = \mathbb{E}_t \hat{f}(\hat{x}_{t-1}, \hat{x}_t, \hat{x}_{t+1}, \hat{r}_t, \varepsilon_t), \qquad \hat{r}_t = \mathbb{E}_t \max\{0, \hat{g}(\hat{x}_{t-1}, \hat{x}_t, \hat{x}_{t+1}, \hat{r}_t, \varepsilon_t)\},$$

which is again in the form of Problem 6. Thus, without loss of generality, we can indeed assume that $h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t) \equiv 0$, meaning that we work with the system:

$$0 = \mathbb{E}_t f(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t), \qquad r_t = \mathbb{E}_t \max\{0, g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\},$$

where $g(\mu_x, \mu_x, \mu_x, \mu_r, 0) \gg 0$.

First order approximations Now, suppose that we believe that a first order approximation gives adequate accuracy away from the bound. This system is locally $d \geq 1$ times differentiable in a neighbourhood of $(\mu_x, \mu_x, \mu_x, \mu_r, 0)$, so we can certainly

take a first order approximation around this point. Doing this gives the following approximation for the equation for r_t:

$$r_t = \mu_r + g_1(x_{t-1} - \mu_x) + g_2(x_t - \mu_x) + g_3\mathbb{E}_t(x_{t+1} - \mu_x) + g_4(r_t - \mu_r) + g_5\varepsilon_t,$$

where $g_1, \dots, g_5$ are the matrices of partial derivatives of g with respect to its first to fifth arguments, respectively. This approximation obviously completely ignores the bound. Thus, we propose to increase its accuracy by imposing the bound on the linearized equations, i.e. by instead working with the equation:

$$r_t = \max\{0, \mu_r + g_1(x_{t-1} - \mu_x) + g_2(x_t - \mu_x) + g_3\mathbb{E}_t(x_{t+1} - \mu_x) + g_4(r_t - \mu_r) + g_5\varepsilon_t\}.$$

This gives a system of equations in nearly the same form as that for which we developed a solution algorithm in section 2, the only difference being the presence of expectations operators and uncertainty. In our base approach, we deal with these following the extended path algorithm of Fair and Taylor (1983). I.e., if we are currently in period t of a simulation, we assume that the agents in the model believe that for all $s > t$, $\varepsilon_s = 0$. Thus, in each period of a simulation run, we merely have to solve a perfect foresight problem of the form of Problem 2, using the methods of section 2. We then advance one period, draw new shocks, and repeat the process.

We can also use a slightly modified form of the representation of Problem 3 to track the endogenous "news" that is coming from the bound, following Holden and Paetz (2012). In particular, we are effectively replacing the bounded equations with equations of the form:

$$r_{a,t} = \mathbb{E}_t\big(g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\big)_a + I_{1,\cdot}y_t^{(a)},$$

for all $a \in \{1, \dots, c\}$, where, for all $a \in \{1, \dots, c\}$:

$$y_{T,t}^{(a)} = \eta_{T,t}^{(a)}, \qquad \forall i \in \{1, \dots, T-1\}, \qquad y_{i,t}^{(a)} = y_{i+1,t-1}^{(a)} + \eta_{i,t}^{(a)},$$

implying that $y_{1,t}^{(a)} = \sum_{i=1}^{T} \eta_{i,t-i}^{(a)}$, where $\eta_{i,t}^{(a)}$ contains the news about the likelihood of the a^{th} bound binding in period $t + i$. Whereas the ys found by the LCP solver will always be positive, the implied $\eta_{i,t}^{(a)}$ need not necessarily be positive, as shocks may hit today which result in the economy moving away from the bound.

Higher order approximations Applying the methods of section 2 to models solved with a higher order approximation is slightly more difficult than in the linear case.

Recall that in linear models, we started by introducing news shocks to the bounded equation(s) and stacking the impulse responses to these news shocks into the M matrix. In order to use these impulse responses to tell us about the path of the bounded variable, we exploited the fact that the impulse response to a linear combination of shocks is the same linear combination of the individual impulse responses. It was this linearity that gave the $q + My$ representation of the path of the bounded variable.

Now, consider what would happen in a pruned or non-pruned second order approximation following a similar linear combination of shocks. Under such an approximation, it is no longer true in general that the impulse response to a linear

combination of shocks is a linear combination of the impulse responses, since the second order approximation captures interactions between the shocks. However, if the partial derivative of f and g with respect to each of the shocks being combined is zero, then the shocks only have second or higher order effects, hence, any interaction between them would be a fourth order effect or higher, and so would not be captured contemporaneously by the second order approximation. The period after the shocks hit, though, linearity would again be broken if a non-pruned second order approximation had been taken, since the slope of the response of the states to their lags vary with the states' levels. This is not true under a second order pruned perturbation approximation though, since under such an approximation, the solution takes the form:

$$x_t = \mu_x + x^{(0)} + x_t^{(1)} + x_t^{(2)}, \qquad x_t^{(1)} = \alpha x_{t-1}^{(1)} + \beta_0 \varepsilon_t,$$
$$x_t^{(2)} = \alpha x_{t-1}^{(2)} + \frac{1}{2}\beta_{22}\left(x_{t-1}^{(1)} \otimes x_{t-1}^{(1)}\right) + \beta_{20}\left(x_{t-1}^{(1)} \otimes \varepsilon_t\right) + \frac{1}{2}\beta_{00}(\varepsilon_t \otimes \varepsilon_t),$$

where $x^{(0)} \in \mathbb{R}^n$ is a constant, $x_t^{(1)}$ is the first order component of the approximation, $x_t^{(2)}$ is the second order component of the approximation, and where β_0 has zero columns corresponding to each shock with respect to which the partial derivatives of f and g is zero (Kim et al. 2008). Thus, $x_t^{(1)}$ does not respond to any shocks for which the partial derivatives of f and g are zero, and hence $x_t^{(2)}$ and x_t are linear in such shocks.

In light of this discussion, in order to preserve the $q + My$ representation, we just need to define M as stacking the impulse responses of the bounded equation(s) to news shocks which hit the bounded equation(s) raised to the power of two, rather than in levels. This generalises to higher order pruned perturbation approximations as one would expect. Hence, in a d^{th} order pruned perturbation approximation, we replace the bounded equations with equations of the form $r_{a,t} = \mathbb{E}_t\left(g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\right)_a + I_{1,\cdot} y_t^{(a)}$, for all $a \in \{1, \dots, c\}$, where, now, for all $a \in \{1, \dots, c\}$:

$$y_{T,t}^{(a)} = \kappa\left(\eta_{T,t}^{(a)}\right)^d, \qquad \forall i \in \{1, \dots, T-1\}, \qquad y_{i,t}^{(a)} = y_{i+1,t-1}^{(a)} + \kappa\left(\eta_{i,t}^{(a)}\right)^d,$$

where κ is a very small constant in order to ensure that the presence of the additional shocks does not have an unwanted risk effect elsewhere in the model. We obviously scale the impulse responses which make up M by $\frac{1}{\kappa}$ to correct for this. In practice, we do not actually need to augment the model we approximate (in, e.g. Dynare) with all of these additional equations. This is because in the limit as $\kappa \to 0$, the generated impulse responses tend to the impulse responses used to construct the M matrix at first order.[7]

Hence, in order to impose the bound in non-linear models solved by perturbation, we can proceed much as we did at first order. At each time step, we first evaluate the

[7] As the instant response must be the same, and the subsequent response is given by $x_t - x^{(0)} = \alpha(x_{t-1} - x^{(0)})$ in both cases.

expected path of the bounded variable(s) in the absence of bounds, and stack the results in q. Thanks to the augmented state space representation of pruned perturbation solutions, this is possible without any Monte-Carlo simulation, as we show in online appendix D. Then, we use the M matrix derived from the first order approximation, and calculated as in section 2.2 in the LCP (q, M). Finally, we use the solution to this LCP to calculate the required offsets to each variable this period, again based on the first order approximation to the model. Since we are not actually augmenting the model's state space, even for high degree approximations to the model, imposing the bound will not slow down simulation much more than it does at order one.

3.2. Integrating over future uncertainty

The downside to the approach discussed in the previous section is that the news shocks that hit the bounded equation(s) will not be conditionally mean zero, i.e. $\mathbb{E}_{t-1}\left(\eta_{i,t}^{(a)}\right)^d \neq 0$. For example, in a model with a zero lower bound on nominal interest rates, stochastic discount factors (i.e. β_t is stochastic), and persistence in these discount factors, the higher is the state of the discount factor, the higher the chances of hitting the bound today, thus $\mathbb{E}_{t-1}\left(\eta_{i,t}^{(a)}\right)^d$ will be decreasing in this state. This failure of rationality with respect to expectations at the bound stems from the fact that we are still treating the bound in a pseudo-perfect foresight manner. In each period, agents act as if they believed that no future news shocks would ever hit the bounded equation. Due to the strict convexity of the $x \mapsto \max\{0, x\}$ mapping and Jensen's inequality, this manifests itself as a systematic downward bias in expectations of r_t.

To rectify this bias, we need to integrate over future uncertainty to calculate the expectation of the cumulated news shocks (the elements of y). We do this following the original stochastic extended path approach of Adjemian and Juillard (2013). In period t, this approach approximates the value of x_t in the model of Problem 6 by the solution to the system:

$$
\begin{aligned}
0 &= \mathbb{E}_t f(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t), \\
r_t &= \mathbb{E}_t \max\{h(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t), g(x_{t-1}, x_t, x_{t+1}, r_t, \varepsilon_t)\}, \\
\forall s \in \mathbb{N}^+, \qquad 0 &= f(x_{t+s-1}, x_{t+s}, x_{t+s+1}, r_{t+s}, \kappa_s \varepsilon_{t+s}), \\
\forall s \in \mathbb{N}^+, & \\
r_t &= \max\{h(x_{t+s-1}, x_{t+s}, x_{t+s+1}, r_{t+s}, \varepsilon_{t+s}), g(x_{t+s-1}, x_{t+s}, x_{t+s+1}, r_{t+s}, \kappa_s \varepsilon_{t+s})\},
\end{aligned}
$$

where $\kappa_0, \kappa_1, \ldots$ control the degree of future uncertainty considered. This is equivalent to supposing that in period t agents believe that in period $t+1$ they will be told the value of all future shocks (i.e. $\varepsilon_{t+1}, \varepsilon_{t+2}, \ldots$). From the perspective of period t, all future shocks are uncertain, meaning that this should capture well the effect of risk. Further justification for this approach comes from the fact that if the model is linear, and $\kappa_0 = \kappa_1 = \cdots$, then by the law of iterated expectations, there is no approximation at all.

However, in our context this will be much easier than in the general fully non-linear context of Adjemian and Juillard (2013). In particular, in the basic algorithm of that paper, to integrate over S periods of future uncertainty, in a model with m shocks, they have to solve the perfect foresight model p^{mS} times, for some constant $p \geq 2$. While they are able to reduce this somewhat through the removal of low weighted quadrature nodes to produce a sparse tree of shocks, the resulting distribution of paths will significantly under-estimate the true variance of the model, and they still have to solve the perfect foresight model many more times when the number of shocks is high.

In our context, we will be able to do much better. In particular, we will be able to attain comparable accuracy with the evaluation of only a polynomial in S number of solutions of the perfect foresight problem, regardless of the number of shocks in the model. Furthermore, since solving an LCP is much easier than solving a general fully non-linear perfect foresight problem, each of these solutions is orders of magnitude faster for us. The key to our invariance to the number of shocks in the model is the fact that in the absence of any bounds, we are able to write down a closed form expression for the conditional covariance of the bounded variables, thanks to the properties of pruned perturbation solutions. To be slightly more specific, suppose that $w_{t,s}$ is the value the bounded variables would take at s if the constraints did not apply from period t onwards. Then, we are able to calculate $\text{cov}_t\left(w_{t,t+i}, w_{t,t+j}\right)$, for $t, i, j \in \mathbb{N}$, without any numerical integration. The derivation of this is contained in online appendix E.

