Efficient Computation of Heterogeneous Agent Models with Aggregate Shocks with an Application to Consumption Dynamics*

Proposal for NBER Macroeconomics Annual 2017

SeHyoun Ahn Benjamin Moll Greg Kaplan Thomas Winberry

July 7, 2016

*Ahn: Princeton University, e-mail sehyouna@princeton.edu, Kaplan: University of Chicago and NBER, e-mail gkaplan@uchicago.edu; Moll: Princeton University, CEPR and NBER, e-mail moll@princeton.edu; Winberry: University of Chicago, e-mail thomas.winberry@chicagobooth.edu. We thank Chris Sims for useful comments. Christian Wolf provided superb research assistance.

1 Introduction

Despite the tremendous degree of household and firm heterogeneity that is routinely found in micro datasets, most quantitative analyses of the effects of shocks and policy on the macroeconomy still employ a representative agent framework, particularly in central banks and other policy institutions. Applied macroeconomists tend to make two excuses for this abstraction. First, they appeal to the "approximate aggregation" hypothesis that, since the seminal work of Krusell and Smith (1998), has become almost folklore in quantitative macroeconomics. This hypothesis says that while heterogeneous agent economies are useful for studying the distributional consequences of shocks and policy, their predictions for aggregate outcomes are often very similar to corresponding models in which heterogeneity is assumed away.[1] Second, macroeconomists argue that because of their high dimensionality, the computational difficulties involved in solving, estimating and simulating heterogeneous agent economies render their use impractical. This lies in stark contrast to the speed and ease with which economists can analyze representative agent economies using widely available software.

Today these excuses hold less validity than they once did. Serious doubt has recently been cast on the scope of the approximate aggregation hypothesis. In economies with frictions (for example, in price setting, asset trading, investment or hiring), the presence of heterogeneity has been shown to substantially alter the aggregate consequences of both shocks and policy. In many cases, the differences between the heterogeneous agent and representative agent models are not merely quantitative in nature, but rather our entire intuition about the relative importance of different economic forces is altered. In other cases, authors have demonstrated the aggregate importance of shocks that simply cannot be adequately modeled in a representative agent setting.[2,3]

[1]For example, in Krusell and Smith (1998), which is basically a heterogenous agent version of a Real Business Cycle (RBC) model, the effects of technology shocks on aggregate output, consumption and investment are essentially the same as those in the standard RBC model

[2]Examples in which approximation aggregation fails on the household side are the recent papers that have have introduced heterogeneity into New Keynesian models (McKay, Nakamura and Steinsson, 2015; Kaplan, Moll and Violante, 2016) and find that intertemporal substitution is much weaker as a monetary transmission mechanism than in otherwise similar representative agent models. Examples of shocks that require a heterogeneous agent framework include Guerrieri and Lorenzoni (2011) who analyze a tightening of household credit constraints, and Bayer et al. (2015) who analyze a shock to the degree of idiosyncratic income risk faced by workers. On the firm side, Winberry (2016) studies a model in which approximate aggregation fails due to capital adjustment costs, and Bloom et al. (2014) study a shock to the dispersion of idiosyncratic productivity which requires a heterogeneous firm framework.

[3]Existing methods for solving these heterogeneous agent economies are substantially slower and more difficult to use than the standard linearization techniques that are typically used to solve representative agent economies (e.g. Dynare). Accordingly, these and other heterogeneous agent models take many shortcuts:

At the same time, there have been several advances in the computational techniques available for solving heterogeneous agent economies that have not yet found their way into the macroeconomist's standard toolkit. In this project, we leverage these advances to develop an efficient and easy-to-use computational method for solving a broad class of general equilibrium heterogeneous agent economies with aggregate shocks. Drawing on ideas from Campbell (1998) and Reiter (2009), we show how to apply familiar linearization techniques to heterogeneous agent models in a way that preserves the aggregate nonlinearities implied by movements in the distribution of individual households and firms. Conceptually, our procedure is straightforward: we compute a non-linear global approximation to the steady state of the discretized model without aggregate shocks, compute a first-order Taylor expansion in the aggregate state variables around this steady state, and solve the resulting linear system. In practice, the high dimensionality of heterogeneous agent models leads to many challenges, which we solve by combining existing ideas from the economics, mathematics, and computer science literatures. To encourage the wide adoption of our methodology – and therefore heterogeneous agent macroeconomics – we have developed a suite of `Matlab` codes implementing the key steps of our method in a model-free way. There are no more excuses; if a researcher can compute the steady state of their model, they can use our codes to compute the dynamics of the model with aggregate shocks with virtually no additional work.[4]

