Perturbation Methods for Incomplete Markets Economies*

David Childers† Keshav Dogra‡

December 5, 2019

Abstract

This paper proposes a perturbation-based algorithm to solve models with both idiosyncratic and aggregate risk and a distribution of agents. The algorithm is described with an application to a version of the canonical model of Krusell and Smith (1998), which features aggregate and idiosyncratic risk, a continuum of agents, and incomplete markets. It proceeds by representing the equilibrium conditions of the model as a system of functional equations, discretizing and linearizing these equations, and solving the resulting finite-dimensional linear model using standard methods. We discuss theoretical convergence properties of the method and present numerical results.

Keywords: perturbation methods, incomplete markets
JEL codes: C62, C63, D52, E32

*The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

†CMU **Email:** dchilder@andrew.cmu.edu

‡Federal Reserve Bank of New York **Email:** keshav.dogra@ny.frb.org

1 Introduction

A large and growing literature studies the interaction between microeconomic heterogeneity and aggregate fluctuations. Studying models with incomplete markets and a continuum of agents has proved challenging, however, because the distribution of wealth in these models is generally an infinite-dimensional state variable. While many methods have been proposed to solve particular models, none have seen widespread adoption in the way that perturbation methods have become the standard method used to solve representative agent DSGE models.

This paper introduces a perturbation-based algorithm to solve models with both idiosyncratic and aggregate risk and a distribution of agents. For ease of exposition, the algorithm is described with an application to a modified version of the canonical model of Krusell and Smith (1998), which features continuously distributed aggregate and idiosyncratic risk, a continuum of agents, and incomplete markets. It proceeds by representing the equilibrium conditions of the model as a system of functional equations, discretizing and linearizing these equations, and solving the resulting finite-dimensional linear model using standard methods akin to Blanchard and Kahn (1980). Thus, like the methods of Reiter (2009), Ahn et al. (2017) and others, the approximation is linear with respect to aggregate shocks, but fully nonlinear with respect to idiosyncratic risk.

Literature review There are many ways to solve models with both aggregate and idiosyncratic risk. Our method is most similar to the subset of this literature that uses a combination of global and local methods to solve these models. Our approach is inspired by the function space perspective on perturbation methods for heterogeneous agent models in Childers (2018b), but substantially simplifies the implementation by allowing all computation to take place directly in finite dimensional Euclidean space. The method introduced here may be considered an application of the general framework described in Childers (2018a), in the special case in which all functions are represented by histograms. Relative to the general case, this choice eliminates the need for several steps of the process, reducing the procedure to application of standard numerical solution methods to a discretized model.

Among other methods applying perturbation methods to heterogeneous agent models, our algorithm is most closely related to Reiter (2009) and Winberry (2018). The difference between our algorithm and their approaches is rather subtle. Both our and their approaches begin by representing a model as a system of functional equations. In incomplete markets economies, these functional equations contain the composition of endogenous functions with other endogenous functions. For example, the Euler equation contains tomorrow's marginal utility of consumption function evaluated at today's saving function. We use a change of variables formula to eliminate compositions of endogenous functions. This means that when we linearize the model, the linearized equilibrium conditions (which are operators M mapping functions $x(w')$ to functions $(Mx)(w)$) are Fredholm integral equations, i.e. take the form

$$(Mx)(w) = \int k(w, w')x(w')dw$$

As we discuss in detail in Section 4, this approach makes it easier to prove theoretical convergence results, since it is relatively straightforward to approximate Fredholm integral equations. In contrast, Reiter (2009) and Winberry (2018) approximate this composition of functions by approximating each function directly,

using splines or Chebyshev polynomials. The derivatives of their equilibrium conditions are not Fredholm integral equations, which makes it harder to prove convergence results. To be clear, our claim is not that our method has superior convergence properties to Reiter (2009) and Winberry (2018). Instead, we view our contribution as a first step in characterizing the theoretical properties of perturbation methods for solving incomplete markets models more generally. Other perturbation-based approaches may well have similar properties. Another closely related method is that of Ahn et al. (2017), who show how to discretize a continuous time model using finite difference methods and then linearize with respect to aggregate shocks.

The remainder of the paper proceeds as follows. Section 2 describes the economic model. Section 3 describes our algorithm. Section 4 describes some of its theoretical properties. Section 5 presents numerical results. Section 6 concludes.

2 Model

We use a variant of the Krusell and Smith (1998) economy to illustrate how to implement our perturbation solution algorithm. Time is discrete. There exists a continuum of households with measure one. Households have preferences

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t^i)^{1-\gamma}}{1-\gamma} \qquad \text{where} \qquad \gamma > 0$$

over the single consumption good, where $\beta < 1$ denotes the discount factor. Markets are incomplete: the only asset households have access to is capital, and they are restricted to hold a non-negative amount of capital. Capital earns a rate of return r_t and depreciates at rate δ. Each household is endowed with s_t^i units of effective labor, where s_t^i is i.i.d. with density $g(s)$, assumed to be infinitely differentiable and with bounded support. Households inelastically supply their endowment of effective labor at a wage ω_t. This yields the budget constraint

$$k_{t+1}^i + c_t^i = \omega_t s_t^i + (r_t + 1 - \delta) k_t^i$$

Defining $R_t = r_t + 1 - \delta$, the household's problem yields the Euler equation

$$(c_t^i)^{-\gamma} \geq \beta \mathbb{E}_t [R_t (c_{t+1}^i)^{-\gamma}] \tag{1}$$

which holds with equality unless $k_{t+1}^i = 0$.

Competitive firms produce output Y_t using a Cobb-Douglas technology combining capital K_t and effective labor input L_t:

$$Y_t = Z_t K_t^{\alpha} L_t^{1-\alpha}$$

where aggregate productivity Z_t follows the process

$$\ln Z_t = \rho_Z \ln Z_{t-1} + \sigma \varepsilon_t \tag{2}$$

and $\varepsilon_t \sim N(0,1)$. Firms maximize profit, yielding $r_t = \alpha Z_t K_t^{\alpha-1} L_t^{1-\alpha}$, $\omega_t = (1-\alpha) Z_t K_t^{\alpha} L_t^{-\alpha}$.

The markets for goods, capital and labor clear:

$$K_{t+1} = \int k_{t+1}^i di = Y_t - \int c_t^i di \tag{3}$$

$$\int s g(s) ds = L_t \tag{4}$$

Note that the supply of effective labor is constant, $L_t = L$.

3 How we solve Krusell Smith

3.1 Writing the model as a set of functional equations

In a recursive equilibrium, household i's decision depends on its individual state variables k_t^i, s_t^i; the aggregate exogenous state variable z_t; and the distribution of wealth. Since s_t^i is independent over time, we can work with 'cash on hand', $w_t^i = R_t k_t^i + \omega_t s_t^i$, as an individual state variable which evolves according to

$$w_{t+1}^i = R_{t+1}(w_t^i - c_t^i) + \omega_{t+1} s_{t+1}^i$$

Let $\mu_t(w)$ denote the cross-sectional density of cash on hand at date t. The t subscript indicates that this density itself depends on the current aggregate shock and the past distribution of wealth.

We rewrite the Euler equation 1 using a version of the parameterized expectations algorithm. Define the expected discounted marginal utility of consumption $\ell_t^i = \beta \mathbb{E}_t[R_{t+1} u'(c_{t+1}^i)]$; we will work with ℓ_t^i rather than c_t^i. Then we can rewrite the Euler equation as

$$\ell_t^i = \beta \mathbb{E}_t[R_{t+1} \min\{\ell_{t+1}^i, (w_{t+1}^i)^{-1/\gamma}\}]$$

Here the expectation is taken over both aggregate and idiosyncratic shocks. We can rewrite this as a functional equation:

$$\ell_t(w) = \beta \mathbb{E}_t \left[R(Z_{t+1}, K_{t+1}) \int c_{t+1} \left(R(Z_{t+1}, K_{t+1})[w - c_t(w)] + \omega(Z_{t+1}, K_{t+1}) s \right)^{-\gamma} g(s) ds \right] \tag{5}$$

where the expectation is taken over aggregate shocks only, and we use the shorthand notation

$$c_t(w) = \min\{\ell_t(w)^{-1/\gamma}, w\} \tag{6}$$

$$R(Z, K) = \alpha Z K^{\alpha - 1} L^{1-\alpha} + 1 - \delta \tag{7}$$

$$\omega(Z, K) = (1 - \alpha) Z K^{\alpha} L^{-\alpha} \tag{8}$$

Note that in evaluating the right hand side of equation (5) for a household with date t wealth w, we must evaluate tomorrow's consumption function $c_{t+1}(\cdot)$ on values which depend on today's consumption function $c_t(\cdot)$. To circumvent this problem, we can use a change of variables to obtain

$$\ell_t(w) = \beta \mathbb{E}_t \left[R(Z_{t+1}, K_{t+1}) \int c_{t+1}(w')^{-\gamma} \frac{1}{\omega(Z_{t+1}, K_{t+1})} g\left(\frac{w' - R(Z_{t+1}, K_{t+1})(w - c_t(w))}{\omega(Z_{t+1}, K_{t+1})} \right) dw' \right] \tag{9}$$

(9) no longer contains the composition of two endogenous functions, but only the composition of an exogenously given function $g(\cdot)$ with the endogenous function $c_t(\cdot)$

Similarly, the distribution of cash on hand $\mu_t(x)$ evolves according to the discrete-time Kolmogorov forward equation

$$\mu_{t+1}(w') = \int \frac{1}{\omega(Z_{t+1}, K_{t+1})} g\left(\frac{w' - R(Z_{t+1}, K_{t+1})(w - c_t(w))}{\omega(Z_{t+1}, K_{t+1})}\right) \mu_t(w) dw \tag{10}$$

The market clearing condition can be written

$$K_{t+1} - \int (w - c_t(w))\mu_t(w) dw = 0 \tag{11}$$

An equilibrium of the model is a collection of sequences $\{\mu_t(\cdot), \ell_t(\cdot), K_t, Z_t\}_{t=0}^{\infty}$ that solve (9), (10), (11), (2).

3.2 Adding lagged variables

Like the methods of Blanchard and Kahn (1980), Klein (2000) and others, our solution method requires us to distinguish between 'predetermined' variables (those with an exogeneously determined initial value) and 'non-predetermined' variables. In the system of functional equations (9), (10), (11), (2), one might think that the policy function $\ell_t(w)$ would be a non-predetermined variable, while the distribution of cash on hand $\mu_t(w)$ would be a predetermined variable. Unfortunately, this is not quite right: $\mu_t(w)$ is not a predetermined variable. While the distribution of *capital* is predetermined, the mapping from capital to cash on hand depends on factor prices, which themselves depend on the aggregate productivity shock. This problem can be solved in the usual way, by augmenting the system to include lagged variables - in this case, lagged policy functions and distributions of wealth, which we define by

$$L\mu_{t+1}(w) = \mu_t(w) \tag{12}$$

$$L\ell_{t+1}(w) = \ell_t(w) \tag{13}$$

The Kolmogorov forward equation becomes

$$\mu_t(w') = \int \frac{1}{\omega(Z_t, K_t)} g\left(\frac{w' - R(Z_t, K_t)(w - Lc_t(w))}{\omega(Z_t, K_t)}\right) L\mu_t(w) dw \tag{14}$$

where again for convenience we define $Lc_t(w) = \min\{L\ell_t(w)^{-1/\gamma}, w\}$. Our complete system of equations is then (9), (14), (12), (13), (11), (2). The predetermined variables are $L\mu_t(\cdot), L\ell_t(\cdot), K_t, Z_t$ and the non-predetermined variables are $\mu_t(\cdot), \ell_t(\cdot)$.

3.3 Discretizing, linearizing and solving the model

Having reduced the model to a set of nonlinear functional equations, the algorithm proceeds as follows. First, we discretize the model, representing the functions $\mu_t(\cdot)$, $\ell_t(\cdot)$ as finite-dimensional vectors, and the functional equations (9), (14) and so forth as nonlinear vector equations. Second, we solve for the deterministic steady state in which aggregate shocks are absent but idiosyncratic shocks remain. Third, we

linearize the system of nonlinear vector equations around the deterministic steady state. Finally, we use a variant of the Klein (2000) method to solve the linearized, discretized model.