For clarity of presentation, let us assume that there is a single bounded variable. As ever, the generalisation to multiple bounded variables will be straightforward. Now, ideally we would like to integrate over infinitely many periods of future uncertainty, but clearly this is not practical in reality. Instead, Adjemian and Juillard (2013) advocate integrating over $S \in \mathbb{N}^+$ periods of future uncertainty, and then ignoring uncertainty from period $S+1$ onwards. By introducing a "discontinuity" in time of this sort, we would risk getting spurious movement in the expected path of variables around S periods into the future. Indeed, this occurred in some early numerical experiments that took this approach. Instead then, we apply a smooth windowing function to the variance of shocks. In particular, if the "true" shock covariance matrix is Σ, then when considering uncertainty at horizon k, we instead use:

$$\widehat{\Sigma}_k := \frac{1}{2}\left(1 + \cos\left(\pi \frac{\min\{k-1, S\}}{S}\right)\right)\Sigma.$$

Even with a time varying covariance matrix, it is still straightforward to calculate:

$$\Omega_t := \text{var}_t[[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]'],$$

following the calculations in online appendix E. Then for the purposes of integration, we make the approximation that:

$$[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]' \sim \text{N}(\mathbb{E}_t[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]', \Omega_t).$$

Of course the cosine window is ad hoc, but so too is the step-function window used by Adjemian and Juillard (2013). The legitimacy of both come from the fact that as S increases, the approximation error should fall. One further argument in favour of our cosine window is that it is widely used in signal processing due to its low distortion in the frequency domain (see e.g. Harris 1978). In this literature, it is termed the Hann or Hanning window. We also note that for most DSGE models, the additional error coming from the normal approximation will be minimal, since it is exact at first order, and higher order approximations are usually dominated by their first order terms.

Given this normal approximation, integration is then relatively straightforward. We first take the Schur decomposition of Ω_t, giving $\Omega_t = UDU'$, where U is an orthogonal matrix and $D \geq 0$ is diagonal, with the elements sorted in decreasing order. To reduce the integration dimension without overly affecting accuracy, we set any elements of D which are less than some small multiple (e.g. 1%) of the maximum element of D to zero, as these components are unlikely to have a big impact. Indeed, when S is very large, it may be advisable to set all but the $\hat{S}$ largest elements of D to zero, which means the cost of integration will scale in $\hat{S}$ not S. After these steps we that that $D = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix}$, where $\dim D_{11} = \hat{S} \times \hat{S}$ for some $\hat{S} \leq S$. Conformably partitioning U as $U = \begin{bmatrix} U_{\cdot 1} & U_{\cdot 2} \end{bmatrix}$,and defining $\Lambda \coloneqq U_{\cdot 1}\sqrt{D_{11}}$, we then have that $\Lambda\Lambda' \approx \Omega_t$. Then if $\zeta \sim \mathrm{N}(0, I_{\hat{S}})$, then $(\mathbb{E}_t[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]') + \Lambda\zeta$ has approximately the same distribution as $[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]'$. We have thus transformed the problem of integrating over the distribution of $[w_{t,t+1} \quad \cdots \quad w_{t,t+S}]'$ to that of integrating over the $\hat{S}$ independent standard normals making up ζ. We do this using either quasi-Monte Carlo methods, the sparse nested Gaussian cubature rules of Genz and Keister (1996), or the equal weight degree 3 monomial cubature rule with $2\hat{S} + 1$ nodes[8]. Of course, ideally we would like to break the domain of integration into pieces on which the integrand was differentiable, but this is not computationally practical for even moderately large $\hat{S}$.[9] We discuss the respective merits of these rules in the following section.

Whichever approach to integration is taken, we end with an approximation to the expected value of the "y" vector of cumulated news shocks needed to impose the bound. This y will imply a set of news shocks that hit today, just as it does when we

[8] While there is a degree 3 monomial cubature rule with only $2\hat{S}$ nodes, including the 0 node generally increases accuracy at no cost, since we are evaluating the point anyway. As well as the 0 node, we evaluate at $\pm\frac{1}{2}\sqrt{2+4\hat{S}}$ with respect to each coordinate, which is easily shown to give a degree 3 rule. The use of monomial rules has been promoted by e.g. Judd and Skrainka (2011).

[9] For example, with $\hat{S} = 1$, there are at least as many discontinuities as there are non-zero elements in Λ. While we could get the full set of discontinuities at arbitrary dimension using a parametric linear complementarity problem solver such as that of Jones and Morrari (2006) which works providing that M is general positive semi-definite, this is computationally intractable for $\hat{S}$ or T bigger than (about) ten, and integrating over all of these regions separately is computationally intractable even for much smaller $\hat{S}$.

ignore future uncertainty. We can thus proceed with the simulation exactly as we do in the case without integrating over future uncertainty.

3.3. Cubature rules

Degree 3 monomial rule The equal weight degree 3 monomial cubature rule with $2\hat{S}+1$ nodes rule exactly integrates all degree 3 monomials in the components of ζ. While a third order approximation to the cumulated news shocks, y, as a function of ζ may do a poor job at capturing this highly non-linear (and even non-differentiable) mapping, in practice the approximation to the integral is often surprisingly accurate. This is in a large part due to the robustness of the integration rule which stems from its equal, positive weights. All known higher degree integration rules that do not use more than polynomial in $\hat{S}$ nodes also feature negative weights on at least some nodes (Cools 2003), which means that their result is not guaranteed to lie within the convex hull of the source evaluations, and, in this case in which we are integrating a positive function (y), it further means the result can have the wrong sign.

Genz and Keister (1996) rules The Genz and Keister (1996) rules allow one to choose the maximum degree of monomial that should be integrated exactly, up to a maximum order of 51. The number of points used is $\mathrm{O}\left(\hat{S}^K\right)$, where $2K+1$ is the degree of monomial that is integrated exactly. When $K>0$ and $\hat{S}>1$, the rule features negative weights on at least one node, which means it is susceptible to the problems mentioned above. However, it has a few points in its favour. Firstly, by using negative weights, the rule is able to ensure that the maximum over the absolute vectors of integration points is independent of $\hat{S}$. This contrasts with the aforementioned rule in which the higher is $\hat{S}$, the further into the tails of the distribution one has to evaluate the integrand. Given the extreme non-linearity of the integrand, evaluating far into the tails can lead the equal weighted integration rule to produce a heavily upwards biased estimate of the integral. Secondly, by using a higher degree rule, we can generally obtain a better approximation to the integrand, despite its non-differentiability. Finally, the Genz and Keister (1996) rules are nested, which means that we can use an adaptive integration degree without wasting evaluations, continuing to increase the degree until approximate convergence. In practice, the results of these rules often repeatedly flip from biased down to biased up as the degree increases, due to the discontinuities. To lessen this, DynareOBC gives the option of averaging integral estimates of adjacent orders, which still integrates polynomials of the lower of the two orders exactly.

Quasi-Monte Carlo The final integration method we consider is quasi-Monte Carlo, generating points from a Sobol sequence (Sobol 1967). Given that the functions we are integrating are absolutely continuous (as they are piecewise polynomial, with a finite number of manifolds of non-differentiability), quasi-Monte Carlo with $2^{1+l}-1$ draws will produce an error that decays as $O\left(\frac{l^{\hat{S}}}{2^l}\right)$. With the Sobol sequence, the choice of $2^{1+l}-1$ integration points for some $l \in \mathbb{N}$ also ensures that the points are exactly mean zero, hopefully lessening overall bias. However, on functions that are well approximated by a polynomial, quasi-Monte Carlo will generally require far more evaluations of the integrand for a similar accuracy than the Genz and Keister (1996) rules would. Which dominates in practice will depend on the precise integrand, which in turn will depend on the model and its current state. At times where the bound is either highly likely to bind or highly likely not to bind, whatever future shocks hit, it is likely that the Genz and Keister (1996) rules will dominate, however, at times when the bound is only binding with moderate probability, quasi-Monte Carlo's "dumb" approach may give it better performance.

3.4. Further details on the DynareOBC toolkit

Code implementing all of the cubature algorithms discussed here is contained in the author's "DynareOBC" toolkit which is a suite of MATLAB files designed to augment the abilities of Dynare (Adjemian et al. 2011). The toolkit may be freely downloaded from http://github.org/tholden/dynareOBC, and this site also contains complete documentation for its assorted options.[10] To use it, one merely has to include a "max", "min" or "abs" in the MOD file describing the DSGE model to be simulated, and then to invoke DynareOBC with the MATLAB command "dynareOBC ModFileName.MOD".

Internally, DynareOBC uses the "YALMIP" (Löfberg 2004) MATLAB toolkit as an interface to a wide variety of open source and commercial mixed integer linear programming solvers. The distribution of DynareOBC comes with a variety of open source solvers, so DynareOBC is certainly not dependent on any particular commercial packages (other than MATLAB itself). DynareOBC also attempts to obtain a parametric solution to the LCP (q, M) for q which only violate the bound in at most the first few periods, using the MPT toolkit (Herceg et al. 2013), which in turn uses an algorithm due to Jones and Morrari (2006). The resulting parametric solution takes the form of a compiled MEX function, which, when passed a q, returns the y that solves the LCP. This reduces the number of times the LCP needs to be solved in inner loops, increasing performance.

[10] A PDF of the toolkit's documentation is available from: https://github.com/tholden/dynareOBC/raw/master/ReadMe.pdf.

Furthermore, DynareOBC includes efficient code for testing whether M is a P-matrix, based on an algorithm of Tsatsomeros and Li (2000), and can also test if M is an S-matrix or (strictly) semi-monotone, properties which all play important roles in the theoretical results on existence and uniqueness established in Holden (2016). Additionally, DynareOBC contains code for accurately simulating integrals over non-linear functions of the model's variables, facilitating, for example, the calculation of Jin and Judd (2002) style accuracy checks. Thus, DynareOBC functions as an easy to use, one stop shop for all queries one might have of a model with OBCs.

4. Performance of our algorithm

4.1. General comments

The accuracy of the numerical algorithm presented in this paper is almost an immediate consequence of results from the prior literature. In particular, Guerrieri and Iacoviello (2015) showed the surprising accuracy of a perfect foresight solution to an otherwise linear approximation to a stochastic non-linear model with occasionally binding constraints. When a first order approximation is taken to the underlying model, and there is a unique solution, our method will produce exactly the same answers as that of Guerrieri and Iacoviello (2015).[11,12] Relative to their method, our method improves along four dimensions. Firstly, our method is guaranteed to produce a result in finite time. Secondly, it gives guarantees about which solution is selected when there are multiple. Thirdly, it also applies to higher order pruned perturbation solutions to the underlying model. Given the evidence that higher order pruned perturbations solutions are considerably more accurate than first order approximations (see e.g. Lan and Meyer-Gohde 2013b), this is likely to produce substantial accuracy gains, particularly as OBCs are usually located far from the steady-state. Finally, our method takes future uncertainty into account, meaning that it captures precautionary effects, unlike the Guerrieri and Iacoviello (2015) method. That integrating over future uncertainty usually increases accuracy in non-linear models has been established by Adjemian and Juillard (2013), so here too substantial accuracy gains are almost certain.

Nonetheless, in the rest of this section, we give some further brief indications of the accuracy and speed of our approach, by applying the implementation of it in the DynareOBC toolkit to three indicative models with OBCs. We restrict ourselves to models for which standard dynamic programming results imply there is a unique solution to the fully nonlinear model, since the presence of multiple solutions makes fair accuracy comparisons nearly impossible, as one cannot guarantee that both

[11] This point was noted by Guerrieri and Iacoviello (2015) with respect to the solution algorithm by Holden and Paetz (2012).
[12] A numerical demonstration of their equivalence in this situation is contained in the "Tests" sub-folder of DynareOBC.

methods are returning the same solution. Likewise, we restrict ourselves to models for which an exact or near exact solution is available, to ensure that our results are not driven by poor quality accuracy measures. Away from these very well behaved problems, other algorithms will often fail to converge, while ours will always give an answer in finite time. For this reason, we do not present speed comparisons, since on a broad enough set of examples, our algorithm would always dominate.

4.2. A model with a closed form solution

We first apply our method to a model for which we can calculate an exact closed form solution, giving very reliable accuracy measures. The model has the property that integrating over a single period of uncertainty is enough for accuracy, making it an ideal test of the performance of different cubature rules.