After developing our computational method, we will illustrate its usefulness by applying it to a richer environment that previously was computationally infeasible. The substantive issue that we will address is the dynamics of aggregate consumption. Twenty-seven years ago, in the NBER Macroeconomics annual, Campbell and Mankiw (1989) showed that the sensitivity of aggregate consumption to real interest rate changes is considerably smaller than suggested by a representative agent model, and that aggregate consumption is instead much more responsive to income changes. To address this inconsistency, they suggested a simple heterogeneous agent model (the "spender-saver" model) in which an exogenous fraction of households are hand-to-mouth and simply consume their disposable income in every period, while the remaining households are permanent income consumers. Christiano (1989), in his discussion of Campbell and Mankiw, points out that the spender-saver model generates an unconditional volatility of aggregate consumption that is too large relative to the

some study only non-stochastic steady states, others study perfect-foresight transitions between steady states, and the remainder use computational methods and algorithms that are highly specific to their particular setting. Often these "shortcuts" are substantially more involved than the general methods we propose – for example solving for perfect-foresight transitions can be more computationally burdensome than solving for the stochastic equilibrium using our proposed methods.

[4]Since our methods deliver a linear representation of the model, they can also be used to estimate the model using likelihood-based techniques as in the representative agent literature.

unconditional volatility of income, in stark contrast to the well-known empirical regularity that aggregate consumption is much smoother than aggregate income. Additionally, and as we discuss in more detail in the main text, the spender-saver model is inconsistent with microeconomic evidence on hand-to-mouth consumption status.

In our application, we will ask whether a realistic process for hand-to-mouth consumption status can reconcile Campbell and Mankiw's (1989) and Christiano's (1989) findings. Following Kaplan and Violante (2014), we will analyze an incomplete-markets model with both liquid and illiquid assets that retains the spirit of the Campbell and Mankiw (1989) insights, but is more consistent with microeconomic evidence. The model endogenously generates a class of "wealthy hand-to-mouth" consumers who hold a substantial amount of illiquid assets but nevertheless have high marginal propensities to consume. We will embed this model in general equilibrium as in Kaplan, Moll and Violante (2016), and incorporate aggregate shocks so as to be able to study the dynamics of aggregate consumption. Until now this last step has been computationally infeasible since the two-asset structure implies that the rich distribution of MPCs at the micro level matters for aggregate dynamics, violating approximate aggregation. However, the methodology developed in this paper can accurately and efficiently solve this model.

This framework holds great promise for being simultaneously consistent with the Campbell and Mankiw (1989) and Christiano (1989) findings. First, despite generating large average marginal propensities to consume, households are nonetheless forward looking rational agents who have a long-run smoothing motive, thereby likely resulting in aggregate consumption being less volatile than aggregate income. Second, Kaplan, Moll and Violante (2016) have shown that the framework generates much smaller direct responses of aggregate consumption to interest rate changes than in representative agent models, consistent with Campbell and Mankiw's observation that expected consumption growth is not associated with expected real interest rates.

The remainder of this proposal details how we will make our two main contributions. Section 2 describes the details of the computational method, which has already been developed, and uses the method to solve a simple Krusell and Smith (1998) model. Section 3 then describes the richer two-asset model and its application to the dynamics of aggregate consumption.

2 Overview of Computational Method

For the sake of concreteness, we describe our computational methods in the context of a simple version of the Krusell and Smith (1998) model. As we show in Section 3, our methods are applicable to a much broader class of models, including those in which approximate aggregation fails.