Discretizing the model We define a grid $\mathcal{W} = \{w_1, w_2, ..., w_n\}$ and represent $\mu_t(\cdot)$ and $\ell_t(\cdot)$ by their values on this grid. We approximate integrals by Riemann sums, or equivalently, by histograms with bin heights given by the values at the grid points, so the discretized Euler equation and Kolmogorov forward equation becomes

$$\ell_t(w_i) = \beta \mathbb{E}_t \left[R(Z_{t+1}, K_{t+1}) \iota \sum_{j=1}^{n} \frac{c_{t+1}(w_j)^{-\gamma}}{\omega(Z_{t+1}, K_{t+1})} g\left(\frac{w_j - R(Z_{t+1}, K_{t+1})(w_i - c_t(w_i))}{\omega(Z_t, K_t)} \right) \right] \tag{15}$$

$$\mu_t(w_j) = \iota \sum_{i=1}^{n} \frac{1}{\omega(Z_t, K_t)} g\left(\frac{w_j - R(Z_t, K_t)(w_i - Lc_t(w_i))}{\omega(Z_t, K_t)} \right) L\mu_t(w_i) \tag{16}$$

where we define $\iota = \frac{w_n - w_1}{n}$. This gives us a system of $4n + 2$ dynamic nonlinear equations in $4n + 2$ variables:

$$\{L\mu_t(w_i)\}_{i=1}^n, \{L\ell_t(w_i)\}_{i=1}^n, K_t, Z_t, \{\mu_t(w_i)\}_{i=1}^n, \{\ell_t(w_i)\}_{i=1}^n$$

Denoting the vector of predetermined variables by x_t and the vector of non-predetermined variables by y_t, we have

$$\mathbb{E}_t f(x_{t+1}, y_{t+1}, x_t, y_t) = 0$$

This system has exactly the same structure as standard representative agent models, which are often solved using perturbation methods. Our method (like Reiter (2009) and Ahn et al. (2017)) exploits this analogy. Assuming that a recursive solution exists, a solution to the discretized model has the form

$$\begin{aligned} y_t &= g(x_t) \\ x_{t+1} &= h(x_t) + \eta \sigma \varepsilon_{t+1} \end{aligned}$$

where η is a $(2n + 2) \times 1$ vector whose elements are all zero except for its last element which equals 1.

Solving for steady state Second, we solve for a deterministic steady state in which aggregate shocks are absent, so $\sigma = 0$ and Z is constant, but idiosyncratic shocks remain. A steady state of the discretized model is a capital stock K, a cash on hand distribution $\{\mu(w_i)\}_{i=1}^n$, and a policy function $\{\ell(w_i)\}_{i=1}^n$. Stacking these vectors as $x = (\mu, \ell, K, 1)$, $y = (\mu, \ell)$, the steady state is a pair (x, y) which solves $f(x, y, x, y) = 0$, if such a pair exists. The steady state can be computed using standard methods as in Aiyagari (1994). The one proviso is that to ensure accuracy guarantees for the full method, the steady state must likewise be computed with sufficient accuracy. For these purposes, it is sufficient to assume that the value of the steady state functions on this grid converges uniformly over all grid points to the true steady state values either (a) at rate $o(n^{-1})$ or (b) at rate $O(n^{-1})$ and the parameterized expectation function $\{\ell(w_i)\}_{i=1}^n$ is monotone over the grid. In all our example applications, monotonicity holds empirically, and the latter convergence rate is standard for discretization methods. For any uniformly accurate approximation method, arbitrarily fast convergence rates can be achieved, if needed, by refining the discretization, and passing on only a subset of grid values to the next step.

Linearization Next, we compute a first order Taylor approximation of the discretized equilibrium conditions around the deterministic steady state.[1] This step can be performed either analytically or using automatic differentiation. Analytical derivatives for our model are described in Appendix A. Automatic differentiation means the numerical computation of derivatives using exact differentiation rules rather than finite-difference approximations. This dominates finite-difference methods in terms of accuracy and symbolic differentiation in terms of speed.[2] If automatic differentiation is used, all the remaining steps in the algorithm can be automated, given a set of nonlinear equations and a steady state provided by the user. Similarly, if analytical derivatives are provided, the remaining steps in the algorithm can be automated. The first order Taylor approximation gives us the system

$$A\mathbb{E}_t \begin{bmatrix} \hat{x}_{t+1} \\ \hat{y}_{t+1} \end{bmatrix} = B \begin{bmatrix} \hat{x}_t \\ \hat{y}_t \end{bmatrix}$$

where hats denote deviations from the deterministic steady state y^*, x^*. We are now almost ready to use Schur decomposition methods, as in Klein (2000), to compute a linear approximation to the policy functions, i.e. the $2n \times (2n+2)$ matrix g_x and the $(2n+2) \times (2n+2)$ matrix h_x. Before doing so, there is one minor transformation which must be made to the model.

Normalization Recall that $\mu_t(x)$ denotes the distribution of cash on hand; we approximate it by a vector $\{\hat{\mu}_t(x_i)\}_{i=1}^n$ denoting the deviations of the distribution from steady state, evaluated on our grid. Since μ_t is a distribution, it should always integrate to one, whatever the sequence of aggregate shocks. Therefore we enforce that $\{\hat{\mu}_t(x_i)\}_{i=1}^n$ always integrates to one, as follows. Define the matrix

$$P = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 1 & 0 & \dots & 1 \end{bmatrix}$$

and take the QR decomposition $P = QR$ of P. We define S to be the transpose of the matrix obtained by deleting the first column of Q. This $(n-1) \times n$ matrix maps a n-vector representing the deviation of a distribution from steady state into a $(n-1)$-vector representing the deviation of a distribution which is constrained to sum to zero.

We then premultiply every equation block by S if it outputs a distribution. Given our ordering of equations, this means that we premultiply the second equation block (16), which defines the current cash on hand distribution $\mu_t(x)$, and the third equation block (12), which updates the lagged cash on hand distribution. Similarly, we postmultiply the first variable ($L\mu_t$) and the fifth variable (μ_t) by S. Formally, we define

$$\tilde{A} = Q_f A \begin{bmatrix} Q_x & 0 \\ 0 & Q_y \end{bmatrix}, \qquad \tilde{B} = Q_f B \begin{bmatrix} Q_x & 0 \\ 0 & Q_y \end{bmatrix}$$

[1] For ease of exposition, we linearize rather than log-linearizing, but nothing prevents us from log-linearizing as is standard.

[2] We use the Julia package `ForwardDiff.jl`, which implements forward mode automatic differentiation (Revels et al., 2016).

where

$$Q_f = \begin{bmatrix} I_n & 0 & 0 & 0 & 0 & 0 \\ 0 & S & 0 & 0 & 0 & 0 \\ 0 & 0 & S & 0 & 0 & 0 \\ 0 & 0 & 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \qquad Q_x = \begin{bmatrix} S & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad Q_y = \begin{bmatrix} S & 0 \\ 0 & I_n \end{bmatrix}$$

$\tilde{A}$ and $\tilde{B}$ are both $4n \times 4n$ matrices. They define the linear system

$$\tilde{A}\mathbb{E}_t \begin{bmatrix} \tilde{x}_{t+1} \\ \tilde{y}_{t+1} \end{bmatrix} = \tilde{B} \begin{bmatrix} \tilde{x}_t \\ \tilde{y}_t \end{bmatrix} \tag{17}$$

where $\tilde{x}_t = Q_x \hat{x}_t$, $\tilde{y}_t = Q_y \hat{y}_t$ denote normalized variables.

Schur decomposition Finally, we apply the generalized Schur decomposition to (17) as in Klein (2000), and then use the results to construct linearized solutions as in Schmitt-Grohe and Uribe (2004). This gives us a solution in terms of normalized variables: $\tilde{y}_t = \tilde{g}_x \tilde{x}_t$, $\tilde{x}_{t+1} = \tilde{h}_x \tilde{x}_t + \eta\sigma\varepsilon_{t+1}$. To recover a solution in terms of un-normalized variables, we simply define[3]

$$g_x = Q_y^{-1} \tilde{g}_x Q_x^{-1}, \qquad g_x = Q_x^{-1} \tilde{h}_x Q_x^{-1}$$

This completes our solution of the model.

4 Theoretical Results

The algorithm described in Section 3 builds on the perturbation approach to solution of rational expectations models with function-valued variables described in Childers (2018b) and the framework for automated solution of such models provided in Childers (2018a). Finite-dimensional perturbation methods (e.g. Schmitt-Grohe and Uribe (2004)) aim to find a local approximation of the solution of a rational expectations model - in other words, they aim to find the derivatives of the policy functions g and h evaluated at a non-stochastic steady state, which are matrices mapping finite vectors to finite vectors. We have the same goal. The difference is that in models with function-valued variables, such as incomplete markets models, the policy functions g and h are infinite-dimensional mappings from functions to functions - for example, mapping a distribution of wealth today to a distribution of wealth tomorrow, or a distribution of wealth today to a consumption function today. In practice, we must approximate these infinite dimensional objects by finite-dimensional objects. Any such approximation is imperfect: our finite-dimensional approximations to g_x and h_x cannot exactly describe how the distribution of wealth or the consumption function would evolve given any sequence of shocks. Nonetheless, we would like our approximation to be accurate in the sense that as we increase the number of grid points in our finite-order approximation, our approximate solution converges to the true solution (i.e. the derivatives of the policy functions of the 'true', infinite-dimensional model).

[3]Note that since Q_x and Q_y are orthogonal, $Q_x^{-1} = Q_x^\top$, $Q_y^{-1} = Q_y^\top$.

Intuitively, we need two conditions to hold in order for our finite-order approximation to the policy functions to converge to the true, infinite-dimensional policy functions. First, we need our finite-dimensional representation of the model's linearized equilibrium conditions to converge to the 'true' infinite-dimensional linearized equilibrium conditions in an appropriate norm. Second, we need the mapping from equilibrium conditions to linearized policy functions to be continuous in an appropriate sense.

Our algorithm is designed to ensure that the first of these conditions holds: our finite-dimensional representation of the model's linearized equilibrium conditions converges to the 'true' infinite-dimensional linearized equilibrium conditions as the number of grid points increases, in a sense to be defined below. This can be demonstrated by applying the general results for approximation of functional derivatives of economic models introduced in Childers (2018a). The main difficulty is that incomplete market models such as Krusell and Smith (1998) have kinks in their equilibrium conditions due to the presence of binding borrowing constraints; in particular, the consumption function has a kink at the point where the borrowing constraint is just binding. Kinks in the equilibrium conditions translate into discontinuities in the functional derivatives of the equilibrium conditions, which makes it hard to show uniform convergence. We discuss how our method ensures that this is the case.

In contrast, there is no guarantee that the second condition described above holds, i.e. that the mapping from linearized equilibrium conditions to linearized policy functions is continuous. This mapping is constructed from an infinite-dimensional version of the Schur decomposition employed in standard linear rational expectations solution methods (Klein (2000), Schmitt-Grohe and Uribe (2004)). In order for our approximate policy functions to converge to true functions, we need to assume - in addition to infinite-dimensional analogues of the standard Blanchard and Kahn (1980) conditions stating that the number of pre-determined variables equals the number of stable eigenvalues - that these eigenvalues are not too sensitive to small changes in the linearized equilibrium conditions. These conditions are formally stated in Childers (2018b) Condition (1). It is not possible to prove analytically whether the Krusell and Smith (1998) model satisfies these conditions, nor can they be directly verified numerically. It is possible to perform tests which detect asymptotically some violations of these conditions, using our finite approximation of the model and its eigenvalues, as described in Appendix B, though these tests cannot detect all possible cases in which the eigenvalues of the true infinite-dimensional model may not be well-behaved. Nonetheless, if Condition (1) holds, our finite-order approximation to the policy functions does indeed converge to the true, infinite-dimensional policy functions.