Suppose the representative household in an economy chooses consumption C_t and zero net supply bond holdings B_t to maximise:

$$\mathbb{E}_t \sum_{k=0}^{\infty} \beta^k \frac{C_{t+k}^{1-\gamma} - 1}{1-\gamma},$$

subject to the restriction that:

$$A_t + R_{t-1} B_{t-1} = C_t + B_t$$

for all $t \in \mathbb{Z}$, where A_t's evolution is given by: $\log A_t = \log A_{t-1} + g_t$, where

$$g_t = \max\{0, (1-\rho)\bar{g} + \rho g_{t-1} + \varepsilon_t\}$$

and $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. This specification may be thought of as capturing the fact that technologies cannot be un-invented. Market clearing implies $A_t = C_t$ and $B_t = 0$ for all $t \in \mathbb{Z}$, and from this, a closed form expression for R_t may be derived.[13] Using this, we define simulation errors as the gap between the true value of $\log R_t$ and the simulated value.[14] In Table 1, we report errors along simulated paths of length 1000, after discarding an initial 100 periods of burn-in, where for simulation, we used the following parameters: $\beta := 0.99$, $\gamma := 5$, $\bar{g} := 0.005$, $\rho := 0.95$ and $\sigma := 0.007$. All cubature runs involve integrating over a single period of future uncertainty, which is sufficient here. We also report errors from the model with the bound removed, for comparison.

[13] In particular, $R_t = \left[\beta\left[\left(1 - \Phi(\frac{\mu_t}{\sigma})\right) + \left(1 - \Phi\left(\frac{\gamma\sigma^2 - \mu_t}{\sigma}\right)\right)\exp\left(\frac{\gamma^2\sigma^2}{2} - \gamma\mu_t\right)\right]\right]^{-1}$, where $\mu_t = (1-\rho)\bar{g} + \rho g_t$ and Φ is the standard normal cumulative distribution function.

[14] We also recorded errors in g_t, but these were essentially zero for all simulation runs.

Bound in Model	*Order*	*Cubature*	*Seconds* [15]	*Log_{10} Mean Abs Error*	*Log_{10} Root Mean Squared Error*	*Log_{10} Max Abs Error*	*Log_{10} Mean Abs Error at Bound* [16]
No	*1*	*N/A*	66	-3.213	-3.213	-3.213	
No	*2*	*N/A*	62	-16.82	-16.63	-15.78	
No	*3*	*N/A*	53	-16.70	-16.57	-15.95	
Yes	1	No	141	-2.435	-2.218	-1.882	-1.882
Yes	2	No	139	-2.425	-2.194	-1.862	-1.862
Yes	3	No	140	-2.425	-2.194	-1.862	-1.862
Yes	1	Monomial, Degree 3	274	-3.136	-3.073	-2.725	-3.131
Yes	2	Monomial, Degree 3	1537	-3.378	-3.172	-2.706	-3.893
Yes	3	Monomial, Degree 3	1397	-3.378	-3.172	-2.706	-3.893
Yes	2	Sparse, Degree 3	1794	-3.016	-2.777	-2.415	-2.415
Yes	2	Sparse, Degree 5	1840	-3.016	-2.777	-2.415	-2.415
Yes	2	Sparse, Degree 7	2009	-3.280	-3.032	-2.663	-2.663
Yes	2	QMC, 15 Points	1965	-3.040	-2.895	-2.664	-2.664
Yes	2	QMC, 63 Points	3184	-3.394	-3.260	-3.020	-3.020
Yes	2	QMC, 1023 Points	5197	-3.804	-3.638	-3.351	-3.351

Table 1: Accuracy in the model of bounded productivity growth

As may be seen from Table 1, our algorithm is generally very fast, as imposing the bound only doubles the running time without cubature. While this is a very simple model, the advantage of our algorithm is that running times are almost independent of the complexity of the model, so similar running times can be expected with even much larger models. These results also show that our algorithm is quite accurate, providing cubature is used. Without cubature, accuracy is below the accuracy of the first order approximation to the model without a bound. Since without cubature, at first order, our method will give identical answers to that of Guerrieri and Iacoviello (2015), this suggests that neither their algorithm nor ours without cubature can deliver comparable accuracy to that delivered by linearization in models without bounds. However, with cubature and a second order approximation, we can deliver errors that are lower in the model with the bound than those in the first order approximation to the model without the bound. As might be expected, quasi-Monte Carlo with many points is the most accurate integration method, but it does come at a significant time cost.

[15] All timings are "wall" time, and include time spent starting the parallel pool, time spent compiling code (although written in MATLAB, DynareOBC generates and compiles C code for key routines), and time spent calculating accuracy. Code was run on one of the following (very similar) twenty core machines: 2x E5-2670 v2 2.5GHz, 64GB RAM; or 2x E5-2660 v3 2.6GHz, 128GB RAM. Use of machines with network attached storage means that there may be some additional variance in these timings.

[16] Errors conditional on the bounded variable being less than 0.0001. The numbers for this column would be identical had we reported root mean squared errors or maximum absolute errors, conditional on being at the bound.

4.3. A model for which log-linearization gives an exact solution in the absence of bounds

We next apply our procedure to a model for which log-linearization gives the exact solution when bounds are removed, but which features quite different behaviour when bounds are included, due to precautionary effects. This helps us to isolate the inaccuracy coming from the presence of an occasionally binding constraint.

The model is a real business cycle model with a 100% depreciation rate (following Brock and Mirman (1972)), endogenous labour supply, and a lower bound on capital. The bound on capital may be thought of as an extreme case of a capital adjustment cost, with output falling to zero if the new level of capital is below the bound. To be more specific, the social planner chooses consumption, C_t, labour effort, L_t, and capital, K_t, to maximise:

$$\mathbb{E}_t \sum_{k=0}^{\infty} \beta^k \left[\log C_{t+k} - \frac{L_{t+k}^{1+\nu}}{1+\nu} \right],$$

subject to the capital constraint $K_t \geq \theta K_{t-1}$, and to the budget constraint,

$$C_t + K_t = Y_t = A_t K_{t-1}^{\alpha} L_t^{1-\alpha},$$

where productivity, A_t, evolves according to $A_t = A_{t-1}^{\rho} \exp \varepsilon_t$, where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$.

The first order conditions of the model imply that $(1-\alpha)\frac{Y_t}{C_t} = L_t^{1+\nu}$ and that:

$$\frac{1}{C_t} - \lambda_t = \alpha\beta\mathbb{E}_t \frac{Y_{t+1}}{C_{t+1}K_t} - \beta\theta\mathbb{E}_t\lambda_{t+1},$$

where λ_t is the KKT-multiplier on the borrowing constraint. To convert the KKT conditions into a form amenable to simulation by our algorithm, note that the bound implies that $\frac{1}{C_t} = \frac{1}{Y_t - K_t} \geq \frac{1}{Y_t - \theta K_{t-1}}$, and the positivity of λ_t implies that $\frac{1}{C_t} \geq \alpha\beta\mathbb{E}_t \frac{Y_{t+1}}{C_{t+1}K_t} - \beta\theta\mathbb{E}_t\lambda_{t+1}$. Furthermore, by the KKT conditions, at least one of these constraints always binds. Hence,

$$\frac{1}{C_t} = \max\left\{ \frac{1}{Y_t - \theta K_{t-1}}, \alpha\beta\mathbb{E}_t \frac{Y_{t+1}}{C_{t+1}K_t} - \beta\theta\mathbb{E}_t\lambda_{t+1} \right\}.$$

In the following, we set $\alpha = 0.3$, $\beta = 0.99$, $\nu = 2$, $\theta = 0.99$, $\rho = 0.95$ and $\sigma = 0.01$.

In order to have a comparator for accuracy tests, we first solve the model globally to a high degree of accuracy, using value function iteration on a fine grid. Full details of the global solution procedure are given in appendix F, where we also plot the value and policy functions for the problem. Note, that as $K_{t-1} \to \infty$, in order to avoid violating the constraint, the solution must feature $L_t \to \infty$. Thus, whereas in the original model without bound, the value function is monotonic increasing in capital, in the model with bound, the value function is decreasing in capital for large enough capital levels. This significantly changes behaviour when away from the bound, as the planner will increase consumption now to avoid having too much capital in future.

To assess the extent to which integrating over future uncertainty captures these effects, we compare the value of consumption implied by the global solution procedure to that implied by our solution procedure with different values for S (the number of periods of future uncertainty considered). In particular, we again report errors along simulated paths of length 1000, after discarding an initial 100 periods of burn-in, where now our error measure is the difference between the value of $\log C_t$ implied by our algorithm at order 2,[17] and that implied by the global procedure.

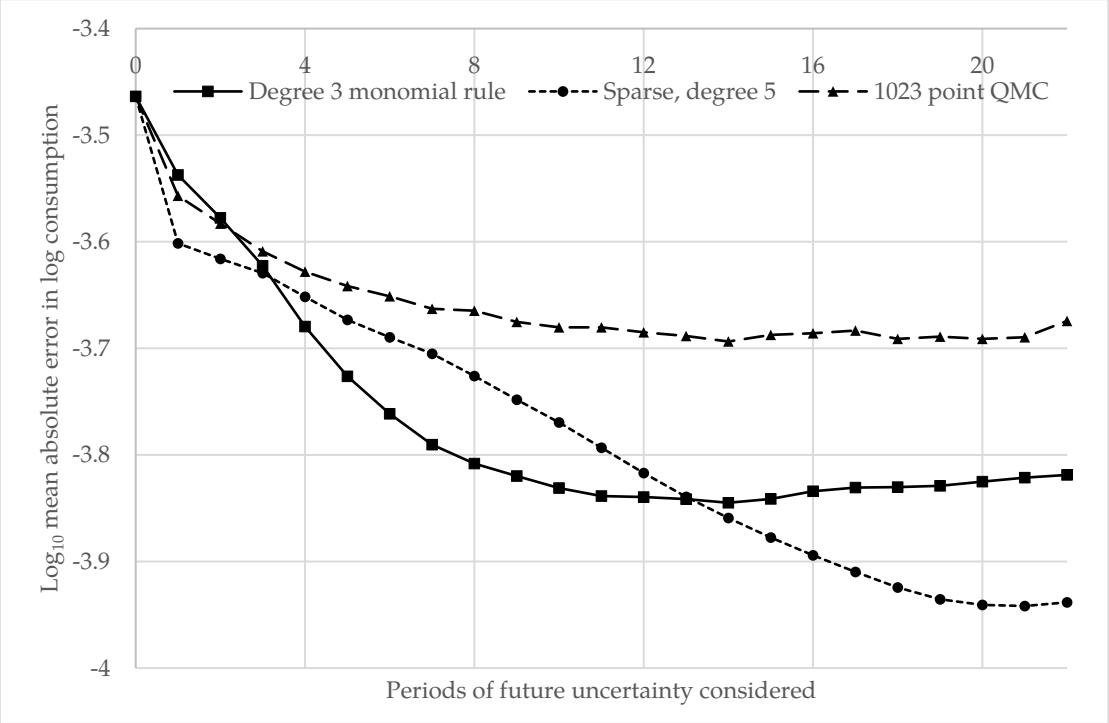

Figure 1: Effect on accuracy of increasing the number of periods of uncertainty considered

The $\log_{10}$ mean absolute values of these errors over the simulation path are shown in Figure 1 for the fastest, median and slowest integration rules used for the previous model, with $S = 0, \ldots, 22$. It may be seen that integrating over future uncertainty produces reasonable accuracy gains in this model, with accuracy initially increasing as more periods of future uncertainty are considered. Accuracy eventually plateaus out for all integration rules as other sources of inaccuracy come to dominate, such as the limited order of the perturbation approximation, the imperfections of the integration, and the fact that the original stochastic extended path algorithm is only an approximation to full rationality. Nonetheless, with ten or more periods of uncertainty, both the monomial and sparse integration rules more than double the accuracy of the algorithm without cubature.

[17] While the model without bounds may be exactly simulated via log-linearization, the bound introduces additional non-log-linearities which means there are substantial gains from higher order approximations.