2.1 Continuous Time

Our methods require the model to be specified in continuous time. While discrete time models pose no conceptual difficulty (indeed, Campbell (1998) and Reiter (2009) originally proposed this general approach in discrete time), continuous time imparts three numerical advantages that we heavily exploit.

1. Continuous time naturally generates sparsity in the matrices that characterize the model's equilibrium conditions. In heterogeneous agent models, these matrices are extremely high-dimensional, so sparsity is essential in order to store and manipulate them.

2. Continuous time allows us to handle non-convexities in individual decision problems much more easily than in discrete time. In many heterogeneous agent environments, non-convexities are important for generating behavior that is consistent with micro data. Examples include menu costs in firms' pricing decisions, adjustment costs in firms' investment decisions, and adjustment costs in households' asset portfolio.

3. Continuous time allows us to more easily capture the nonlinear relationship between aggregate variables and the distribution of agents in the economy, despite using a first-order approximation to compute how individual decisions depend on the aggregate state. As we discuss in Section 2.4, the restrictions imposed by first-order approximations are most severe immediately upon impact of the shock, so the shorter is a model time period, the shorter the impact period.

2.2 Description of Krusell and Smith Model for Illustration

Environment There is a fixed mass of households $j \in [0,1]$ who have preferences represented by the expected utility function

$$\mathbb{E}_0 \int_0^\infty e^{-\rho t} \frac{c_{jt}^{1-\gamma}}{1-\gamma} dt,$$

where ρ is the rate of time preference and γ is the coefficient of relative risk aversion. At each instant t, a household's idiosyncratic labor productivity is $z_{jt} \in \{z_L, z_H\}$ with $z_L < z_H$. Households switch between the two values for labor productivity according to a Poisson process with arrival rates λ_L and λ_H.[5] A household with labor productivity z_{jt} earns labor income $w_t z_{jt}$. Markets are incomplete; households can only trade in one asset a_{jt}, productive capital, subject to the borrowing constraint $a_{jt} \geq \underline{a}$.

There is a representative firm which has access to the Cobb-Douglas production function

$$Y_t = e^{Z_t} K_t^\alpha N_t^{1-\alpha},$$

where Z_t is aggregate productivity, K_t is aggregate capital and N_t is aggregate efficiency units of labor. Aggregate productivity follows the Ornstein-Uhlenbeck process

$$dZ_t = -\nu Z_t dt + \sigma dW_t,$$

where dW_t is the innovation to a standard Brownian motion, ν captures the rate of mean reversion, and σ captures the size of innovations.

Equilibrium The household-level state variables are asset holdings a and the idiosyncratic labor productivity shock z. The aggregate state variables are the aggregate productivity shock Z_t and the cross-sectional distribution of households over their individual states $g_t(a,z)$. For notational convenience we denote an equilibrium object conditional on a particular realization of the aggregate state $(g_t(a,z), Z_t)$ with a subscript t. An equilibrium of

[5]The assumption that idiosyncratic shocks follow a Poisson process is for simplicity of exposition - the methods can also handle diffusion or jump-diffusion shock processes.

the model is characterized by the following equations:

$$\begin{aligned}
\rho v_t(a,z) &= \max_c \; u(c) + \partial_a v_t(a,z)\left(w_t z + r_t a - c\right) \\
&\quad + \lambda_z\left(v_t(a,z') - v_t(a,z)\right) + \frac{1}{dt}\mathbb{E}_t\left[dv_t(a,z)\right] && \text{(HJB)} \\
\partial_t g_t(a,z) &= -\partial_a\left[s_t(a,z)\, g_t(a,z)\right] - \lambda_z g_t(a,z) + \lambda_{z'} g_t(a,z') && \text{(KFE)} \\
dZ_t &= -\nu Z_t dt + \sigma dW_t && \text{(TFP)} \\
w_t &= (1-\alpha)\, e^{Z_t} K_t^{\alpha} \bar{N}^{-\alpha}, && \text{(AUX 1)} \\
r_t &= \alpha e^{Z_t} K_t^{\alpha-1} \bar{N}^{1-\alpha} - \delta, && \text{(AUX 2)} \\
K_t &= \int a g_t(a,z) da dz, && \text{(AUX 3)} \\
1 &= \int g_t(a,z) da dz, && \text{(AUX 4)}
\end{aligned}$$