Finally, it is worth noting that all our approximation results concern convergence of the finite-dimensional linearized model to the infinite-dimensional linearized model. The infinite-dimensional linearized model is still only a first order approximation to the 'true', infinite-dimensional nonlinear model.

4.1 Guarantees for Functional Derivatives of the Equilibrium Conditions

For shorthand, denote by

$$\mathcal{F}(x_{t+1}, y_{t+1}, x_t, y_t) = 0$$

the system of equations (9), (14), (12), (13), (11), (2), $x_t = \{L\mu_t(.), L\ell_t(.), K_t, Z_t\}$ be the vector of predetermined variables, $y_t = \{\mu_t(.), \ell_t(.)\}$ the vector of jump variables, and $\mathcal{H}_x = \{L_0^2 \times L^2 \times \mathbb{R} \times \mathbb{R}\}$, $\mathcal{H}_y = \{L_0^2 \times L^2\}$ the spaces of square integrable functions on which these objects live. Note that $\mathcal{F}$ is a functional equation.

The algorithm described in Section 3 produces a pair of matrices $(\tilde{A}, \tilde{B})$ which are discretized approxi-

mations of the functional derivatives (A, B) of $\mathcal{F}$ with respect to (x_{t+1}, y_{t+1}) and (x_t, y_t) respectively.[4] In order to ask whether our finite approximation converges to the true (A, B) as the number of grid points becomes large, we need to define the distance between our finite-dimensional matrices and the true infinite dimensional functional derivatives, i.e. we need to choose a metric. The metric we choose is based on the operator norm. For any linear map $M[.]$ defined over a Hilbert space $\mathcal{H}$ of functions with norm $\|.\|_{\mathcal{H}}$, the operator norm is defined as

$$\|M\|_{op} := \sup_{\{x \in \mathcal{H}: \|x\|_{\mathcal{H}}=1\}} \|M[x]\|_{\mathcal{H}}$$

We focus on convergence in operator norm, rather than a weaker notion of convergence, because the equilibrium conditions must converge in this strong sense in order for our approximation to the model's policy functions g_x, h_x to converge. However, the operator norm is defined on functions, and can be used to describe the distance between two functions; it cannot describe the distance between a function and a vector. So in order to describe a sense in which our finite-dimensional matrices $(\tilde{A}, \tilde{B})$ converge to the infinite-dimensional operators (A, B), we need to interpret the matrices $(\tilde{A}, \tilde{B})$ as operating on a space of functions. One way to do so is to describe an interpolation scheme which takes a finite vector and outputs an infinite-dimensional functions.[5] In particular, the interpolation scheme we will use in our proofs maps vectors to piecewise constant functions or histograms.

More concretely: let $\tilde{K}$ be a $n \times n$ matrix mapping n-vectors to n-vectors. We want to interpret it as an operator $\tilde{M}$ which, given an input function $x(w)$, produces an output function $y(w) = (\tilde{M}x)(w)$. We do so as follows. Given an input function $x(w)$ and a grid $\mathcal{W}_n := \{w_1, ..., w_n\}$, define the n-vector x by

$$x_i = \frac{1}{w_i - w_{i-1}} \int x(w) \mathbf{1}\{w_{i-1} \leq w < w_i\} dw, i = 1, ..., n,$$

where $w_0 = -\infty$, define the n-vector $y = \tilde{K}x$, and define the function $y(w)$ by

$$y(w) = \sum_{i=1}^{n} \mathbf{1}\{w_{i-1} \leq w < w_i\} y_i$$

Equivalently,

$$(\tilde{M}x)(w) = \int \frac{1}{w_j - w_{j-1}} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{1}\{w_{i-1} \leq w < w_i, w_{j-1} \leq w' < w_j\} x(w') dw' \quad (18)$$

$$= \int \tilde{k}(w, w') x(w') dw' \quad (19)$$

Having defined the operator mapping functions to functions in this way, we can then compute the distance between this operator and another operator mapping functions to functions - in particular, the true infinite-dimensional linearized equilibrium set of equilibrium conditions A. This is how we will conceptualize

[4] Precisely, as $L\mu$ and μ are density functions and so are defined only on the cone of non-negative functions integrating to 1, derivatives with respect to these arguments may be interpreted as the unique extension of the right (functional) derivatives defined over L_0^2, the space of square integrable functions which integrate to 0: see Childers (2018b) Appendix C.2. Derivatives with respect to all other arguments may be taken to be Fréchet derivatives.

[5] We do not actually need to construct this interpolation scheme in practice. It is a device used to ensure that $(\tilde{A}, \tilde{B})$ converge to (A, B) in a sense which is strong enough to ensure that our approximations to g_x, h_x converge, assuming of course that Condition (1) happens to hold.

the distance between our finite representation of the equilibrium conditions and the true infinite-dimensional equilibrium conditions.

Having *defined* convergence in operator norm, actually showing convergence in operator norm is hard for general maps A. However, the economic model described in Section 2, and in particular our representation of the model a set of functional equations in Section 3, have a special structure which allows us to show that convergence obtains. To explain why, we need to linearize the functional equations in function space.[6] The reason that the infinite dimensional operator can be well approximated by finite matrices is that it takes the form of a set of integral equations. After linearization in function space, each subcomponent of the conditions acting on a space of functions correponds to an integral operator of the form

$$(Mx)(w) = \int k(w, w')x(w')dw' \tag{20}$$

where the kernel function $k(w, w')$ is a function defined in terms of derivatives of equilibrium conditions at the steady state. Here the notation means that the operator M applied to the function $x(w')$ outputs another function $M[x]$. Evaluating this function at the particular value w gives the value $M[x](w)$. For example, the derivative of the Euler equation 9 with respect to $\ell_{t+1}(w')$ is the mapping

$$\mathcal{F}^1_{\ell_{t+1}}[\ell_{t+1}](w) := \int k(w, w')\ell_{t+1}(w')dw'$$

where the kernel function $k(w, w')$ is defined as[7]

$$k(w, w') = \frac{\beta R}{\omega}\mathbf{1}(\ell(w)^{-1/\gamma} < w)g\left(\frac{w' - R(w - c(w))}{\omega}\right)$$

The true linearized equilibrium conditions (20) and the approximate equilibrium conditions (18) have the same form: they are both integral equations. Approximating the operator M by the operator $\tilde{M}$ is equivalent to approximating the kernel function $k(w, w')$ by the histogram $\tilde{k}(w, w')$. In order to show that $\tilde{M}$ converges to M in operator norm, it suffices to show that $\tilde{k}$ converges to k in a particular norm, which is dominated by the L^∞ norm.[8] It is well known that histogram approximations of sufficiently regular functions converge in L^∞ (see, e.g. Nickl (2013)). Unfortunately, the presence of borrowing constraints induces discontinuities in some of our kernel functions k and the location of these discontinuities - the level of cash on hand w^* at which individuals are just liquidity constrained in the nonstochastic steady state of the true infinite-dimensional model - is not known, but must be approximated numerically. As a consequence, convergence of $\tilde{k}$ to k in L^∞ is not guaranteed.[9]

We deal with this issue as follows. In order to prove that approximate equilibrium conditions converge to

[6] Again, this step needs to be performed in the proof, but not in practice. The algorithm first discretizes the model, then linearizes, so it is never necessary to actually compute the functional derivatives of the model, although this step is straightforward in principle.

[7] The derivatives of the discretized model are presented in Appendix A. The derivatives of the infinite-dimensional model are essentially the same, but with weighted sums replaced with integrals.

[8] This follows from Young's inequality: see Johnstone (2013) Theorem C.26.

[9] While results for substantially weaker norms, such as L^2 norm, are readily available in this case (Mallat (2008)), these cannot control the approximation error of the map in terms of the operator norm. Intuitively, this is because uniformity over all function inputs must permit functions with substantial mass concentrated precisely at the region between the true discontinuity and the approximated value of the discontinuity, in which true and approximate functions must differ by an amount bounded away from 0.

true equilibrium conditions in a useful sense, we do not actually need to implement a particular interpolation scheme (i.e. a particular grid $\mathcal{W}_n$) in practice. We only need to show that some interpolation scheme exists. So we can choose - purely in theory - a grid $\mathcal{W}_n$ which includes w^* as one of its endpoints, and is evenly spaced (possibly with different widths) on each side. Call this the *transformed histogram representation* of the kernel function. Since the kernel function is continuous on either side of this discontinuity, standard histogram interpolation results imply that our transformed histogram representation of the kernel function $\tilde{k}$ converges uniformly to the true kernel function k. This in turn applies that the approximate operator $\tilde{M}$ associated with the transformed histogram representation converges to the true M in operator norm.

Since we do not know w^* exactly, we cannot actually implement the transformed histogram representation. Instead, we work with a known histogram representation with evenly spaced grid points, which will not include the breakpoint w^*. *This* $\tilde{k}$ may not converge uniformly to k. However, it may still be the case that the integral operator $\tilde{M}$ associated with $\tilde{k}$ converges to the true integral operator M, provided that we only apply these operators to sufficiently smooth input functions $x(w)$. If x is smooth, the known and transformed histogram representations of $\tilde{M}x$ will differ little. Since the transformed histogram representation is accurate with respect to operator norm, the known histogram representation will also be accurate with respect to L^2 norm. The upshot of this sequence of transforms and inverse transforms is that $\tilde{M}x$ is an accurate representation of Mx for sufficiently smooth functions x. Similarly, under the conjecture that the solution map from approximate derivatives of equilibrium conditions to derivatives of policy functions is continuous, operator norm accuracy of the transformed histogram representation will also ensure that the approximate solution of the model $\tilde{g}_x, \tilde{h}_x$ is accurate, in the sense that $\tilde{g}_x x$, $\tilde{h}_x x$ are close to $g_x x, h_x x$ for sufficiently smooth x. This means that objects of direct economic interest, like impulse responses to shocks and initial conditions away from the steady state, are also represented accurately.

Finally, before formalizing the consistency in the above sense of the functional derivatives, two more transformations need to be made, both of which are, at least in this model, purely formal in nature, in the sense that, although they change the form of the matrices $(\tilde{A}, \tilde{B})$, they leave the solutions (g_x, h_x) numerically identical, and so need not be implemented in practice. The first relates to the fact that some derivatives do not take the form described above, as kernel integral operators. In some cases, the functional derivative is simply an identity; in these cases, the numerical derivative is an identity matrix, which creates no issues with respect to the representation as a map on the space of functions. For others, including specifically the derivative of the Euler equation with respect to $\ell_t(w)$, the derivative does not take the form of an integral or an identity, but produces a diagonal matrix corresponding to multiplication by a function. Fortunately, this can be transformed to the other form by dividing all elements of the row of derivatives of the Euler equation by this function, and so dividing the kernel functions by this function. Approximation can then be measured in terms of the properties of this new kernel function. Numerically, this corresponds to multiplying by the inverse of the diagonal matrix of derivatives with respect to $\ell_t(w_i)$. As noted in Childers (2018a), left multiplication of a row by an invertible matrix leaves the solution unchanged.

Second, the procedure described in Childers (2018a) works explicitly with matrices which map coefficients of an orthonormal basis function representation of the input function to the coefficients of the basis function representation of the output function. It does this by including an extra step, which multiplies the matrices corresponding to the values of the kernel functions on a grid by interpolation matrices which provide the coefficient representation of these kernel functions. While this allows alternative choices of function

representation map beyond histograms, it requires distinguishing which parts of derivatives correspond to kernel functions and which correspond to identity maps. The useful property of evenly spaced histograms is that the coefficients can be chosen to be just the values of the grid points, scaled by a constant equal to the square root of the width of the bin, to assure that the L^2 norm of each term equals 1. It can therefore be shown (see Appendix) that the coefficient representation is just a constant rescaling of the rows and columns. As noted above, a rescaling of the rows produces identical solutions. A related result shows that a rescaling of the columns simply rescales the solution matrices; the corresponding function representation remains the same, though one should interpret the entries of the solution matrices as on the scale of values rather than of coefficients. Although in this case one must use a transformed histogram representation which is not necessarily evenly spaced to derive the bounds, in terms of implementation one can continue to use the same rescaled identity map to produce approximate coefficients which still produce sup norm accurate function approximations, albeit with some cost to accuracy relative to the infeasible case in which the exact location of the discontinuities were known.