For some idea of the relative speeds of these integration rules, we note that with 22 periods of uncertainty considered, solving and simulating 1100 periods with the monomial rule required a total of 2778 seconds, doing so with the sparse rule took a total of 4361 seconds, and doing the same with the quasi-Monte Carlo rule completed in a total of 13350 seconds.[18] All of these times are fast enough that they are unlikely to cause any real inconvenience to users of the algorithm, and of course running times were much faster with fewer periods of uncertainty. We also stress again that in larger models, the user can still expect similar running times, as the cost of our algorithm is related to the number of periods at the bound, rather than directly to the number of state variables.

Surprisingly, for this model, it seems that the faster, monomial rule actually produces the most accurate results of the three approaches, at medium horizons, in terms of mean absolute error.[19] This is perhaps due to the fact that the monomial rule places more weight on the tails of the distribution, better capturing the area in which the model is at the bound. These results suggest that this fast monomial rule may often have adequate performance in practice, removing the need to ever use slower integration rules such as quasi-Monte Carlo.

4.4. A model which is otherwise linear

The previous two models featured non-linearities other than the bound, giving an additional source of inaccuracy. We finish by considering a model which is linear other than its bounds, to give the clearest picture of the effects of increasing the integration horizon. We choose one with four bounds, to further illustrate that our method can readily scale up to several bounds.

Suppose the social planner in a small open economy chooses consumption C_t, disposal D_t and bond holdings B_t, to maximise:

$$\mathbb{E}_t \sum_{k=0}^{\infty} \beta^k \left[-\frac{1}{2}(1 - C_t)^2 - \frac{\phi}{2} B_t^2 \right],$$

subject to the budget constraint:

$$C_t + D_t + B_t - RB_{t-1} = Y_t = \max\{\underline{Y}, A_t\},$$

the positivity constraints:

$$0 \leq C_t, \qquad 0 \leq D_t,$$

and the certain repayment of interest constraint:

$$\forall k \in \mathbb{N}^+, \qquad \Pr_t\big((R-1)B_t \leq Y_{t+k}\big) = 1.$$

Preferences here feature a cost of holding large positive or negative asset positions, as is common in the open economy literature (see e.g. Schmitt-Grohé and Uribe 2003), which will ensure that agents cannot accumulate enough savings to stay at $C_t = 1$

[18] These are again wall times, from a run on the hardware listed in footnote 15.

[19] The sparse and quasi Monte Carlo rules have very slightly lower maximum absolute error at long horizons, however.

forever, with probability one. The production function incorporates a back-stop technology (e.g. agriculture), to ensure that income is strictly positive with probability one.

We suppose that productivity evolves according to:

$$A_t = (1-\rho)\mu + \rho A_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim \text{NIID}(0, \sigma^2)$, with $\sigma > 0$. Hence, there is a positive probability that $A_t < \underline{Y}$, regardless of the value of A_{t-1}. Consequently, the certain repayment of interest condition implies that:

$$B_t \geq -\frac{\underline{Y}}{R-1}.$$

Writing $\lambda_{C,t}$ and $\lambda_{D,t}$ for the Lagrange multipliers on the positivity constraints on consumption and disposal, respectively, and $\lambda_{Y,t}$ for the Lagrange multiplier on the budget constraint, the first order conditions of the problem imply that $\lambda_{D,t} = \lambda_{Y,t}$, $\lambda_{C,t} = \lambda_{Y,t} - (1 - C_t)$, and that $\phi B_t + \lambda_{Y,t} = (R-1)\lambda_{B,t} + \beta R \mathbb{E}_t \lambda_{Y,t+1}$. Now, if $C_t > 0$, then $\lambda_{C,t} = 0$, so $C_t = 1 - \lambda_{Y,t}$. Hence, $C_t = \max\{0, 1 - \lambda_{Y,t}\}$. Similarly, if $B_t > -\frac{\underline{Y}}{R-1}$, then $\lambda_{B,t} = 0$, so $B_t = \frac{1}{\phi}[\beta R \mathbb{E}_t \lambda_{Y,t+1} - \lambda_{Y,t}]$. Hence:

$$B_t = \max\left\{-\frac{\underline{Y}}{R-1}, \frac{1}{\phi}[\beta R \mathbb{E}_t \lambda_{Y,t+1} - \lambda_{Y,t}]\right\}.$$

For convenience, we define:

$$X_t := C_t + D_t.$$

Then, from the complementary slackness condition of D_t and $\lambda_{D,t}$:

$$0 = \min\{D_t, \lambda_{D,t}\} = \min\{X_t - C_t, 1 - C_t + \lambda_{C,t}\}.$$

If $C_t = 0$, then we have that $0 = \min\{X_t, 1 + \lambda_{C,t}\} = X_t$, so $C_t = 0 = \min\{X_t, 1\}$, and if $C_t > 0$, then $\lambda_{C,t} = 0$, so we have that $0 = \min\{X_t - C_t, 1 - C_t\}$, implying that $C_t = \min\{X_t, 1\}$ in this case too. Hence, in either case we can rewrite the C_t equation as:

$$\min\{1, X_t\} = C_t = \max\{0, 1 - \lambda_{Y,t}\}.$$

Finally, the budget constraint may be rewritten in terms of X_t as:

$$X_t + B_t - RB_{t-1} = Y_t = \max\{\underline{Y}, A_t\}.$$

This gives a system of equations in B_t, X_t, $\lambda_{Y,t}$ and A_t involving four occasionally binding constraints. In the following, we set $\beta = 0.99$, $\mu = 0.5$, $\rho = 0.95$, $\sigma = 0.05$, $\underline{Y} = 0.25$, $R = \beta^{-1}$ and $\phi = R - 1$.[20]

As before, we first solve the model globally to a high degree of accuracy, using value function iteration on a fine grid. Full details of the global solution procedure are given in appendix G, where we again plot the value and policy functions for the problem. To assess the improvement in accuracy from integrating over future uncertainty, we compare the value of X_t implied by the global solution procedure to that implied by our solution procedure at order 1, with different values for S (the number of periods

[20] $R = \beta^{-1}$ implies that $B_t = 0$ in steady-state. $\phi = R - 1$ simplifies some expressions for the solution in the absence of bounds.

of future uncertainty considered), along simulated paths of length 1000, after discarding an initial 100 periods of burn-in.

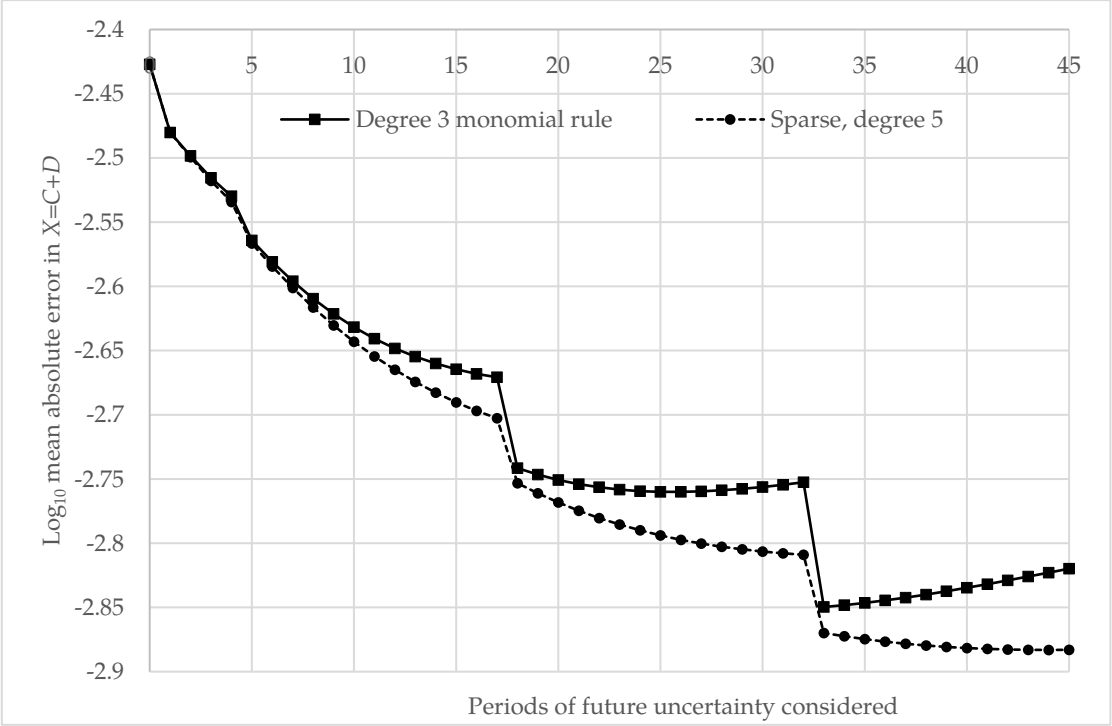

Figure 2: Effect on accuracy of increasing the number of periods of uncertainty considered

The $\log_{10}$ maximum absolute values of these errors over the simulation path are shown in Figure 2 for the monomial and sparse integration rules, with $S = 0, \ldots, 45$. Since there are no other sources of inaccuracy here, integrating over future uncertainty produces substantial accuracy gains, with accuracy nearly three times higher with large numbers of periods of considered uncertainty, compared to the case without cubature. The eventual plateau in accuracy is driven by the imperfections of our integration rules and the fact that the original stochastic extended path algorithm is only an approximation to full rationality. With unbounded computer power, we could run quasi-Monte Carlo integration with (e.g.) $2^{2S+1} - 1$ points, and the result would give an arbitrarily accurate approximation to the integral with high enough S. In practice, this is unnecessary however, as our results here suggest that there are already very substantial returns to integrating over future uncertainty using integration rules with far fewer points, and even extremely accurate integration would not solve the fact that the original stochastic extended path algorithm is only an approximation to full rationality.

Due to the presence of four bounds in this model, the LCPs which are solved each period are larger than those in a model with a single bound. Furthermore, the structure and parameters of the model mean that the model spends long runs at the

bound, further increasing the computational cost. However, despite this, the running times are still fast enough to be practical. With 45 periods of considered uncertainty, if the monomial integration rule is used, it completes a simulation of 1100 periods in a total of 1383 seconds, and with the sparse rule, it completes in a total of 2855 seconds.[21] This illustrates that our algorithm readily scales up to models with multiple bounds.

5. Conclusion

This paper has presented the first algorithm for the perfect foresight solution of otherwise linear models with occasionally binding constraints that always completes in finite time, and which is able to detect when the model has no solution. The paper went on to exploit this underlying solver to produce a robust, accurate and scalable simulation algorithm for general nonlinear models with occasionally binding constraints that accounted for future uncertainty. Code implementing all of the algorithms discussed here is contained in the author's "DynareOBC" toolkit which augments the abilities of Dynare (Adjemian et al. 2011) with the ability to solve models with OBCs.

We also derived results on the computational complexity of the problem which suggest that for large models, there is unlikely to be an alternative computational approach which is significantly faster. These computational complexity results also imply that when the model does not always have a unique solution (i.e. M is not a P-matrix, in the language of Holden (2016)), then computing expectations is a computationally difficult problem. Given that Holden (2016) showed that in standard New Keynesian models with a ZLB, under inflation targeting there is not always a unique solution, though there is under price-level targeting, then this implies that computing expectations is likely to be much easier for agents under price-level targeting than under inflation targeting. Given that there are often substantial welfare costs to failures in forming expectations rationally, this provides further arguments in favour of price level targeting to those contained in Holden (2016).

6. References

Adjemian, Stéphane, Houtan Bastani, Fréderic Karamé, Michel Juillard, Junior Maih, Ferhat Mihoubi, George Perendia, Johannes Pfeifer, Marco Ratto, and Sébastien Villemot. 2011. *Dynare: Reference Manual Version 4*. CEPREMAP.

Adjemian, Stéphane, and Michel Juillard. 2013. 'Stochastic Extended Path Approach'. CEIS Tor Vergata.

———. 2016. 'The Stochastic Extended Path Approach'. National Bank of Belgium.

Beaudry, Paul, and Franck Portier. 2006. 'Stock Prices, News, and Economic Fluctuations'. *American Economic Review* 96 (4): 1293–1307.

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

Bodenstein, Martin, Luca Guerrieri, and Christopher J. Gust. 2013. 'Oil Shocks and the Zero Bound on Nominal Interest Rates'. *Journal of International Money and Finance* 32 (February): 941–967.