and where $s_t(a,z)$ is the optimal saving policy function corresponding to the household optimization problem (HJB). For detailed derivations of these equations, see Achdou et al. (2015). (HJB) is the household's Hamilton-Jacobi-Bellman equation. Due to our use of time-dependent notation with respect to aggregate states, $\mathbb{E}_t$ denotes the conditional expectation with respect to aggregate states only. (KFE) describes how the mass of households at a point in the individual state space evolves over time. (TFP) describes the evolution of the aggregate shock. Lastly, (AUX 1) to (AUX 3) defines prices given the aggregate state, and (AUX 4) normalizes g to be a valid distribution.

2.3 Outline of the Method

Our method broadly follows the steps one would take to linearize a representative agent model. First, we solve for the steady state of the model without aggregate shocks. Second, we take a first-order Taylor expansion of the equilibrium conditions around that steady state, yielding a system of stochastic linear differential equations. Third, we solve the linear system of differential equations. Conceptually, extending each of these steps to the heterogeneous agent context is straightforward. However, the size of the system with heterogeneity leads to a number of computational challenges.

Step 1: Compute the Steady State We compute the Taylor expansion around the steady state with no aggregate shocks but a full set of idiosyncratic shocks. We use a fully non-linear approximation to solve for this steady state in order to retain the rich non-linearities and heterogeneity in the decision rules at the individual level. The steady state is

characterized by a stationary value function $v(a,z)$ and cross-sectional distribution $g(a,z)$. To solve for these objects, we use the finite difference methods outlined in Achdou et al. (2015). Finite difference methods are a fast, accurate, and robust approach to solving for the steady state; however, any other method to compute the steady state could also be used in this step.

Specifically, we approximate the value function and cross-sectional distribution over a discretized grid of asset holdings $\mathbf{a} = (a_1 = \underline{a}, a_2, ..., a_N)'$. The value function and distribution are then represented by their values along this discrete grid, which we denote by $\mathbf{v} = (v(a_1, z_L), ..., v(a_N, z_L), v(a_1, z_H), ..., v(a_N, z_H))'$ and $\mathbf{g} = (g(a_1, z_L), ..., g(a_N, z_H))'$. We then solve the (HJB) and (KFE) at each point along this discrete grid and approximate the derivatives in these equations with finite difference. If the finite difference is chosen in the correct way, the discretized steady state can be written compactly as the following system of matrix equations (see Achdou et al., 2015)[6]:

$$\rho \mathbf{v} = \mathbf{u}(\mathbf{v}) + \mathbf{A}(\mathbf{v}; \mathbf{p})\mathbf{v} \qquad \text{(HJB SS)}$$

$$\mathbf{0} = \mathbf{A}(\mathbf{v}; \mathbf{p})'\mathbf{g} \qquad \text{(KFE SS)}$$

$$\mathbf{p} = \mathbf{F}(\mathbf{g}). \qquad \text{(PRICE SS)}$$

(HJB SS) is the approximated (HJB) for each point on the discretized grid, collapsed into vector notation. The vector $\mathbf{u}(\mathbf{v}; \mathbf{p})$ is the maximized utility function over the grid, and the matrix multiplication $\mathbf{A}(\mathbf{v}; \mathbf{p})\mathbf{v}$ captures the remaining terms in (HJB). These terms are differences of the value function at different points in the grid, weighted by how fast households move around the individual state space. (KFE SS) is the approximated (KFE) in steady state and imposes that the distribution is stationary. Finally, (PRICE SS) defines the prices $\mathbf{p} = (r, w)'$ as a function of aggregate capital through the distribution $\mathbf{g}$.