Denote by $(\tilde{A}^m, \tilde{B}^m)$ the matrices resulting from applying the two transforms (division of the Euler equation by the diagonal matrix of the derivative with respect to $\ell_t(w)$ and scaling the rows and columns by the square root of the width). Again, these need not be computed in practice but the guarantees must be defined in terms of this representation of the derivatives.

The following lemma formalizes the above discussion, demonstrating that the method described produces Jacobian matrices $(\tilde{A}^m, \tilde{B}^m)$ such that the operators constructed from the evenly spaced histogram representations of these matrices converge uniformly over the class of smooth input functions. Specifically, define the classes of smooth functions $\mathcal{H}_x \supset \Lambda_x^\alpha := \{(L\mu(.), L\ell(.), K, Z) \in L_0^2 \times L^2 \times \mathbb{R} \times \mathbb{R} : \ L\mu(.) \in \Lambda^\alpha(w), L\ell(.) \in \Lambda^\alpha(w)\}$, $\mathcal{H}_y \supset \Lambda_y^\alpha := \{(\mu(.), \ell(.)) \in L_0^2 \times L^2 : \ \mu(.) \in \Lambda^\alpha(w), \ell(.) \in \Lambda^\alpha(w)\}$. In words, these are the subsets of the function spaces in which all functions are Hölder continuous with exponent $\alpha \geq \frac{1}{2}$.

1. *There exist some C_1, N such that for $n > N$ the maps $P_n = (P_n^x, P_n^y) : \ \mathcal{H}_x \times \mathcal{H}_y \to \mathbb{R}^{2n+1} \times \mathbb{R}^{2n-1}$ which map each $\ell(w)$ to its histogram representation $P[\ell] := \{\int \ell(w) 1\{\frac{w_n - w_1}{n}(i-1) \leq w < \frac{w_n - w_1}{n} i\} dw\}_{i=1}^n$ and each $\mu(w)$ to its normalized histogram representation $SP[\mu]$, satisfy,*

$$\sup_{\{(x,y) \in \Lambda_x^\alpha \times \Lambda_y^\alpha : \|(x,y)\|_{\mathcal{H}_x \times \mathcal{H}_y} = 1\}} \max\{\left\| [P_n^* \tilde{A}^m P_n - A](x,y) \right\|_{\mathcal{H}_x \times \mathcal{H}_y}, \left\| [P_n^* \tilde{B}^m P_n - B](x,y) \right\|_{\mathcal{H}_x \times \mathcal{H}_y}\} \leq C_1 n^{-\frac{1}{4}}$$

Remark. The proof of the lemma follows by first verifying the conditions of Childers (2018a) Lemma (10), which ensures that the set of kernel functions defined by the set of derivatives of the equilibrium equations satisfy sufficient regularity to be approximated at the given rate in operator norm by a function representation piecewise over the regions between discontinuities. One may then apply Lemma (25) of Childers (2018a), which demonstrates that over Hölder smooth inputs, replacing coefficients with respect to an evenly spaced histogram with those with respect to one aligned with the discontinuities contributes approximation error which is asymptotically of no higher order than the original error. The claimed rate of convergence is derived from the error contributed by terms due to histogram approximation of the kernel function, quadrature error from numerical integration, and numerical error in constucting the steady state, as well as from "stretching" the original input function and the histogram bins when converting between the evenly spaced and the transformed representation. This rate is slower than the $O(n^{-1/2})$ rate that could be obtained had the location of a discontinuity been exactly known, but still ensures consistency.

4.2 Conjecture regarding model parameters and implications of such conjecture

While the method described ensures accurate computation of the functional derivatives in the sense defined in Lemma (1), which provides some reassurance regarding the numerical results as reflecting the properties of the model, the accuracy of the results depends not only on the numerical properties of the system, but also on the numerical properties of the solution algorithm. As discussed earlier, a sufficient condition for the accuracy of the solutions (g_x, h_x) as numerical approximations of the functional derivatives of the policy functions of the model around the non-stochastic steady state is the continuity of the map defining these solutions as a function of the derivatives of the equilibrium conditions. Unfortunately, this map need not be continuous nor even be well defined for all models or parameter values. Sufficient conditions for continuity to hold are provided in Childers (2018b) Condition (1), which extends the Blanchard Kahn eigenvalue conditions for existence and uniqueness of a stable solution of the linearized equations to the function space setting and additionally imposes a stability condition which ensures that the numerical stability result for the QZ algorithm argued by Klein (2000) to provide strong justification for its use in the solution of linear rational expectations models extends to the infinite dimensional setting. As these are properties of the exact model and depend on the true spectrum, which can be approximated but not, in general, computed exactly nor have its properties verified outside of a few tractable special cases, whether they hold in the Krusell Smith model for the parameters chosen is unknown. Nevertheless, such a possibility is not ruled out nor, in the authors opinions, does it appear to be implausible in cases where the diagnostic criteria of Appendix B do not indicate a clear violation. For this reason, it is stated as a conjecture:

2. *For the parameters at which the model is evaluated, the system of equilibrium conditions $\mathcal{F}$ satisfies Condition (1) in Childers (2018b)*

More specifically, this imposes the following: that the true system of (functional) derivatives of the equilibrium conditions has generalized eigenvalues bounded away from the unit circle, with this separation defined according to a particular measure defined in that paper, and the space corresponding to the set of eigenvalues outside the unit circle is isomorphic to the space spanned by the jump variables $y = (\mu_t(.), \ell_t(.))$.

Remark. While not fully verifiable in this case, there are observable conditions which might rule out or provide evidence against the veracity of said conjecture. In particular, for finite dimensional isolated subsystems of the equations, this condition requires that the standard Blanchard Kahn conditions hold. So, for example, if an exogenous component possesses unstable dynamics, then no values of the other variables in the system can ensure that it is stable, and so the conditions would not be satisfied. Concretely, in this model, $|\rho_Z| > 1$ would result in a failure of the conjecture. Numerically, the (asymptotic) tests provided in Appendix B consist of relaxations of the Blanchard Kahn conditions requiring the set of generalized eigenvalues outside the unit circle to be being equal to the cardinality of the space of jump variables as applied to the finite dimensional matrices $(\tilde{A}, \tilde{B})$. In cases where this is violated or *nearly* violated even for large n, one should not proceed to use the resulting numerical solutions, in the same way one would discard such parameter values in the finite case. The additional continuity condition is harder to check, even asymptotically. Although a modulus of continuity from the derivatives to the solution can be approximated in some cases, and so when close to 0 can indicate violations of the continuity requirement, it is not guaranteed that all failures of continuity will result in a small estimated modulus, and so one should proceed with caution even if this indicator does not suggest any concerns.

If conjecture (2) holds, the output of the proposed algorithm then can be assured to satisfy approximation bounds comparable in structure and in rate to those satisfied by the approximate functional derivatives. Intuitively, as the solution is then continuous in operator norm, operator norm convergence of the representation of the functional derivatives in the space of functions implies that the approximate solution operators generated by the same mapping from matrices to the space of functions applied to the derivatives maps the solution matrices g_x, h_x to operators which also converge in operator norm to the true functional derivatives of the solutions. Moreover, by the same guarantees relating the transformed representation to the known representation of functions in terms of histograms over evenly spaced intervals for smooth functions, this implies that over the space of appropriately smooth initial conditions $x \in \Lambda_x$, the solution maps produce accurate approximations of the response of the economy to these states. Stated precisely:

Proposition 3. *Assume that the conditions for Lemma (1) hold and that the model parameters are such that Conjecture (2) is true. Then, there exists some C_2, C_3, N such that for $n > N$*

For the same histogram representation maps operators $P_n^x : \mathcal{H}_x \to \mathbb{R}^{2n+1}$, $P_n^y : \mathcal{H}_y \to \mathbb{R}^{2n-1}$ as in Lemma (1)

$$\sup_{\{x \in \Lambda_x^\alpha : \|x\|_{\mathcal{H}_x} = 1\}} \|[P_n^{y*} \tilde{g}_x P_n^x - g_x](x)\|_{\mathcal{H}_y} \leq C_2 n^{-\frac{1}{4}}$$

$$\sup_{\{x \in \Lambda_x^\alpha : \|x\|_{\mathcal{H}_x} = 1\}} \left\|[P_n^{x*} \tilde{h}_x P_n^x - h_x](x)\right\|_{\mathcal{H}_x} \leq C_3 n^{-\frac{1}{4}}$$

Remark. Provided the conditions hold, for any sufficiently smooth function, the histogram representation of that function when passed through the solution matrices produced by the algorithm can be compared to the true response generated by the model to that perturbation, and the distributions of wealth, consumption policy rules, and aggregate capital generated by the numerical approximation should differ by an amount which is small, and which becomes more accurate as a finer discretization is applied.

Often, beyond the policy functions themselves, objects of economic interest include maps which are defined by repeated application of the policy function, such as impulse response functions. Because the above approximation results describe accuracy over a limited class of inputs, proposition (3) does not directly imply that impulse response functions beyond the initial response are calculated accurately. Nevertheless, the same results can be extended to iterated applications of the policy maps, and so can provide numerical convergence results for IRFs.

Proposition 4. *Assume that the conditions for Proposition (3) hold. Then there exists some C_4, N such that for $n > N$, for any non-negative integer k*

$$\sup_{\{x \in \Lambda_x^\alpha : \|x\|_{\mathcal{H}_x} = 1\}} \left\|[P_n^{y*} \tilde{h}_x^k P_n^x - h_x^k](x)\right\|_{\mathcal{H}_y} \leq C_4 n^{-\frac{1}{4}}$$

$$\sup_{\{x \in \Lambda_x^\alpha : \|x\|_{\mathcal{H}_x} = 1\}} \left\|[P_n^{y*} \tilde{g}_x \tilde{h}_x^k P_n^x - g_x h_x^k](x)\right\|_{\mathcal{H}_y} \leq C_5 n^{-\frac{1}{4}}$$

5 Numerical Results

Our parameterization of the model is relatively standard. We set $\beta = 0.95$, $\gamma = 3$, $\alpha = 1/3$, $\delta = 0.2$. The aggregate productivity shock has mean 1, persistence $\rho = 0.95$ and standard deviation $\sigma = 0.01$. The idiosyncratic shock to households' labor endowment is the sum of two i.i.d. shocks. The first shock is a truncated log-normal with variance 0.5, lower bound 0.5 and upper bound 1.5. The distribution of the second shock is a mollifier: a smooth function with compact support.[10] We calculate integrals over the log-normal component of skill s using Curtis-Clenshaw quadrature with a grid of 50 points. We use 160 grid points for cash on hand w.

Figure 1a plots the steady state distribution of cash on hand $\mu(w)$. The distribution is modestly skewed to the right. Figure 1b plots the steady state consumption policy function. As is standard, the consumption function is concave: households at the borrowing limit have a marginal propensity to consume (MPC) $c'(w)$ equal to 1; households slightly above the limit have an MPC which is high but below 1; and households with much higher wealth have a substantially lower MPC.