[21] Once more, this is wall time, using the hardware given in footnote 15.

Brock, William A, and Leonard J Mirman. 1972. 'Optimal Economic Growth and Uncertainty: The Discounted Case'. *Journal of Economic Theory* 4 (3) (June): 479–513.

Chung, S.J. 1989. 'NP-Completeness of the Linear Complementarity Problem'. *Journal of Optimization Theory and Applications* 60 (3) (March 1): 393–399.

Cools, Ronald. 2003. 'An Encyclopaedia of Cubature Formulas'. *Oberwolfach Special Issue* 19 (3) (June): 445–453.

Cottle, Richard W. 2009. 'Linear Complementarity Problem'. In *Encyclopedia of Optimization*, edited by Christodoulos A. Floudas and Panos M. Pardalos, 1873–1878. Springer US.

Coxson, Gregory E. 1994. 'The P-Matrix Problem Is Co-NP-Complete'. *Mathematical Programming* 64 (1–3) (March 1): 173–178.

Csizmadia, Zsolt, and Tibor Illés. 2006. 'New Criss-Cross Type Algorithms for Linear Complementarity Problems with Sufficient Matrices'. *Optimization Methods and Software* 21 (2) (April 1): 247–266.

Fair, Ray C, and John B Taylor. 1983. 'Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models'. *Econometrica* 51 (4) (July): 1169–85.

Fukuda, Komei. 2015. 'Toward a Strongly Polynomial Algorithm for the Linear Complementarity Problems with Sufficient Matrices'. *Unpublished* (September 22).

Genz, Alan, and Bradley D. Keister. 1996. 'Fully Symmetric Interpolatory Rules for Multiple Integrals over Infinite Regions with Gaussian Weight'. *Journal of Computational and Applied Mathematics* 71 (2): 299–309.

Guerrieri, Luca, and Matteo Iacoviello. 2015. 'OccBin: A Toolkit for Solving Dynamic Models with Occasionally Binding Constraints Easily'. *Journal of Monetary Economics* 70 (March): 22–38.

Harris, F.J. 1978. 'On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform'. *Proceedings of the IEEE* 66 (1) (January): 51–83.

Hebden, James, Jesper Lindé, and Lars Svensson. 2011. 'Optimal Monetary Policy in the Hybrid New Keynesian Model under the Zero Lower Bound'. Unpublished.

Herceg, Martin, Michal Kvasnica, Colin N. Jones, and Manfred Morari. 2013. 'Multi-Parametric Toolbox 3.0'. In *Control Conference (ECC), 2013 European*, 502–510. IEEE.

Holden, Tom. 2010. *Products, Patents and Productivity Persistence: A DSGE Model of Endogenous Growth*. Economics Series Working Paper. University of Oxford, Department of Economics.

———. 2016. 'Existence and Uniqueness of Solutions to Dynamic Models with Occasionally Binding Constraints.' Unpublished.

Holden, Tom, and Michael Paetz. 2012. 'Efficient Simulation of DSGE Models with Inequality Constraints'. School of Economics Discussion Papers.

Illés, Tibor, Marianna Nagy, and Tamás Terlaky. 2010. 'A Polynomial Path-Following Interior Point Algorithm for General Linear Complementarity Problems'. *Journal of Global Optimization* 47 (3) (July 1): 329–342.

Jin, Hehui, and Kenneth L. Judd. 2002. 'Perturbation Methods for General Dynamic Stochastic Models'.

Jones, C.N., and M. Morrari. 2006. 'Multiparametric Linear Complementarity Problems'. In *Decision and Control, 2006 45th IEEE Conference on*, 5687–5692.

Judd, Kenneth L., Lilia Maliar, and Serguei Maliar. 2012. *Merging Simulation and Projection Approaches to Solve High-Dimensional Problems*. National Bureau of Economic Research, Inc.

Judd, Kenneth L., and Ben Skrainka. 2011. *High Performance Quadrature Rules: How Numerical Integration Affects a Popular Model of Product Differentiation*. Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Karp, Richard M. 1972. 'Reducibility among Combinatorial Problems'. In *Complexity of Computer Computations*, edited by Raymond E. Miller, James W. Thatcher, and Jean D. Bohlinger, 85–103. The IBM Research Symposia Series. Springer US.

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

Kojima, M., N. Megiddo, T. Noma, and A. Yoshise. 1991. *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*. Springer.

Kojima, M., Shinji Mizuno, and Akiko Yoshise. 1989. 'A Polynomial-Time Algorithm for a Class of Linear Complementarity Problems'. *Mathematical Programming* 44 (1–3) (May 1): 1–26.

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

———. 2013b. *Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations*. SFB 649 Discussion Papers. Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Löfberg, J. 2004. 'YALMIP : A Toolbox for Modeling and Optimization in MATLAB'. In *Proceedings of the CACSD Conference*. Taipei, Taiwan.

Magnus, Jan R., and H. Neudecker. 1979. 'The Commutation Matrix: Some Properties and Applications'. *Ann. Statist.* (2) (March): 381–394.

Pardalos, P. M., and J. B. Rosen. 1988. 'Global Optimization Approach to the Linear Complementarity Problem'. *SIAM J. Sci. Stat. Comput.* 9 (2): 341–354.

Rosen, J. B. 1990. 'Minimum Norm Solution to the Linear Complementarity Problem'. *Functional Analysis, Optimization and Economics*: 208–216.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2003. 'Closing Small Open Economy Models'. *Journal of International Economics* 61 (1) (October): 163–185.

Sims, ChristopherA. 2002. 'Solving Linear Rational Expectations Models'. *Computational Economics* 20 (1–2) (October 1): 1–20.

Sobol, Ilya M. 1967. 'On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals'. *USSR Computational Mathematics and Mathematical Physics* (7): 86–112.

Tsatsomeros, Michael J., and Lei Li. 2000. 'A Recursive Test for P-Matrices'. *BIT Numerical Mathematics* 40 (2) (June 1): 410–414.

Tseng, Paul. 2000. 'Co-NP-Completeness of Some Matrix Classification Problems'. *Mathematical Programming* 88 (1) (June 1): 183–192.

Online Appendices to: "Computation of solutions to dynamic models with occasionally binding constraints."

Tom D. Holden, School of Economics, University of Surrey

A. Construction of a model with arbitrary *M* matrix

Let $\mathcal{M} \in \mathbb{R}^{T\times T}$. Consider a model with the following equations:

$$a_t = \max\{0, b_t\},$$
$$a_t = 1 + \sum_{j=1}^{T}\sum_{k=1}^{T} \mathcal{M}_{j,k}(c_{j-1,k-1,t} - c_{j,k,t}) + d_{0,t},$$
$$c_{0,0,t} = a_t - b_t,$$
$$c_{0,k,t} = \mathbb{E}_t c_{0,k-1,t+1}, \qquad \forall k \in \{1,\dots,T\},$$
$$c_{j,k,t} = c_{j-1,k,t-1}, \qquad \forall j \in \{1,\dots,T\}, k \in \{0,\dots,T\},$$
$$d_{k,t} = d_{k+1,t-1}, \qquad \forall k \in \{0,\dots,T-1\},$$
$$d_{T,t} = 0$$

with steady-state $a_\cdot = b_\cdot = 1$, $c_{j,k,\cdot} = 0$, $d_{k,\cdot} = 0$ for all $j,k,\in \{0,\dots,T\}$. Defining:

$$x_t := [a_t \quad b_t \quad (\operatorname{vec} c_{\cdot,\cdot,t})' \quad d_{\cdot,t}']'$$

and dropping expectations, this model is then in the form of Problem 2.

Now consider the model's Problem 3 type equivalent, in which for $t \in \mathbb{N}^+$:

$$a_t = \begin{cases} b_t + y_{t,0} & \text{if } t \leq T \\ b_t & \text{if } t > T' \end{cases}$$

where $y_{\cdot,\cdot}$ is defined as in Problem 3. Thus, if $c_{j,k,0} = 0$ and $d_{k,0} = 0$ for all $j,k \in \{0,\dots,T\}$, then for all $t \in \mathbb{N}^+$, $j,k \in \{0,\dots,T\}$:

$$c_{0,k,t} = \begin{cases} y_{t+k,0} & \text{if } t+k \leq T \\ 0 & \text{if } t+k > 0' \end{cases}$$
$$c_{j,k,t} = \begin{cases} c_{0,k,t-j} & \text{if } t-j > 0 \\ 0 & \text{if } t-j \leq 0 \end{cases} = \begin{cases} y_{t+k-j,0} & \text{if } t-j > 0, t+k-j \leq T \\ 0 & \text{otherwise} \end{cases}.$$

Hence, for all $t \in \mathbb{N}^+$, $j,k \in \{1,\dots,T\}$:

$$c_{j-1,k-1,t} - c_{j,k,t} = \begin{cases} y_{t+k-j,0} & \text{if } t-j = 0, t+k-j \leq T \\ 0 & \text{otherwise} \end{cases} = \begin{cases} y_{k,0} & \text{if } t = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, for all $t \in \{1,\dots,T\}$:

$$a_t - 1 = \sum_{k=1}^{T} \mathcal{M}_{t,k} y_{k,0}.$$

Consequently, if $y_{k,0} = I_{\cdot,l}$ for some $l \in \{1,\dots,T\}$, then $a_t - 1 = \mathcal{M}_{t,l}$ (i.e. the relative impulse response to a news-shock at horizon l) is the l^{th} column of $\mathcal{M}$.

Finally, note that in the model's Problem 1 equivalent, if $c_{j,k,0} = 0$ for all $j,k \in \{0,\dots,T\}$, then for all $t \in \mathbb{N}^+$, $a_t = b_t = d_{0,t} = d_{t,0}$. Hence, if $d_{\cdot,0} = q$ for some $q \in \mathbb{R}^T$, then $q = q$ for this model.

B. Proof that the existence of a polynomial time approximate solution would imply P=NP

Suppose that we have some solution procedure which accepts problems in the form of Problem 2 (and possibly other problems) together with a radius κ and an accuracy level ϵ, and, in time polynomial in n, returns a set $\mathcal{D}_{\kappa,\epsilon} \subseteq \mathbb{R}^n$ membership of which may be evaluated in time polynomial in n, and a policy function $p_{\kappa,\epsilon}: \mathcal{D}_{\kappa,\epsilon} \subseteq \mathbb{R}^n \to \mathbb{R}^n$ that may be evaluated in time polynomial in n and that satisfies:

$$\left\| p^*(x) - p_{\kappa,\epsilon}(x) \right\|_\infty < \epsilon,$$

for all $x \in \mathcal{D}^* \subseteq \mathcal{D}_\epsilon$ with $\|x - \mu\|_\infty < \kappa$, and where $p^*: \mathcal{D}^* \to \mathcal{D}^*$ is an exact policy function, i.e. a function satisfying:

1. For all $x \in \mathcal{D}^*$:

$$x_1 = \max\{0, I_{1,\cdot}\mu + A_{1,\cdot}(x - \mu) + (B_{1,\cdot} + I_{1,\cdot})(p^*(x) - \mu) + C_{1,\cdot}(p^*(p^*(x)) - \mu)\},$$
$$(A_{-1,\cdot} + B_{-1,\cdot} + C_{-1,\cdot})\mu = A_{-1,\cdot}x + B_{-1,\cdot}p^*(x) + C_{-1,\cdot}p^*(p^*(x)).$$

2. For all $x_0 \in \mathcal{D}^*$, if $x_t = p^*(x_{t-1})$ for all $t \in \mathbb{N}^+$, then $x_t \to \mu$ as $t \to \infty$.
3. For all $x_0 \notin \mathcal{D}^*$, there is no $x_1, x_2, \ldots \in \mathbb{R}^n$ which solve this instance of Problem 2.

Now consider the following "knapsack"-type problem from Chung (1989):

Problem 7 Suppose $a_1, \ldots, a_{T-2}, b \in \mathbb{N}^+$ are given. Find $z_1, \ldots, z_{T-2} \in \{0,1\}$ such that $\sum_{j=1}^{T-2} a_j z_j = b$.