Figure 2 plots the impulse responses of aggregate variables to a 1 standard deviation TFP shock ε_t. While these aggregate variables are not all explicitly included in the model, it is straightforward to compute their impulse response functions by taking a first order approximation of aggregate variables as a function of the endogenous functional variables. For example, aggregate consumption can be defined as

$$C_t = \int c_t(w)\mu_t(w)dw$$

and approximated as

$$C_t = \iota \sum_{i=1}^{n} c_t(w_i)\mu_t(w_i)$$

Linearizing this expression, we have

$$\hat{C}_t = \iota \sum_{i=1}^{n} \hat{c}_t(w_i)\mu_t(w_i) + \iota \sum_{i=1}^{n} c(w_i)\hat{\mu}_t(w_i)$$

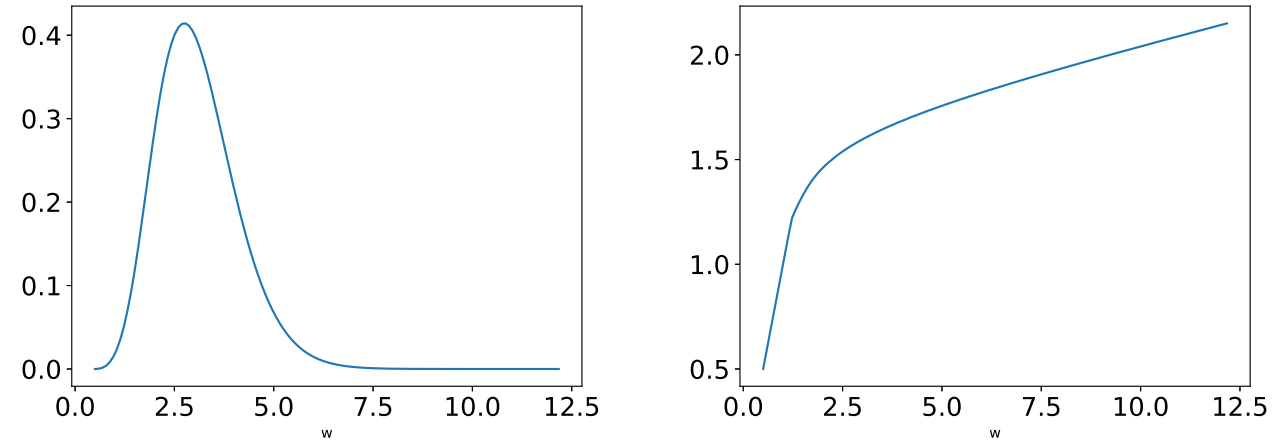
where hats denote deviations from steady state and functions without a t subscript denote steady state values.

Since our model features a distribution of households, we can also display how the whole distributions of consumption and wealth respond to this shock. Figure 3 shows the 'impulse response surface' of the cash on hand distribution to the TFP shock. That is, at each time t, this figure plots the deviation of $\mu_t(w)$ from steady state $\mu(w)$ as a function of w. On impact, the distribution of cash on hand shifts to the right, with a decline in the mass of the distribution near 2 (to the left of the steady state mean) and an increase in the mass of the distribution near 4 (to the right of the steady state mean).

[10] Formally, the second shock has distribution

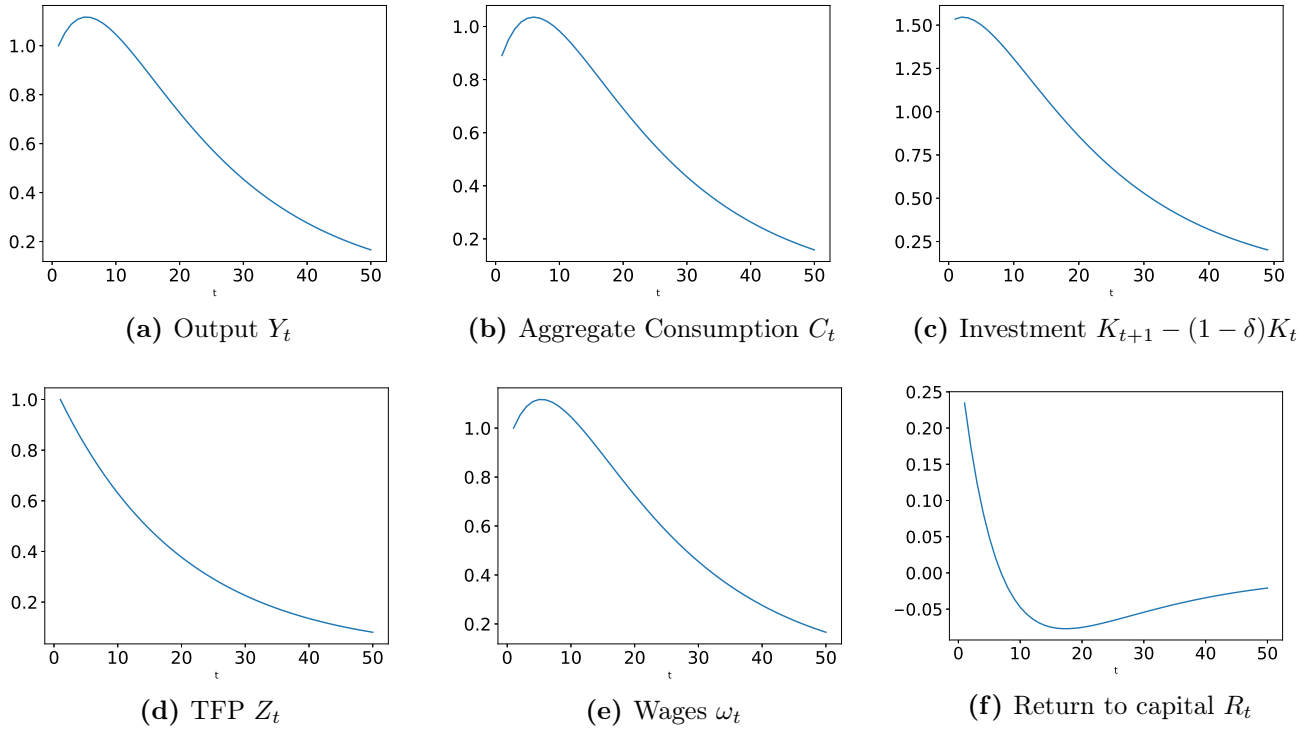
$$\varphi(z) = \left(\frac{2}{\bar{z} - \underline{z}}\right) \exp\left\{\frac{-1}{1 - \left(-1 + 2\frac{z - \underline{z}}{\bar{z} - \underline{z}}\right)^2}\right\} \frac{1}{I_n}$$

with support $[\underline{z}, \bar{z}]$, where the numerical constant I_n ensures normalization. We set $\underline{z} = 0$, $\bar{z} = 2$.



(a) Distribution of cash on hand $\mu(w)$ **(b)** Consumption Decision Rule $c(w)$

Figure 1. Steady State



(a) Output Y_t **(b)** Aggregate Consumption C_t **(c)** Investment $K_{t+1} - (1-\delta)K_t$

(d) TFP Z_t **(e)** Wages ω_t **(f)** Return to capital R_t

Figure 2. Impulse response of aggregate variables to a unit standard deviation increase in TFP. All figures show percentage deviations from steady state.

We compute the impulse response surface of the distribution of consumption as follows. Define a grid for consumption, $\{c_1, ..., c_m\}$. Approximate the mass of households in bin j at time t as

$$\iota \sum_{i=1}^{n} \mathbf{1}(c_j < c_t(w_i) \leq c_{j+1}) \mu_t(w_i)$$

where $c_t(w_i)$ is approximated as $c(w_i) + \hat{c}_t(w_i)$ and $\mu_t(w_i)$ is approximated as $\mu(w_i) + \mu c_t(w_i)$. Figure 4 plots

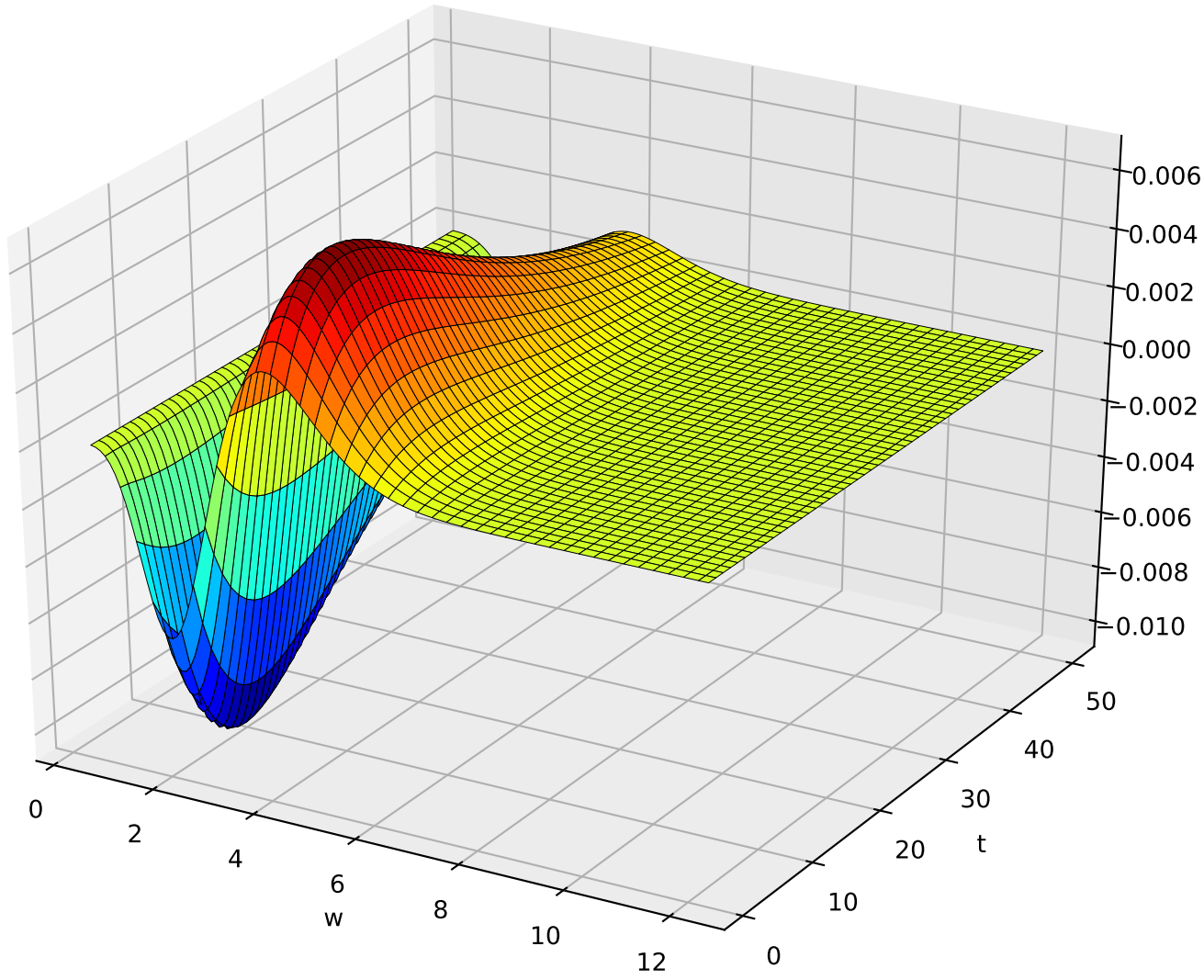


Figure 3. Impulse response of the distribution of cash on hand to an increase in TFP

the deviation of this object from its steady state value. While the mass of the consumption distribution shifts to the right, the shift is smaller than the shift in the cash on hand distribution.

Table 1 reports the time our algorithm takes to solve for steady state, compute derivatives and perform generalized Schur decomposition, for various numbers of cash on hand grid points n and with both analytical and automatic differentiation.[11] Using analytical derivatives, the slowest part of the code is the computation of steady state. Automatic differentiation takes substantially longer, especially for larger numbers of grid points.

[11] These computations were performed on a Windows machine with an Intel Core i7-7600U CPU 2.80 GHz processor with 4 cores and 16 GB of RAM running Windows 10 Enterprise.