Chung (1989) shows that there exists $q \in \mathbb{Z}^T$, $\mathcal{M} \in \mathbb{Z}^{T \times T}$, such that Problem 7 has a solution if and only if the LCP $(q, \mathcal{M})$ has a solution, where q and $\mathcal{M}$ may be computed from $a_1, \ldots, a_{T-2}$ and b in time polynomial in T. Furthermore, the details of the proof in Chung (1989) reveal that for any $y \in \mathbb{R}^T$ that solves the LCP $(q, \mathcal{M})$, $y_{T-1} = y_T = 0$, and setting $z_t = \frac{y_t}{a_t}$ for $t \in \{1, \ldots, T-2\}$ solves Problem 7. Since $a_1, \ldots, a_{T-2} \in \mathbb{N}^+$ and $z_1, \ldots, z_{T-2} \in \{0,1\}$, this implies that $y_t \in \{0, a_t\} \subseteq \mathbb{N}$ for $t \in \{1, \ldots, T-2\}$. Moreover, by Proposition 2, given $a_1, \ldots, a_{T-2}$ and b we can thus construct a model in the form of Problem 2 in polynomial time in T, featuring polynomial in T state variables, and such that for an appropriately chosen initial state, setting $z_t = \frac{x_{1,t}}{a_t}$ for $t \in \{1, \ldots, T-2\}$ solves Problem 7 if it has a solution (where $x_{1,t}$ gives the path of the bounded variable in the constructed model). Additionally, by inspecting the proof of Proposition 2 from appendix A, we see that since $q \in \mathbb{Z}^T$, $\mathcal{M} \in \mathbb{Z}^{T \times T}$, and $y_t \in \mathbb{N}$ for all $t \in \{1, \ldots, T-2\}$, it must be the case that for all $t \in \mathbb{N}$, $x_t \in \mathbb{Z}^n$ (for some $n \in \mathbb{N}$), so the model is always integer valued.

To complete the proof, we set $\epsilon := \frac{1}{2}$, $\kappa := \|x_0 - \mu\|_\infty + 1$ and construct the policy function $p_{\kappa,\epsilon}$ for the constructed model. By assumption, we can do this in time polynomial in T. Using this we can construct an exact solution for z_t as follows. Set $\hat{x}_0 := x_0$. Now suppose we have defined $\hat{x}_{t-1}$ for some $t \in \mathbb{N}^+$. We first test if $\hat{x}_{t-1} \in$

$\mathcal{D}_{\kappa,\epsilon}$. If it is not, then we terminate the procedure with a "no solution" message. Otherwise, we set each element of $\hat{x}_t$ to be equal to the nearest integer to the corresponding element of $p_{\kappa,\epsilon}(\hat{x}_{t-1})$. We then test if $\|\hat{x}_t - \mu\| < \kappa$. If it is, we proceed, otherwise, we redefine $\kappa := \|\hat{x}_t - \mu\|_\infty + 1$, and restart. After an amount of time bounded by a polynomial in T, we will have either terminated with a "no solution" message, or have successfully defined $\hat{x}_t$ for $t \in \{0, \dots, T-2\}$, where for all $t \in \{0, \dots, T-2\}$, we will have found some κ for which $\|\hat{x}_t - \mu\| < \kappa$. In this case, we then define $\hat{z}_t = \frac{\hat{x}_{1,t}}{a_t}$ for $t \in \{1, \dots, T-2\}$ and test whether it solves Problem 7. If it does, report a "solution found" message, otherwise, report a "no solution" message. We now prove that this procedure works.

We first prove that if $x_0 \in \mathcal{D}^* \cap \mathbb{Z}^n$ then $\hat{x}_t = x_t \in \mathcal{D}^* \cap \mathbb{Z}^n$ for all $t \in \mathbb{N}$, by induction on t. The inductive base case is trivial. Suppose for the inductive step that for some $t \in \mathbb{N}^+$, $\hat{x}_{t-1} = x_{t-1} \in \mathcal{D}^* \cap \mathbb{Z}^n$. Then $\hat{x}_{t-1} \in \mathcal{D}_{\kappa,\epsilon}$, so $\hat{x}_t$ is defined and is equal to the (elementwise) nearest integer to $p_{\kappa,\epsilon}(\hat{x}_{t-1})$. Then since as $\hat{x}_{t-1} \in \mathcal{D}^*$ and $\|\hat{x}_{t-1} - \mu\| < \kappa$, $\left\|p^*(\hat{x}_{t-1}) - p_{\kappa,\epsilon}(\hat{x}_{t-1})\right\|_\infty < \frac{1}{2}$, and $p^*: \mathcal{D}^* \cap \mathbb{Z}^n \to \mathcal{D}^* \cap \mathbb{Z}^n$, it must be the case that $\hat{x}_t = p^*(\hat{x}_{t-1}) \in \mathcal{D}^* \cap \mathbb{Z}^n$. This establishes the inductive hypothesis, and hence if we set $\hat{z}_t = \frac{\hat{x}_{1,t}}{a_t}$ for $t \in \{1, \dots, T-2\}$ then test whether it solves Problem 7, we will find that it does, giving a solution, after only polynomial in T calculations.

Now suppose that $x_0 \in \mathbb{Z}^n$, but $x_0 \notin \mathcal{D}^*$. There are two possibilities. Either we will find some $t \in \{1, \dots, T-1\}$ for which $\hat{x}_t \notin \mathcal{D}_\epsilon$, and hence $\hat{x}_t \notin \mathcal{D}^*$, or we will successfully calculate $\hat{x}_{1,t}$ for $t \in \{1, \dots, T-2\}$. By the previous result, in the former case we will have found in polynomial time in T a proof that $x_0 \notin \mathcal{D}^*$, and hence that there is no solution to Problem 7. In the latter case, we can again set $\hat{z}_t = \frac{\hat{x}_{1,t}}{a_t}$ for $t \in \{1, \dots, T-2\}$ and test whether it solves Problem 7, and we will find (in polynomial in T calculations) that it does not, giving an alternative polynomial time in T proof that $x_0 \notin \mathcal{D}^*$, and hence that there is no solution to Problem 7.

We have thus established that the procedure described (calculating $\hat{x}_1, \dots, \hat{x}_{T-2}$ and then testing whether $\hat{z}_t = \frac{\hat{x}_{1,t}}{a_t}$ solves Problem 7), enables us to answer the question of whether Problem 7 has a solution in an amount of time that is polynomial in T. But Problem 7 is NP-complete (Karp 1972), and hence this implies that P=NP.

C. Special cases with polynomial time solutions

Polynomial time algorithms exist for the LCP if M is general positive semi-definite (Kojima, Mizuno, and Yoshise 1989). However, it appears that M is general positive semi-definite in only very few macroeconomic models, so this is of minimal relevance. Furthermore, if either condition 1 or condition 2 of Proposition 5 of Holden (2016) is known to be satisfied, then we can find out if a solution exists in polynomial time, by solving the feasibility problem. Moreover, a polynomial time algorithm exists (Illés,

Nagy, and Terlaky 2010) which will give a certificate that one of the following is true, for a given q and real number $\bar{\kappa} \geq 0$:

- For any $\kappa \leq \bar{\kappa}$, M is not a member of the matrix class $P^*(\kappa)$, defined in the paper. (*Note that for $\kappa_1 < \kappa_2$, $P^*(\kappa_1) \subseteq P^*(\kappa_2)$, and that the class of sufficient matrices is the union of the classes of $P^*(\kappa)$ matrices for all $\kappa \geq 0$.)*
- The LCP (q, M) has no solution.
- The LCP (q, M) has the solution y.

Thus for "most" sufficient matrices we can find a solution (or a certificate that there is none), in polynomial time. It has been conjectured that in fact this holds for all sufficient matrices (Fukuda 2015).

Unfortunately, no algorithm is known for finding out if M is sufficient in polynomial time. Indeed, it has also been shown (Coxson 1994; Tseng 2000) that it is "co-NP complete" to test if M is non-degenerate, a P-matrix, a P_0-matrix, semi-monotone, strictly semi-monotone, column sufficient or row sufficient, where all of these classes are as defined in Holden (2016). This means that were a polynomial time (in T) algorithm available for these things then we would have a proof that P=NP.

D. The augmented state-space representation of a pruned perturbation solution

We seek to convert the model into the form:

$$z_t = o + P\tilde{z}_{t-1} + Q\xi_t,$$
$$x_t = u + Vz_t,$$

where $\mathbb{E}_{t-1}\xi_t = 0$, and where throughout, $\tilde{}$s over variables denote the subset of state variables. We proceed by taking each order of approximation in turn. We assume that the original model has l state variables. Of the assorted algorithms available for pruning, it appears that Lan and Meyer-Gohde's (2013a) algorithm is the most accurate (Lan and Meyer-Gohde 2013b), and so both the discussion below, and the implementation in DynareOBC is based on this approach, however, everything we say would also go through with alternative pruning algorithms.

Order 1 At order 1:

$$x_t^{(1)} = \alpha\tilde{x}_{t-1}^{(1)} + \beta_0\varepsilon_t,$$
$$x_t = \mu_x + x_t^{(1)},$$

so if we define:

$$z_t := x_t^{(1)}, \qquad \tilde{z}_t := \tilde{x}_t^{(1)}, \qquad o := 0, \qquad P := \alpha, \qquad Q := \beta_0, \qquad \xi_t := \varepsilon_t, \qquad u := \mu_x,$$
$$V := I_n,$$

then we are done.

Order 2 At order 2:

$$x_t^{(1)} = \alpha \tilde{x}_{t-1}^{(1)} + \beta_0 \varepsilon_t,$$
$$x_t^{(2)} = \alpha \tilde{x}_{t-1}^{(2)} + \frac{1}{2}\beta_{22}\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \beta_{20}\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) + \frac{1}{2}\beta_{00}(\varepsilon_t \otimes \varepsilon_t),$$
$$x_t = \mu_x + x^{(0)} + x_t^{(1)} + x_t^{(2)},$$

for some constant $x^{(0)}$.

Now, note that:

$$\begin{aligned}
\tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} &= \left(\tilde{\alpha}\tilde{x}_{t-1}^{(1)} + \tilde{\beta}_0\varepsilon_t\right) \otimes \left(\tilde{\alpha}\tilde{x}_{t-1}^{(1)} + \tilde{\beta}_0\varepsilon_t\right) \\
&= \tilde{\alpha}\tilde{x}_{t-1}^{(1)} \otimes \tilde{\alpha}\tilde{x}_{t-1}^{(1)} + \tilde{\alpha}\tilde{x}_{t-1}^{(1)} \otimes \tilde{\beta}_0\varepsilon_t + \tilde{\beta}_0\varepsilon_t \otimes \tilde{\alpha}\tilde{x}_{t-1}^{(1)} + \tilde{\beta}_0\varepsilon_t \otimes \tilde{\beta}_0\varepsilon_t \\
&= (\tilde{\alpha} \otimes \tilde{\alpha})\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \left(\tilde{\alpha} \otimes \tilde{\beta}_0\right)\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) + \left(\tilde{\beta}_0 \otimes \tilde{\alpha}\right)\left(\varepsilon_t \otimes \tilde{x}_{t-1}^{(1)}\right) \\
&\quad + \left(\tilde{\beta}_0 \otimes \tilde{\beta}_0\right)(\varepsilon_t \otimes \varepsilon_t) \\
&= (\tilde{\alpha} \otimes \tilde{\alpha})\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \left(\left(\tilde{\alpha} \otimes \tilde{\beta}_0\right) + \left(\tilde{\beta}_0 \otimes \tilde{\alpha}\right)K_{m,l}\right)\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) \\
&\quad + \left(\tilde{\beta}_0 \otimes \tilde{\beta}_0\right)(\varepsilon_t \otimes \varepsilon_t) \\
&= (\tilde{\alpha} \otimes \tilde{\alpha})\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \left(\left(\tilde{\alpha} \otimes \tilde{\beta}_0\right) + K_{l,l}(\tilde{\alpha} \otimes \tilde{\beta}_0)K_{l,m}K_{m,l}\right)\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) \\
&\quad + \left(\tilde{\beta}_0 \otimes \tilde{\beta}_0\right)(\varepsilon_t \otimes \varepsilon_t) \\
&= (\tilde{\alpha} \otimes \tilde{\alpha})\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + (I_{l^2} + K_{l,l})\left(\tilde{\alpha} \otimes \tilde{\beta}_0\right)\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) \\
&\quad + \left(\tilde{\beta}_0 \otimes \tilde{\beta}_0\right)(\varepsilon_t \otimes \varepsilon_t),
\end{aligned}$$