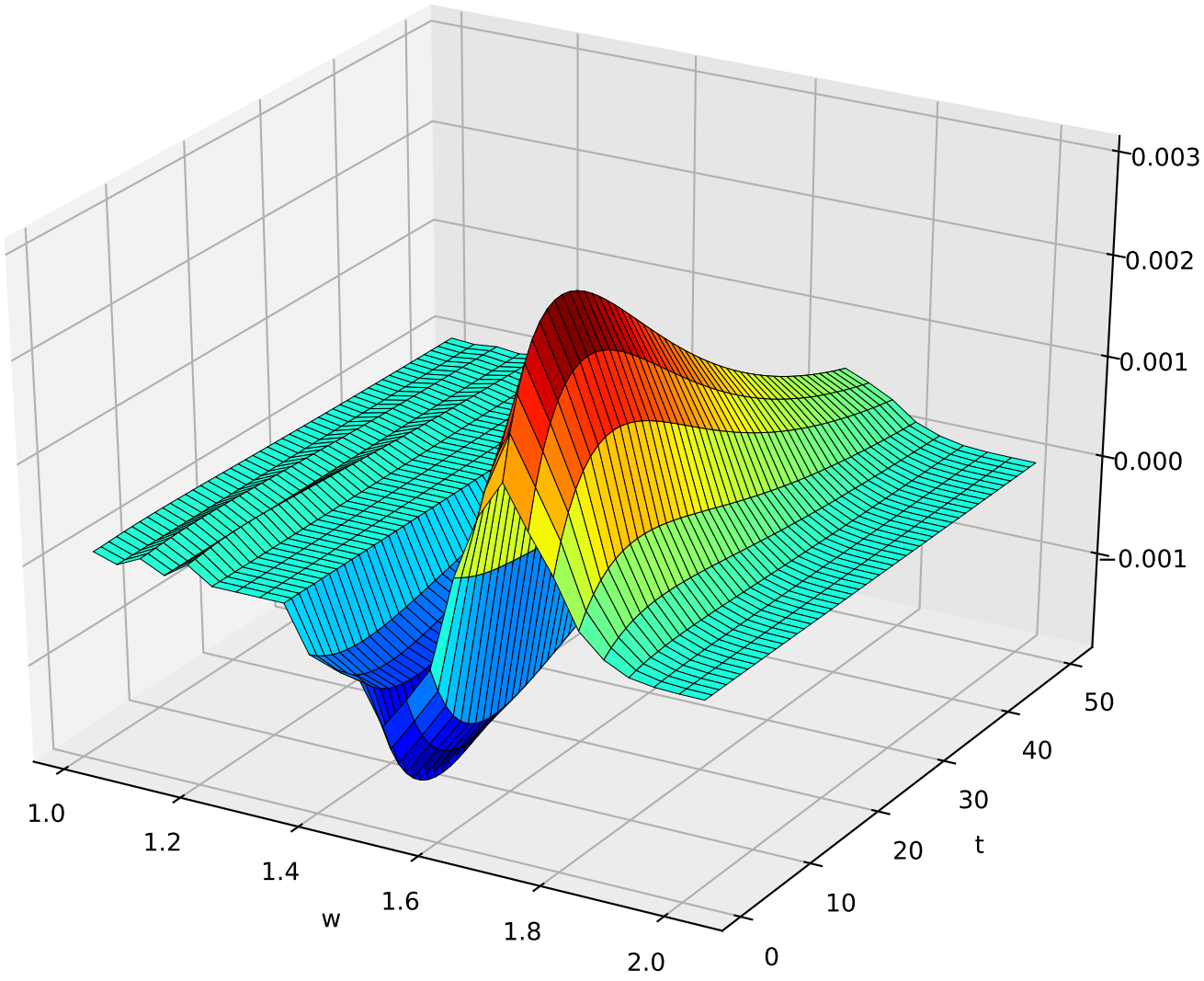


Figure 4. Impulse response of the distribution of consumption to an increase in TFP

Table 1
Run time (seconds)

	Analytical derivatives			Automatic differentiation		
n	80	160	320	80	160	320
steady state	2.3	9.4	35.8	2.2	8.3	33.0
derivatives	0.5	2.9	7.4	73.8	367.6	2313.6
Schur decomposition	0.3	2.4	19.9	0.3	2.2	20.3
Total	3.4	15.1	63.4	76.6	378.4	2367.8

6 Conclusion

In this paper we have described how perturbation methods, similar to those used to solve representative agent DSGE models, can be applied to solve a model with both aggregate and idiosyncratic risk and a continuum of agents. While the method was described in a variant of the Krusell and Smith (1998)

economy, it can easily be applied to a broader class of heterogeneous agent models.

References

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf, "When Inequality Matters for Macro and Macro Matters for Inequality," in "NBER Macroeconomics Annual 2017, volume 32" NBER Chapters, National Bureau of Economic Research, Inc, 2017.

Aiyagari, S. Rao, "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

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

Canay, Ivan A, Andres Santos, and Azeem M Shaikh, "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 2013, *81* (6), 2535–2559.

Childers, David, "Automated Solution of Heterogeneous Agent Models," Technical Report, Carnegie Mellon University 2018.

_ , "Solution of Rational Expectations Models with Function Valued States," Technical Report, Carnegie Mellon University 2018.

Johnstone, Iain M., *Gaussian estimation: Sequence and wavelet models*, Unpublished manuscript, 2013.

Klein, Paul, "Using the generalized Schur form to solve a multivariate linear rational expectations model," *Journal of Economic Dynamics and Control*, September 2000, *24* (10), 1405–1423.

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

Mallat, Stephane, *A wavelet tour of signal processing: the sparse way*, third ed., Academic press, 2008.

Nickl, Richard, *Statistical Theory* 2013.

Reiter, Michael, "Solving heterogeneous-agent models by projection and perturbation," *Journal of Economic Dynamics and Control*, March 2009, *33* (3), 649–665.

Revels, Jarrett, Miles Lubin, and Theodore Papamarkou, "Forward-Mode Automatic Differentiation in Julia," *CoRR*, 2016, *abs/1607.07892.*

Schmitt-Grohe, Stephanie and Martin Uribe, "Solving dynamic general equilibrium models using a second-order approximation to the policy function," *Journal of Economic Dynamics and Control*, January 2004, *28* (4), 755–775.

Stewart, Gilbert W and Ji guang Sun, "Matrix perturbation theory," 1990.

Winberry, Thomas, "A Method for Solving and Estimating Heterogeneous Agent Macro Models," Technical Report January 2018.

Appendix

A Functional Derivatives

In this appendix we report the derivatives of our discretized functional equations. For convenience, we repeat the functional equations:

$$\ell_t(w_i) = \beta \mathbb{E}_t \left[R(Z_{t+1}, K_{t+1}) \iota \sum_{j=1}^{n} \frac{c_{t+1}(w_j)^{-\gamma}}{\omega(Z_{t+1}, K_{t+1})} g\left(\frac{w_j - R(Z_{t+1}, K_{t+1})(w_i - c_t(w_i))}{\omega(Z_t, K_t)} \right) \right] \tag{21}$$

$$\mu_t(w_j) = \iota \sum_{i=1}^{n} \frac{1}{\omega(Z_t, K_t)} g\left(\frac{w_j - R(Z_t, K_t)(w_i - Lc_t(w_i))}{\omega(Z_t, K_t)} \right) L\mu_t(w_i) \tag{22}$$

$$L\mu_{t+1}(w_i) = \mu_t(w_i), i = 1, ..., n \tag{23}$$

$$L\ell_{t+1}(w_i) = \ell_t(w_i), i = 1, ..., n \tag{24}$$

$$K_{t+1} = \iota \sum_{i=1}^{n} \mu_t(w_i)(w_i - c_t(w_i)) \tag{25}$$

$$\ln Z_t = \rho_Z \ln Z_{t-1} + \sigma \varepsilon_t \tag{26}$$

In what follows, hats denote deviations from steady state values. The linearized Euler equation (21) is

$$\left[1 + \beta R \iota \sum_{j=1}^{n} \frac{c(w_j)^{-\gamma}}{\omega} g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) \frac{R}{\gamma \omega} \mathbf{1}(\ell(w_j)^{-1/\gamma} < w_j) \right] \hat{\ell}_t(w_i)$$

$$= \left[\beta \iota \sum_{j=1}^{n} \frac{c(w_j)^{-\gamma}}{\omega} \left(g\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) + g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) \frac{w_i - c(w_i)}{\omega} \right) \right] (R_Z \mathbb{E}_t \hat{Z}_{t+1} + R_K \hat{K}_{t+1})$$

$$+ \beta R \iota \sum_{j=1}^{n} \frac{1}{\omega} \mathbf{1}(\ell(w_j)^{-1/\gamma} < w_j) g\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) \hat{\ell}_{t+1}(w_j)$$

$$- \left[\beta R \iota \sum_{j=1}^{n} \frac{c(w_j)^{-\gamma}}{\omega^2} g\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) \right] (\omega_Z \mathbb{E}_t \hat{Z}_{t+1} + \omega_K \hat{K}_{t+1})$$

$$- \left[\beta R \iota \sum_{j=1}^{n} \frac{c(w_j)^{-\gamma}}{\omega^3} + g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega} \right) (w_j - R(w_i - c(w_i))) \right] (\omega_Z \mathbb{E}_t \hat{Z}_{t+1} + \omega_K \hat{K}_{t+1})$$

where $R_Z, R_K, \omega_Z, \omega_K$ denote the derivatives of the factor prices (7) and (8) with respect to their arguments.

The linearized Kolmogorov forward equation (22) is

$$\begin{aligned}
\hat{\mu}_t(w_j) &= -\left[\iota\sum_{i=1}^{n}\frac{1}{\omega^2}g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega}\right)(w_i - c(w_i))\mu(w_i)\right](R_Z\hat{Z}_t + R_K\hat{K}_t) \\
&\quad -\left[\iota\sum_{i=1}^{n}\frac{\mu(w_i)}{\omega^2}\left(g\left(\frac{w_j - R(w_i - c(w_i))}{\omega}\right) + g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega}\right)\frac{w_i - c(w_i)}{\omega}\right)\right](\omega_Z\hat{Z}_t + \omega_K\hat{K}_t) \\
&\quad -\iota\sum_{i=1}^{n}\frac{1}{\omega^2}g'\left(\frac{w_j - R(w_i - c(w_i))}{\omega}\right)R\mu(w_i)\frac{1}{\gamma}\ell(w_i)^{-\frac{1}{\gamma}-1}\mathbf{1}(\ell(w_j)^{-1/\gamma} < w_j)\widehat{L\ell}_t(w_i) \\
&\quad +\iota\sum_{i=1}^{n}\frac{1}{\omega}g\left(\frac{w_j - R(w_i - c(w_i))}{\omega}\right)\widehat{L\mu}_t(w_i)
\end{aligned}$$

The remaining linearized equations are:

$$\begin{aligned}
\widehat{L\mu}_{t+1}(w_i) &= \hat{\mu}_t(w_i) \\
\widehat{L\ell}_{t+1}(w_i) &= \hat{\ell}_t(w_i) \\
\hat{K}_{t+1} &= \iota\sum_{i=1}^{n}[\hat{\mu}_t(w_i)(w_i - c(w_i)) + \frac{1}{\gamma}\mu(w_i)\ell(w_i)^{-\frac{1}{\gamma}-1}\mathbf{1}(\ell(w_j)^{-1/\gamma} < w_j)\hat{\ell}_t(w_i)] \\
\mathbb{E}_t\hat{Z}_{t+1} &= \rho_Z\hat{Z}_t
\end{aligned}$$

B Diagnostic Criteria

While the conditions in Condition (1) regarding continuity and stability of the solution with respect to the approximated operators are properties of the exact model and cannot in general be verified analytically, under certain conditions the values of the properties which must be checked are consistently approximable, and so can be checked numerically by diagnostic criteria which can asymptotically detect any violation. In particular, they are properties of the spectrum of the operator pair (A, B), which can be approximated consistently by the spectrum of $(\hat{A}, \hat{B}) := (\tilde{A}^m + \tilde{A}^\perp, \tilde{B}^m + \tilde{B}^\perp)$, where $(\tilde{A}^\perp, \tilde{B}^\perp)$ are additional components of the functional derivatives orthogonal to the finite space of histograms spanned by the original representations as defined in Childers (2018b), which can be calculated analytically, though in practice do not need to be computed for any of the results in the body of this paper. This permits a set of tests which, for a subset of models, have power to reject the validity of the criterion. If any of these tests fails, one should treat the results as suspicious. The absence of a warning, however, cannot be interpreted as indicating validity. Let ζ_n be an upper bound on the operator norm convergence rate $(\hat{A}, \hat{B}) \to (A, B)$; this can be guaranteed without reference to Condition (1).

Criterion 5. *Spectral Gap: Let $\sigma(\hat{A}+\tilde{A}^\perp, \hat{B}+\tilde{B}^\perp)$ denote the generalized spectrum of $(\hat{A}, \hat{B})$, Γ the complex unit circle, and $d(\lambda, S)$ for $S \subset \mathbb{C}_\infty$ the minimum distance in the extended complex plane between a point and a set. Let γ_n be a bound on $\chi(\sigma(\hat{A} + \tilde{A}^\perp, \hat{B} + \tilde{B}^\perp), \sigma(A, B))$ the maximum minimum distance between the spectra of the two operator pairs, which is guaranteed to exist by continuity of the generalized spectrum (Stewart and Sun (1990) Ch VI). Choose $\epsilon_n \searrow 0$ at a rate such that $\frac{\epsilon_n}{\gamma_n} \to \infty$. Define $z_n = \inf\{d(\lambda, \Gamma) : \lambda \in \sigma(\hat{A}, \hat{B})\}$. Reject Condition (1)(i) if $z_n < \epsilon_n$.*

Claim. If $\sigma(A,B)\cap\Gamma\neq\emptyset$, $\exists N$ such that $\forall n>N$, Criterion (5) rejects. Further, suppose Condition (1)(i) is true, then $\exists N$ such that $\forall n>N$, Criterion (5) does not reject.

Essentially, this says that whether there exists an eigenvalue in the unit circle can be tested by extending out the circle slightly. For close enough approximations, the eigenvalues of the true and approximate derivatives will be close enough to distinguish whether there is an eigenvalue in a small neighborhood of this value. By letting this neighborhood contract at a rate slower than the convergence rate of the operators themselves, one can ensure that any element of the spectrum which is truly bounded away from the unit circle will eventually be detected as such. Note that by orthogonality and diagonality, $\sigma(\hat{A}+\tilde{A}^{\perp},\hat{B}+\tilde{B}^{\perp})=\sigma(\hat{A},\hat{B})\cup\sigma(\tilde{A}^{\perp},\tilde{B}^{\perp})$, so one can construct the spectra separately for each component when performing the test.

If Criterion (5) fails, then one should not proceed to the other two; the other conditions will not even be well defined.

If Criterion (5) does not reject, one would also like to be able to verify Condition (1)(ii). Unfortunately, this condition is only testable in one direction, in the sense that we can construct a test which, if the criterion is true, will not reject, but may not have power to detect all possible violations of the condition. As a result, a rejection should be taken as an absolute sign that there is an issue with the model, but a failure to reject should not be interpreted as an indication that the condition actually holds. The content of the condition is the claim that a certain constant is bounded away from 0; the test constructs an approximation of the constant. In cases where the condition holds, this estimator is consistent, and this constant will converge to its true value, above 0. However, if the condition does not hold, the estimate need not be consistent, and so while a small value, which could never occur if the truth were bounded away from 0, is a clear sign of failure, failure may also result in values which are far from 0.

Criterion 6. *Dif Constant. Let* $(\hat{S}_{11},\hat{T}_{11})$, $(\hat{S}_{22},\hat{T}_{22})$, *and* (S_{11},T_{11}), (S_{22},T_{22}), *denote the components inside and outside the unit circle, respectively of the generalized Schur decomposition of* $(\hat{A},\hat{B})$ *and* (A,B), *respectively, and let* $(\tilde{S}_{11}^{\perp},\tilde{T}_{11}^{\perp})$ *and* $(\tilde{S}_{22}^{\perp},\tilde{T}_{22}^{\perp})$ *denote the corresponding subcomponent of the decomposition of* $(\tilde{A}^{\perp},\tilde{B}^{\perp})$. *Letting* $\|(Q,P)\|_{\mathcal{B}}=\max(\{\|Q\|_{op},\|P\|_{op}\}$ $T(P,Q):=(QT_{11}-T_{22}P,QS_{11}-S_{22}P)$, $\hat{T}(P,Q):=(Q(\hat{T}_{11}+\tilde{T}_{11}^{\perp})-(\hat{T}_{22}+\tilde{T}_{22}^{\perp})P,Q(\hat{S}_{11}+\tilde{S}_{11}^{\perp})-(\hat{S}_{22}+\tilde{S}_{22}^{\perp})P)$. *Define* $\delta=\left\|T^{-1}\right\|_{\mathcal{B}}^{-1}$, *which is the true dif constant, and define* $\hat{\delta}_n=\left\|\hat{T}^{-1}\right\|_{\mathcal{B}}^{-1}$.

Choose $\epsilon_n\searrow 0$ *at a rate such that* $\frac{\epsilon_n}{\zeta_n}\to\infty$. *Reject Condition (1)(ii) if* $\hat{\delta}_n<\epsilon_n$.

Claim. Suppose Conditions (1)(i) and (1)(ii) are true. If $\delta>0$, $\exists N$ such that $\forall n>N$, Criterion (6) does not reject.

Note that no claim is made regarding what happens when Condition (1)(ii) is false, in which case $\delta=0$ or is not defined. However, because in the case where the condition is true, it will not reject, rejection must be due to a failure of the condition. If this test fails, one should assume that the solution is not reliable, and not proceed to the next test. Beyond rejection or lack thereof, the value of this test statistic may be informative as well. Following Childers (2018b) Appendix B Thm 3, if γ is the operator norm error in the approximation of the equilibrium conditions, can be bounded using standard function approximation tools, then the operator norm error in the solution is bounded by a term which asymptotes to $\frac{2\gamma}{\delta}$. This means that a consistent approximation of δ can provide a bound the constant term in the solution error. If this value is small, not only might it suggest the possibility that conditions for consistency do not hold; it also

suggests that even if they do, the quality of the approximation may be poor. So, small values of the test statistic which do not meet the threshold for rejection may also be indicators of low solution quality.

While a test which does not have power to reject in all cases when the solution fails may not provide full confidence in results, there may be reason not to worry. In the case where a model is finite dimensional, whenever Condition (1)(i) holds, Condition (1)(ii) also holds (Stewart and Sun (1990) Thm VI.1.11). The same is true for operator pairs which are diagonalizable, or for which either A or B is invertible, regardless of dimension. For such models, the condition is redundant and need not be tested separately from Condition (1)(i). It is possible that a version of this result also holds in the infinite dimensional case, at least for operator pairs, like those to which our method applies, which are approximately diagonal. While neither confirmation nor counterexamples are yet known to the authors regarding this conjecture, if it were to hold, then it would be sufficient to test Condition (1)(i) only, which can be done consistently. If this were true, the results of Criterion (6) could still be interpreted along the lines in Remark (B), but then no additional test would be needed beyond that of Criterion (5) to test the condition.

While Condition (1)(i) is consistently testable directly in any case in which the derivatives can be shown to converge, Condition (1)(iii), which requires the invertibility of a map between the space of jump variables and the space of stable eigenvalues, and so effectively tests for existence and uniqueness of equilibria, is consistently testable only under certain conditions. First, Condition (1)(i) and (1)(ii) must hold: this allows consistent approximation of the operator which must be inverted. Second, the component of the approximation which lies orthogonal to the span of the basis functions used, defined in terms of $(\tilde{A}^{\perp}, \tilde{B}^{\perp})$ only, must be well behaved, in the sense that its inverse must be bounded away from 0. This allows avoiding the case where the inverse we would like to verify exists is ill posed. Canay et al. (2013) show that in such a situation, no consistent test of invertibility can be constructed. Here, because we work with a potentially broader set of operators, in some cases informative tests can be constructed. Fortunately, this condition can be verified analytically, and in the Krusell Smith model defined here does not hold, so in this model the test is consistent, at least if the other two conditions hold.

Criterion 7. *Stable spectrum isomorphism: Choose $\epsilon_n \searrow 0$ at a rate such that $\frac{\epsilon_n}{\zeta_n} \to \infty$. Let $\hat{U}_{22}$ and U_{22} denote the component of the generalized Schur decomposition of $(\hat{A}, \hat{B})$ and (A, B), respectively corresponding to the component of the spectrum outside the unit circle acting on the space of jump variables y, and let $\tilde{U}_{22}^{\perp}$ denote the corresponding subcomponent of the decomposition of $(\tilde{A}^{\perp}, \tilde{B}^{\perp})$. Suppose $\left\|\tilde{U}_{22}^{\perp -1}\right\|_{op}^{-1} > 0$. Define $z_n = \left\|\hat{U}_{22}^{-1}\right\|_{op}^{-1}$. Reject Condition (1)(iii) if $z_n < \epsilon_n$.*

Claim. Suppose Conditions (1)(i) and (1)(ii) are true. If $\left\|\tilde{U}_{22}^{\perp -1}\right\|_{op}^{-1} > 0$ and U_{22} does not have bounded inverse. Then $\exists N$ such that $\forall n > N$, Criterion (7) rejects. Further, if $\left\|\tilde{U}_{22}^{\perp -1}\right\|_{op}^{-1} > 0$ and $\left\|U_{22}^{-1}\right\|_{op}^{-1} \neq 0$, then $\exists N$ such that $\forall n > N$, Criterion (7) does not reject.

Note: the condition and claim make no statements about what happens when $\left\|\tilde{U}_{22}^{\perp -1}\right\|_{op}^{-1} = 0$ or does not exist. In this case, no information is conveyed by the event $z_n < \epsilon_n$; this relates to the impossibility result for testing invertibility of certain operators shown by Canay et al. (2013).

B.1 Proofs of Claims

Proof. of claim regarding Criterion (5)

By assumption, $\chi(\sigma(\hat{A}+\tilde{A}^{\perp},\hat{B}+\tilde{B}^{\perp}),\sigma(A,B)) \leq \gamma_n$ for large enough n so by continuity of the generalized spectrum , $\exists N_1$ and constant C such that $\forall n > N_1 \sup_{\lambda\in\sigma(\hat{A}+\tilde{A}^{\perp},\hat{B}+\tilde{B}^{\perp})} \inf d(\lambda,\sigma(A,B)) \leq C\gamma_n$.

If Condition (1)(i) is true, then, by the fact that the resolvent set is open, $\inf d(\sigma(A,B),\Gamma) = \delta > 0$.

By the triangle inequality, $|z_n - \delta| \leq C\gamma_n$ Then $z_n - \epsilon_n \geq \delta - C\gamma_n - \epsilon_n \to \delta > 0$. So, $\exists N$ such that $\forall n \geq N$, the test does not reject.

If Condition (1)(i) is false, then, $\inf d(\sigma(A,B),\Gamma) = 0$, so by the triangle inequality, $z_n \leq C\gamma_n$ Then $z_n - \epsilon_n \leq C\gamma_n - \epsilon_n$. By assumption, $\frac{\epsilon_n}{\zeta_n} \to \infty$, so, for any C, there exists N such that $\forall n \geq N$, $C\gamma_n - \epsilon_n \leq 0$ and the test rejects. □

Proof. of claim regarding Criterion (6)

Suppose Condition (1)(i) is true and $\left\|T^{-1}\right\|_{\mathcal{B}}^{-1} = \delta > 0$. Then, by Childers (2018b) Appendix B Thm 4,

$\max\{\left\|\hat{T}_{11}+\tilde{T}_{11}^{\perp}-T_{11}\right\|_{op}, \left\|\hat{T}_{22}+\tilde{T}_{22}^{\perp}-T_{22}\right\|_{op}, \left\|\hat{S}_{11}+\tilde{S}_{11}^{\perp}-S_{11}\right\|_{op}, \left\|\hat{S}_{22}+\tilde{S}_{22}^{\perp}-S_{22}\right\|_{op}\} \leq C\zeta_n$ for $n \geq N_1$, for some N_1.