where $K_{l,m} \in \mathbb{R}^{lm \times lm}$ is the commutation matrix for $l \times m$ matrices, i.e. it is the unique matrix such that for all $D \in \mathbb{R}^{l \times m}$, $K_{l,m} \operatorname{vec} D = \operatorname{vec} D'$ (Magnus and Neudecker 1979). Thus, if we define:

$$z_t := \begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ \tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} \end{bmatrix}, \qquad \tilde{z}_t := \begin{bmatrix} \tilde{x}_t^{(1)} \\ \tilde{x}_t^{(2)} \\ \tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} \end{bmatrix},$$

$$P := \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & \frac{1}{2}\beta_{22} \\ 0 & 0 & \tilde{\alpha} \otimes \tilde{\alpha} \end{bmatrix}, \qquad Q := \begin{bmatrix} \beta_0 & 0 & 0 \\ 0 & \frac{1}{2}\beta_{00} & \beta_{20} \\ 0 & \tilde{\beta}_0 \otimes \tilde{\beta}_0 & (I_{l^2} + K_{l,l})(\tilde{\alpha} \otimes \tilde{\beta}_0) \end{bmatrix},$$

$$\xi_t := \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes \varepsilon_t - \operatorname{vec}\Sigma \\ \tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t \end{bmatrix}, \qquad o := \begin{bmatrix} 0 \\ \frac{1}{2}\beta_{00}\operatorname{vec}\Sigma \\ (\tilde{\beta}_0 \otimes \tilde{\beta}_0)\operatorname{vec}\Sigma \end{bmatrix},$$

$$u := \mu_x + x^{(0)}, \qquad V := [I \quad I \quad 0],$$

then we are done.

Order 3 At order 3:

$$x_t^{(1)} = \alpha\tilde{x}_{t-1}^{(1)} + \beta_0\varepsilon_t,$$

$$x_t^{(2)} = \alpha\tilde{x}_{t-1}^{(2)} + \frac{1}{2}\beta_{22}\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \beta_{20}\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) + \frac{1}{2}\beta_{00}(\varepsilon_t \otimes \varepsilon_t),$$

$$x_t^{(1,\sigma^2)} = \alpha\tilde{x}_{t-1}^{(1,\sigma^2)} + \frac{1}{2}\beta_{\sigma^2,0}\varepsilon_t + \frac{1}{2}\beta_{\sigma^2,1}\tilde{x}_{t-1}^{(1)},$$

$$\begin{aligned} x_t^{(3)} = \alpha\tilde{x}_{t-1}^{(3)} &+ \frac{1}{6}\beta_{333,1}\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)}\right) + \frac{1}{6}\beta_{000}(\varepsilon_t \otimes \varepsilon_t \otimes \varepsilon_t) \\ &+ \frac{1}{2}\beta_{330,1}\left(\tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t\right) + \frac{1}{2}\beta_{300}\left(\tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t \otimes \varepsilon_t\right) + \beta_{22}\left(\tilde{x}_{t-1}^{(2)} \otimes \tilde{x}_{t-1}^{(1)}\right) \\ &+ \beta_{20}\left(\tilde{x}_{t-1}^{(2)} \otimes \varepsilon_t\right), \end{aligned}$$

$$x_t = \mu_x + x^{(0)} + x_t^{(1)} + x_t^{(2)} + x_t^{(1,\sigma^2)} + x_t^{(3)},$$

for the same constant $x^{(0)}$ as at order 2, providing the shocks have zero skewness (e.g. they are normally distributed). By similar calculations to those at second order, we then have that if we define:

$$z_t^3 := \mathbb{E}_s\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ \tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} \\ x_t^{(1,\sigma^2)} \\ x_t^{(3)} \\ \tilde{x}_t^{(2)} \otimes \tilde{x}_t^{(1)} \\ \tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} \otimes \tilde{x}_t^{(1)} \end{bmatrix},$$

$$P := \begin{bmatrix} \alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & \frac{1}{2}\beta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\alpha} \otimes \tilde{\alpha} & 0 & 0 & 0 & 0 \\ \frac{1}{2}\beta_{\sigma^2,1} & 0 & 0 & \alpha & 0 & 0 & 0 \\ \frac{1}{2}\beta_{300}(I_l \otimes \operatorname{vec}\Sigma) & 0 & 0 & 0 & \alpha & \beta_{22} & \frac{1}{6}\beta_{333,1} \\ \left(\tilde{\beta}_{20} \otimes \tilde{\beta}_0 + \frac{1}{2}K_{l,l}(\tilde{\alpha} \otimes \tilde{\beta}_{00})\right)(I_l \otimes \operatorname{vec}\Sigma) & 0 & 0 & 0 & 0 & \tilde{\alpha} \otimes \tilde{\alpha} & \frac{1}{2}\tilde{\beta}_{22} \otimes \tilde{\alpha} \\ \left((I_{l^2} + K_{l,l}) \otimes I_l + K_{l^2,l}\right)(\tilde{\alpha} \otimes \tilde{\beta}_0 \otimes \tilde{\beta}_0)(I_l \otimes \operatorname{vec}\Sigma) & 0 & 0 & 0 & 0 & 0 & \tilde{\alpha} \otimes \tilde{\alpha} \otimes \tilde{\alpha} \end{bmatrix},$$

$$Q_{11} := \begin{bmatrix} \beta_0 & 0 & 0 \\ 0 & \frac{1}{2}\beta_{00} & \beta_{20} \\ 0 & \tilde{\beta}_0 \otimes \tilde{\beta}_0 & (I_{l^2} + K_{l,l})(\tilde{\alpha} \otimes \tilde{\beta}_0) \\ \frac{1}{2}\beta_{\sigma^2,0} & 0 & 0 \end{bmatrix},$$

$$Q_{22} := \begin{bmatrix} \beta_{20} & \frac{1}{2}\beta_{330,1} & \frac{1}{2}\beta_{300} & \frac{1}{6}\beta_{000} \\ \tilde{\alpha} \otimes \tilde{\beta}_0 & \frac{1}{2}\tilde{\beta}_{22} \otimes \tilde{\beta}_0 + (\tilde{\beta}_{20} \otimes \tilde{\alpha})(I_l \otimes K_{m,l}) & \tilde{\beta}_{20} \otimes \tilde{\beta}_0 + \frac{1}{2}K_{l,l}(\tilde{\alpha} \otimes \tilde{\beta}_{00}) & \frac{1}{2}\tilde{\beta}_{00} \otimes \tilde{\beta}_0 \\ 0 & (I_l \otimes (I_{l^2} + K_{l,l}) + K_{l,l^2})(\tilde{\alpha} \otimes \tilde{\alpha} \otimes \tilde{\beta}_0) & \left((I_{l^2} + K_{l,l}) \otimes I_l + K_{l^2,l}\right)(\tilde{\alpha} \otimes \tilde{\beta}_0 \otimes \tilde{\beta}_0) & \tilde{\beta}_0 \otimes \tilde{\beta}_0 \otimes \tilde{\beta}_0 \end{bmatrix},$$

$$Q := \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix},$$

$$\xi_t := \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \otimes \varepsilon_t - \operatorname{vec}\Sigma \\ \tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t \\ \tilde{x}_{t-1}^{(2)} \otimes \varepsilon_t \\ \tilde{x}_{t-1}^{(1)} \otimes \tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t \\ \tilde{x}_{t-1}^{(1)} \otimes (\varepsilon_t \otimes \varepsilon_t - \operatorname{vec}\Sigma) \\ \varepsilon_t \otimes \varepsilon_t \otimes \varepsilon_t \end{bmatrix}, \qquad o := \begin{bmatrix} 0 \\ \frac{1}{2}\beta_{00}\operatorname{vec}\Sigma \\ (\tilde{\beta}_0 \otimes \tilde{\beta}_0)\operatorname{vec}\Sigma \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$u := \mu_x + x^{(0)}, \qquad V := [I \quad I \quad 0 \quad I \quad I \quad 0 \quad 0],$$

then again we are done.

E. The conditional covariance of future variables under a pruned perturbation solution

First, suppose that:

$$z_t = o + Pz_{t-1} + Q\xi_t$$

where $\mathbb{E}_t\xi_{t+k} = 0$ for $k > 0$. Then:

$$\mathbb{E}_t z_{t+k} = \sum_{j=0}^{k-1} P^j o + P^k z_t,$$

so:

$$z_{t+k} - \mathbb{E}_t z_{t+k} = \sum_{j=1}^{k} P^{k-j} Q \xi_{t+j}.$$

Consequently:

$$\operatorname{cov}_t(z_{t+a}, z_{t+b}) = \sum_{i=1}^{a}\sum_{j=1}^{b} P^{a-i} Q\big(\mathbb{E}_t \xi_{t+i}\xi'_{t+j}\big) Q' P'^{b-j}$$

If $\mathbb{E}_t\xi_{t+i}\xi'_{t+j} = 0$ for $i \neq j$, then this simplifies to:

$$\operatorname{cov}_t(z_{t+a}, z_{t+b}) = P^{a-\min\{a,b\}}\left[\sum_{i=1}^{\min\{a,b\}} P^{\min\{a,b\}-i} Q(\mathbb{E}_t\xi_{t+i}\xi'_{t+i})Q' P'^{\min\{a,b\}-i}\right] P'^{b-\min\{a,b\}}.$$

Now, in the previous section of these appendices (D), we showed that at order 1, 2 and 3 the pruned perturbation solutions may be represented in the form:

$$z_t = o + P\tilde{z}_{t-1} + Q\xi_t$$

where $\mathbb{E}_t\xi_{t+k} = 0$ for $k > 0$. It is trivial to add zero columns to P so that we instead have:

$$z_t = o + Pz_{t-1} + Q\xi_t,$$

thus, we just need to evaluate $\mathbb{E}_t\big(\xi_{t+i}\xi'_{t+j}\big)$ in order to have a closed form expression for $\operatorname{cov}_t(z_{t+a}, z_{t+b})$, then from this and the fact that $x_t = u + Vz_t$, we would have that:

$$\operatorname{cov}_t(x_{t+a}, x_{t+b}) = \operatorname{cov}_t(u + Vz_{t+a}, u + Vz_{t+b}) = V \operatorname{cov}_t(z_{t+a}, z_{t+b})\, V'.$$

We now proceed to evaluate $\mathbb{E}_t\xi_{t+i}\xi'_{t+j}$ for orders 1 and 2. We skip the third order case as a second order approximation to the conditional covariance is normally sufficient for reasonable accuracy, and as the third order conditional covariance is very slow to calculate.