As a result, for $n \geq N_1$

$$\begin{aligned}\left\|\hat{T}(P,Q)-T(P,Q)\right\|_{\mathcal{B}} &= \max\{\left\|Q(\hat{T}_{11}+\tilde{T}_{11}^{\perp}-T_{11})-(\hat{T}_{22}+\tilde{T}_{22}^{\perp}-T_{22})P\right\|_{op}, \\ &\quad \left\|Q(\hat{S}_{11}+\tilde{S}_{11}^{\perp}-S_{11})-(\hat{S}_{22}+\tilde{S}_{22}^{\perp}-S_{22})P)\right\|_{op}\} \\ &\leq C\zeta_n \left\|(P,Q)\right\|_{\mathcal{B}}\end{aligned}$$

So by $\left\|T^{-1}\right\|_{\mathcal{B}}^{-1} = \delta > 0$, there exists n large enough such that $\left\|\hat{T}^{-1}\right\|_{\mathcal{B}}$ exists, and

$$\begin{aligned}\left|\left\|\hat{T}^{-1}\right\|_{\mathcal{B}}^{-1} - \left\|T^{-1}\right\|_{\mathcal{B}}^{-1}\right| &\leq \frac{1}{\left\|\hat{T}^{-1}\right\|_{\mathcal{B}}\left\|T^{-1}\right\|_{\mathcal{B}}}\left|\left\|\hat{T}^{-1}\right\|_{\mathcal{B}} - \left\|T^{-1}\right\|_{\mathcal{B}}\right| \\ &\leq \frac{2}{\left\|\hat{T}^{-1}\right\|_{\mathcal{B}}\left\|T^{-1}\right\|_{\mathcal{B}}}\left\|\hat{T}(P,Q)-T(P,Q)\right\|_{\mathcal{B}} \\ &\leq C_2\zeta_n\end{aligned}$$

for some C_2. As a result, for $\hat{\delta}_n - \epsilon_n \geq \delta - C_2\zeta_n - \epsilon_n \to \delta > 0$. So, $\exists N$ such that for $n \geq N$, the test does not reject. □

Proof. of claim regarding Criterion (7).

Suppose Conditions (1)(i) and (1)(ii) are true and $\left\|\tilde{U}_{22}^{\perp-1}\right\|_{op}^{-1} = b > 0$ Then, by Childers (2018b) Appendix B Thm 4, $\exists N_1$ s.t. for $n \geq N_1$

$$\left\|\hat{U}_{22}+\tilde{U}_{22}^{\perp}-U_{22}\right\|_{op} \leq C\zeta_n$$

Suppose $\left\|U_{22}^{-1}\right\|_{op}^{-1} = d > 0$. Then for large enough n, $\hat{U}_{22}+\tilde{U}_{22}^{\perp}$ is invertible and for some C

$$\left\|(\hat{U}_{22}+\tilde{U}_{22}^{\perp})^{-1}-U_{22}^{-1}\right\|_{op}^{-1} \leq C\zeta_n$$

By orthogonality

$$(\hat{U}_{22} + \tilde{U}_{22}^{\perp})^{-1} = \hat{U}_{22}^{-1} + \tilde{U}_{22}^{\perp -1}$$

and

$$\left\|(\hat{U}_{22} + \tilde{U}_{22}^{\perp})^{-1}\right\|_{op}^{-1} \leq \min\{\left\|\hat{U}_{22}^{-1}\right\|_{op}^{-1}, \left\|\tilde{U}_{22}^{\perp -1}\right\|_{op}^{-1}\}$$

So by the triangle inequality

$$z_n - \epsilon_n \geq \left\|(\hat{U}_{22} + \tilde{U}_{22}^{\perp})^{-1}\right\|_{op}^{-1} - \epsilon_n \to d > 0$$

and so $\exists N$ st $\forall n \geq N$, the test does not reject.

Next suppose U_{22} does not have bounded inverse.

Then for n large, it must be that $\left\|(\hat{U}_{22} + \tilde{U}_{22}^{\perp})^{-1}\right\|_{op} \geq \frac{1}{C\zeta_n}$, and $\left\|\hat{U}_{22}^{-1}\right\|_{op} \geq \frac{1}{C\zeta_n} - b$. So

$$z_n - \epsilon_n \leq \frac{1}{\frac{1}{C\zeta_n} - b} - \epsilon_n$$

because $\frac{\epsilon_n}{\zeta_n} \nearrow \infty$, there must exist N such that for all $n \geq N$ the above term is less than 0, and so the test rejects. □

C Proofs of Propositions

To formalize the claim that the rescaling may be ignored provided in Section 4, we first demonstrate an auxiliary result, which shows that the procedure described in this paper produces results numerically equivalent to those of the algorithm in Childers (2018a), which includes as an additional step procedures to transform from grid points to coefficients before applying the rational expectations algorithm to the approximate model. The use of a histogram representation, for which the map from grid points to coefficients is simply an identity matrix rescaled by a constant, does no more than a rescaling, which is undone by the rational expectations solver, at least if the results are then placed on a consistent scale.

8. *Consider a model satisfying the conditions of Theorem (11) in Childers (2018a), using tensor product histogram approximations as interpolation schemes. Then, the guarantees of Theorem (11) are preserved (in a sense defined in the proof) if in Algorithm (3) in Childers (2018a), step 5 is replaced with step 5′ defined below.*

Algorithm. *Step 5′: $\forall \ell = 1 \dots d_2$, $\forall j = 1 \dots 2d_2$, represent the (ℓ, j) block of the equilibrium equations as $\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)} + \mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}$ (that is, as simply the functional derivatives of the equation with respect to the grid points)*

Proof. of lemma (8)

This result applies from an application of Childers (2018a) Lemma (28) to show that the representation in Step 5′ produces solution equivalent to the representation in Step 5 of the algorithm in Childers (2018a). Consider a block (ℓ, j) of form $(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)} + M_{[\ell^o]}(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}(\Pi^{[j]})^{-1}M^*_{[j]}$ or of form $\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)} + M_{[\ell^o]}\mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}(\Pi^{[j]})^{-1}M^*_{[j]}$ depending on whether Childers (2018a) Condition (5)(b)(ii) or (5)(b)(i)

is satisfied for row ℓ. Under a histogram approximation, $M_{[\ell^o]} = \frac{\sqrt{K_{[\ell^o]}}}{\sqrt{c_{\ell^o}}} I_{K_{[\ell^o]}}$, $M_{[j]} = \frac{\sqrt{K_{[j]}}}{\sqrt{c_j}} I_{K_{[j]}}$, and $\Pi^{[j]} = \frac{c_j}{K_{[j]}} I_{K_{[j]}}$, where $K_{[j]}$ and $K_{[\ell^o]}$ are the cardinality of the set of grid points over argument sets $[j]$ and $[\ell^o]$, respectively and c_j and c_{ℓ^o} are the width of the interval spanned by the corresponding functions. By Lemma (28), multiplying a row by an invertible matrix does not change the solution, and so after multiplying each row ℓ satisfying Condition (5)(b)(ii) by $\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}}$, each block (ℓ, j) of the algorithm using Step 5 is given by $\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)} + \frac{\sqrt{K_{[\ell^o]} c_j}}{\sqrt{K_{[j]} c_{\ell^o}}} \mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}$. Applying again Lemma (28), multiplying each row ℓ by $\frac{\sqrt{c_{\ell 0}}}{\sqrt{K_{[\ell^o]}}} I_{K_{[\ell^o]}}$ and each column j by $\frac{\sqrt{K_{[j]}}}{\sqrt{c_j}} I_{K_{[j]}}$ produces a solution for which h_x and g_x becomes $W_x^{-1} h_x W_x$ and $W_y^{-1} g_x W_y$, where W is $\frac{\sqrt{K_{[j]}}}{\sqrt{c_j}} I_{K_{[j]}}$ on the (j, j) block and 0 elsewhere, and W_x and W_y denote the submatrices corresponding to x and y variables. Further, note that $\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)}$ is nonzero only if $\frac{\sqrt{K_{[j]}}}{\sqrt{c_j}} = \frac{\sqrt{K_{[\ell^o]}}}{\sqrt{c_{\ell^o}}}$ for any model satisfying the conditions, as otherwise identity maps are not defined between these spaces. As a result, after these transformations, the (ℓ, j) block is of form $\mathcal{F}^{\ell}_{\overrightarrow{g}_j(out)} + \mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}$ exactly as in Step 5′. So, the solutions using either representation are equivalent up to multiplication by components of matrix W, which simply rescales each grid point according to the dimension of the space. □

Proof. of Lemma 1,

The proof of this lemma is nearly identical to the first part of the proof of Childers (2018a) Lemma (15), and so only the differences will be described. Childers (2018a) Lemma (15) first proceeds by verifying the sup norm convergence of the kernel function based on the properties of the approximation of the steady state, smoothness properties of the derivatives, and the convergence properties of the transformed interpolation scheme. the only difference here is that the discontinuous function in this model takes a different functional form, viz. $1\{\ell(w)^{-1/\gamma} < w\}$.

In order to define a transformation such that the boundary of the transformed histogram coincides with the location w^* of the discontinuity, note that by the assumed uniform convergence of $\ell(w_i)$ at rate $O(n^{-1})$ and either the monotonicity or an $o(n^{-1})$ convergence rate and the monotonicity of the true $\ell^*(w)$ which follows from standard results for consumption savings problems with fixed interest rate, for large enough n there exists a threshold grid point such that $1\{\ell(w_i)^{-1/\gamma} < w_i\}$ switches, and so one may construct a map $\tau^n(w)$ which stretches this grid point to the true boundary w^* and the transformed histogram interpolation representation $\tilde{\ell}(w)$ using this map satisfies $1\{\ell(w)^{-\frac{1}{\gamma}} < w\} = 1\{\tilde{\ell}(w)^{-\frac{1}{\gamma}} < w\}$.

Using the existence and accuracy properties of such a map following the argument in Childers (2018a) Lemma (15) ensures the convergence of the representation of the derivatives in terms of at convergence rate $O(n^{-1/4})$ over the space of transformed histograms of order n, or equivalently operator norm convergence to $(A, B) - (\tilde{A}^{\perp}, \tilde{B}^{\perp})$, where $(\tilde{A}^{\perp}, \tilde{B}^{\perp})$ described in Childers (2018b), can be computed analytically from the derivatives and extend the representation from the space of functions represented by a finite set of basis functions to the component of the full function spacce orthogonal to this space.

An application of the result bounding the error incurred by mapping the transformed histogram representation to the uniform histogram representation, Lemma (25) in Childers (2018a), obtain

$$\sup_{\{(x,y)\in\Lambda_x^{\alpha}\times\Lambda_y^{\alpha}:\|(x,y)\|_{\mathcal{H}_x\times\mathcal{H}_y}=1\}} \max\{\left\|[P_n^*\tilde{A}^m P_n - (A - \tilde{A}^{\perp})](x,y)\right\|_{\mathcal{H}_x\times\mathcal{H}_y}, \left\|[P_n^*\tilde{B}^m P_n - (B - \tilde{B}^{\perp})](x,y)\right\|_{\mathcal{H}_x\times\mathcal{H}_y}\}$$

$$\leq Cn^{-\frac{1}{4}}$$

Finally, steps analogous to those of the proof of Corollary (26) of Childers (2018a) demonstrate that $\sup_{\{(x,y)\in\Lambda_x^\alpha\times\Lambda_y^\alpha:\|(x,y)\|_{\mathcal{H}_x\times\mathcal{H}_y}=1\}} \max\{\left\|\tilde{A}^\perp(x,y)\right\|_{\mathcal{H}_x\times\mathcal{H}_y}, \left\|\tilde{B}^\perp(x,y)\right\|_{\mathcal{H}_x\times\mathcal{H}_y}\} \leq O(n^{-1/2})$ and so these terms can be removed, proving the result. □

Proof. of Propositions 3 and 4

By the proof of lemma 1, the representation of the functional derivatives from $(\tilde{A}^m, \tilde{B}^m)$ in terms of transformed histogram representation converges in operator norm at rate $O(n^{-1/4})$. By Lemma (8) and the invariance to left multiplaction, the algorithm produces solution matrices with operator representation identical to that which would be produced by application to matrices $(\tilde{A}^m, \tilde{B}^m)$. Therefore by the Conjecture, this procedure satisfies all the conditions of Childers (2018a) Theorem (11) and so converges in operator norm to the true solution operator, minus a term corresponding to the component orthogonal to the bases, which adds error of smaller order by the same reasoning as above. The convergence of the IRFs follows directly from Childers (2018a) Corollary (26). □