Order 1 At order 1, $\xi_t := \varepsilon_t$, thus:

$$\mathbb{E}_t\xi_{t+i}\xi'_{t+j} = \begin{cases}\Sigma & \text{if } i = j\\ 0 & \text{if } i \neq j\end{cases}.$$

Order 2 At order 2:

$$\xi_t := \begin{bmatrix}\varepsilon_t \\ \varepsilon_t \otimes \varepsilon_t - \operatorname{vec}\Sigma \\ \tilde{x}^{(1)}_{t-1} \otimes \varepsilon_t\end{bmatrix},$$

thus $\mathbb{E}_t \xi_{t+i} \xi_{t+j}' = 0$ if $i \neq j$, and by theorem 4.3 of Magnus and Neudecker (1979):

$$\begin{aligned}\mathbb{E}_s \xi_t \xi_t' &= \mathbb{E}_s \begin{bmatrix} 1 \otimes \varepsilon_t \\ \varepsilon_t \otimes \varepsilon_t - \operatorname{vec} \Sigma \\ \tilde{x}_{t-1}^{(1)} \otimes \varepsilon_t \end{bmatrix} \begin{bmatrix} 1 \otimes \varepsilon_t' & \varepsilon_t' \otimes \varepsilon_t' - (\operatorname{vec} \Sigma)' & \tilde{x}_{t-1}^{(1)'} \otimes \varepsilon_t' \end{bmatrix} \\ &= \begin{bmatrix} \Sigma & 0 & \mathbb{E}_s \tilde{x}_{t-1}^{(1)'} \otimes \Sigma \\ 0 & (I_{m^2} + K_{m,m})(\Sigma \otimes \Sigma) & 0 \\ \mathbb{E}_s \tilde{x}_{t-1}^{(1)} \otimes \Sigma & 0 & \mathbb{E}_s \left(\tilde{x}_{t-1}^{(1)} \tilde{x}_{t-1}^{(1)'} \right) \otimes \Sigma \end{bmatrix}.\end{aligned}$$

F. Global solution procedure for the capital constrained model

The value function is:

$$\begin{aligned}V(K, A) &= \max_{\substack{C, L \text{ s.t.} \\ AK^\alpha L^{1-\alpha} - C \geq \theta K}} \left[\log C - \frac{L^{1+\nu}}{1+\nu} + \beta \mathbb{E} V(AK^\alpha L^{1-\alpha} - C, A^\rho \exp \varepsilon) \right] \\ &= \max_{\substack{C \text{ s.t.} \\ \left((AK^\alpha)^{1+\nu} \left(\frac{1-\alpha}{C} \right)^{1-\alpha} \right)^{\frac{1}{\nu+\alpha}} - C \geq \theta K}} \left[\log C - \frac{1}{1+\nu} \left(\frac{1-\alpha}{C} A K^\alpha \right)^{\frac{1+\nu}{\nu+\alpha}} \right. \\ &\qquad \left. + \beta \mathbb{E} V \left(\left((AK^\alpha)^{1+\nu} \left(\frac{1-\alpha}{C} \right)^{1-\alpha} \right)^{\frac{1}{\nu+\alpha}} - C, A^\rho \exp \varepsilon \right) \right]\end{aligned}$$

where the second line comes from substituting in the labour first order condition, and where $\varepsilon \sim \mathrm{N}(0, \sigma^2)$.

We construct a uniform rectangular grid on $\log A$ and $\log K$ as follows. For both variables, the grid is centred on the model's non-stochastic steady-state. For productivity, the grid extends to ± 4 times the standard deviation of $\log A_t$, which covers more than 99.99% of its stationary distribution. For capital, the grid extends to ± 16 times the standard deviation of $\log K_t$ in the model without bounds, which also covers (a lot) more than 99.99% of its stationary distribution in the model with bounds. Furthermore, it is wide enough to cover the area in which the value function is highly curved, as it goes from increasing to decreasing. The grid has 256 points along the productivity axis, and 1024 points along the capital axis, making for a total of 262,144 points.

We use linear interpolation/extrapolation for points off the grid. Then, due to piecewise linearity of the integrand in the value function, and the Gaussianity of ε, integration can be performed exactly. For speed, we precompute the associated weights at each productivity level, so exact integration just requires a dot product between the weights, and the value function interpolated to the future capital level.

We initialize the value function iteration algorithm with the exact solution in the absence of bounds, e.g.:

$$V = F + G \log K + H \log A,$$

where:

$$F = \frac{\log(1-\alpha\beta)}{1-\beta} + \frac{1-\alpha}{(1-\alpha\beta)(1-\beta)(1+\nu)}\left[\log\left(\frac{1-\alpha}{1-\alpha\beta}\right) - 1\right] + \frac{\alpha\beta\log\alpha\beta}{(1-\alpha\beta)(1-\beta)},$$

$$G = \frac{\alpha}{1-\alpha\beta},$$

$$H = \frac{1}{(1-\alpha\beta)(1-\beta\rho)}.$$

To facilitate solving with the bound, in the first iteration, we set $\theta = 0$, and then we increase θ by 0.005 with each iteration until it gets to 0.99, in order to "homotope" from the solution without the bound to the solution with the bound at the correct level. We then continue with conventional fixed point iterations until the maximum absolute change in the value function over the grid ceases to decrease. We report the penultimate value function, i.e. we discard the final one which was a bigger step away. Within these iterations, at each grid node, we first solve for the C at which the constraint binds exactly, then maximise C over the interval from 0 to the found bound. Both procedures will deliver a result accurate to somewhere (roughly) between 10^{-8} and 10^{-16}.

As one indication of accuracy, the final value function iteration step changed the value function by at most 1.73×10^{-11} at all grid points, and the implied policy function changed by at most 1.38×10^{-7} at all grid points. By way of comparison, when the same algorithm was run on the model without a bound, the algorithm made two steps (the minimum possible), the last of which changed the value function by at most 4.41×10^{-13} and which changed the policy function by at most 1.83×10^{-7}.

In Figure 3 and Figure 4 we plot the value and policy functions in terms of $\log K$ for a variety of productivity levels, including both the highest and lowest productivities on the grid. These illustrate the extent of the departure from the model without bounds.

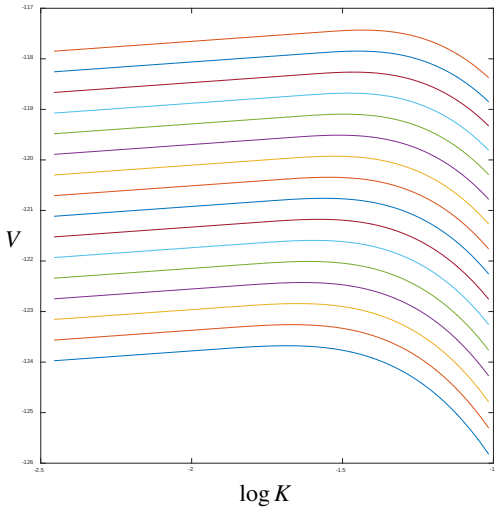

Figure 3: Value function

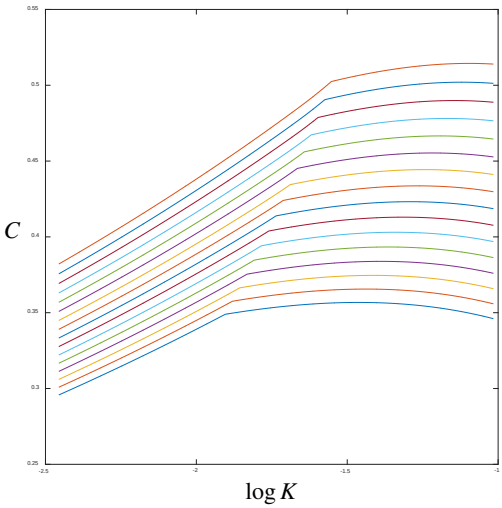

Figure 4: Policy function

G. Global solution procedure for the borrowing constrained model

In terms of $X = C + D$, the value function is:

$$V(B,A) = \max_{X \in \Gamma(B,A)} \left[-\frac{1}{2}[\max\{0,1-X\}]^2 - \frac{\phi}{2}[RB + \max\{\underline{Y},A\} - X]^2 + \beta \mathbb{E} V(RB + \max\{\underline{Y},A\} - X, (1-\rho)\mu + \rho A + \varepsilon) \right]$$

where $\varepsilon \sim N(0,\sigma^2)$ and where:

$$\Gamma(B,A) = \left[0, RB + \max\{\underline{Y},A\} + \frac{\underline{Y}}{R-1}\right].$$

We construct a uniform rectangular grid on A and B as follows. For productivity, the grid is centred on μ and extends to ± 4 times the standard deviation of A, which covers more than 99.99% of its stationary distribution. For bond holdings, the minimum of the grid is at the minimum possible value of bond holdings, namely $-\frac{\underline{Y}}{R-1}$. The maximum of the grid is chosen as a point close to, but higher than the point at which the value function becomes flat in B at all points on the productivity grid.[22] With our parameters, this corresponds to an upper bound of $B = 7.5$. The grid has 256 points along the productivity axis, and 1024 points along the bond holdings axis, making for a total of 262,144 points.

We again use linear interpolation/extrapolation along the productivity axis, but here we use shape preserving cubic spline interpolation along the bond holding axis. Using cubic splines is not problematic here since no extrapolation in bond holdings is required. By producing a smoother value function, it also made numerical maximisation easier, reducing numerical errors. As before, due to piecewise linearity of the integrand in the value function, and the Gaussianity of ε, integration can be performed exactly, with precomputation of weights as before.

[22] For any productivity level, for sufficiently high bond holdings, $X = 1$ and the value function is flat in B.

We initialize the value function iteration algorithm with the cumulative maximum (in A and B) of the exact solution in the absence of bounds. The problem in the absence of bounds is a standard quadratic dynamic programming problem, and features linear policy functions and a quadratic optimal value function. We then proceed with conventional fixed point iterations until the maximum absolute change in the value function over the grid ceases to decrease. We report the penultimate value function, i.e. we discard the final one which was a bigger step away. Within these iterations, for each productivity level we first solve for the level of B at which it is optimal to choose $X = 1$. Then, for any higher B we know that the value function will be flat and that $C = 1$. This both reduces the number of times the value function must be maximised, and reduces numerical errors.

As before, as an indication of accuracy we note that the final value function iteration step changed the value function by at most 1.32×10^{-11} at all grid points, the implied policy function changed by at most 4.60×10^{-7} at all grid points, and the function giving the value of B at which it is optimal to choose $X = 1$ changed by a maximum of 4.56×10^{-7}.

In Figure 5 and Figure 6 we plot the value and policy functions in terms of B for a variety of productivity levels, including both the highest and lowest productivities on the grid. Note the high curvature in the policy function below the $X = 1$ kink; it is certainly not the case that the optimal policy function is piecewise linear here. Figure 7 plots the function giving the level of B at which it is optimal to set $X = 1$.

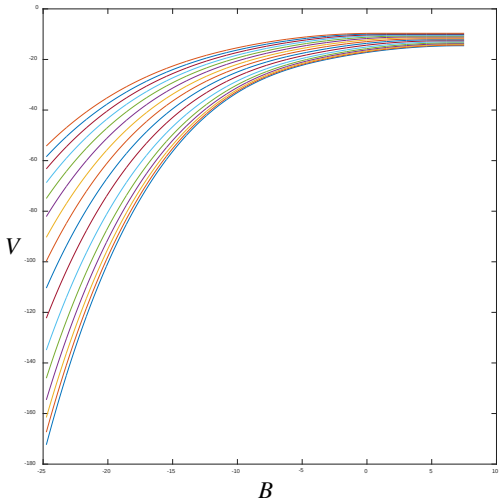

Figure 5: Value function

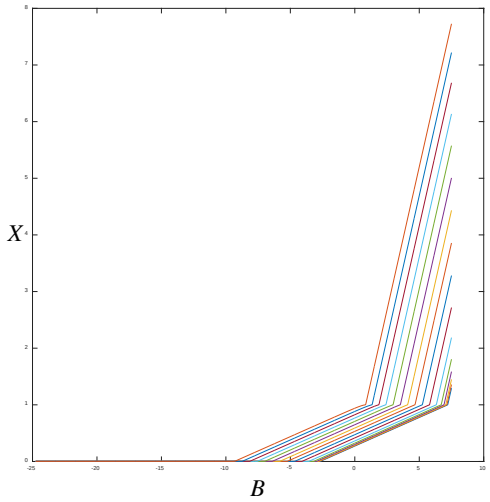

Figure 6: Policy function

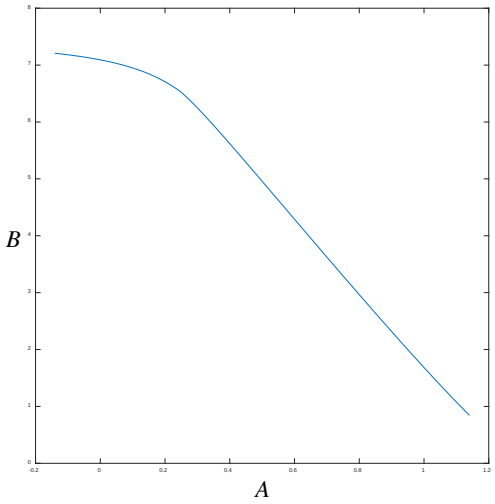

Figure 7: Value of B at which $X = 1$