The practical challenge in solving for the steady state values of $\mathbf{v}$ and $\mathbf{g}$ is the size of the system (HJB SS) to (PRICE SS). Since $\mathbf{v}$ and $\mathbf{g}$ each have $2N$ entries, where N is the size of the individual asset grid, the total system has $4N + 3$ equations in $4N + 2$ unknowns.[7] In simple models like this, highly accurate solutions can be obtained with as little as $N = 50$ if the grids are chosen carefully. However, models with more complex shock process and multiple state variables, N can easily grow into the tens of thousands. For such large systems, it is important to exploit the sparsity of the transition matrix $\mathbf{A}(\mathbf{v}; \mathbf{p})$ to

[6] Normalization condition (AUX 4) has been omitted to simplify the discussion.

[7] With K_t in (AUX 3) substituted into (AUX 1) and (AUX 2). Even though there is one more equation than unknowns, the solution still exists because $\mathbf{A}(\mathbf{v}; \mathbf{g})'$) in (KFE SS) is linearly dependent. Intuitively, (KFE SS) determines $\mathbf{g}$ only up to normalization.

efficiently solve the system.

Step 2: Compute First-Order Taylor Expansion The second step is to compute a first-order Taylor expansion of the model's discretized equilibrium conditions around the steady state equilibrium. With aggregate shocks, the discretized equilibrium takes the form:

$$\begin{aligned}
\rho \mathbf{v}_t &= \mathbf{u}\left(\mathbf{v}_t\right) + \mathbf{A}\left(\mathbf{v}_t;\mathbf{p}_t\right)\mathbf{v}_t + \frac{1}{dt}\mathbb{E}_t d\mathbf{v}_t \\
\frac{d\mathbf{g}_t}{dt} &= \mathbf{A}\left(\mathbf{v}_t;\mathbf{p}_t\right)'\mathbf{g}_t \\
\mathbf{p}_t &= \mathbf{F}\left(\mathbf{g_t}; Z_t\right) \\
dZ_t &= -\nu Z_t dt + \sigma dW_t.
\end{aligned} \qquad \text{(NON-LINEAR)}$$

Outside of steady state, the aggregate state $(\mathbf{g}_t, Z_t)$ fluctuates over time, inducing fluctuations in the marginal products of capital and labor, and therefore prices $\mathbf{p}_t = \mathbf{F}\left(\mathbf{g_t}; Z_t\right)$. Prices in turn induce fluctuations in households' decisions and therefore in $\mathbf{v}_t$ and the transition matrix $\mathbf{A}\left(\mathbf{v}_t; \mathbf{p}_t\right)$.

The challenge in computing the Taylor expansion is the size of the system ($4N + 2$ equations in $4N+2$ unknowns).[8] To compute the derivatives we use automatic differentiation, which is faster than symbolic or numerical differentiation and is accurate up to machine precision.[9] For this differentiation to be numerically feasible, it is essential to exploit the sparsity of the transition matrix $\mathbf{A}\left(\mathbf{v}_t;\mathbf{p}_t\right)$. To the best of our knowledge, there was no open-source automatic differentiation package for `Matlab` that exploits sparsity, so we wrote our own package that will be made available for others to use.

After taking these derivatives, the model's equilibrium conditions can be written in linear form:[10]

[8] After plugging in K_t, and dropping one grid point of $\mathbf{g}_t$ to capture the normalization condition (AUX 4).

[9] Automatic differentiation is a technique for numerical differentiation that works as follows. Any function is represented on a computer as a composition of various elementary functions with known derivatives, such as addition, multiplication, or exponentiation. Automatic differentiation evaluates the derivatives of the original function at the same time as evaluating the function itself by iteratively applying the chain rule.

[10] We have here taken the expectation of the entire system. This results in the loss of some information but also in more compact notation that is useful below. Alternatively, the system can be written with an expectation in the first equation only and an additional term involving σdW_t. To arrive at (LINEAR), we first rearrange (NON-LINEAR) so that all time derivatives are on the left-hand side. Second, we take expectation of the entire system (not just the first equation) using the fact that the expectation of a Brownian

$$\mathbb{E}_t \begin{bmatrix} d\widehat{\mathbf{v}}_t \\ d\widehat{\mathbf{g}}_t \\ \mathbf{0} \\ dZ_t \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{B}_{vv} & \mathbf{B}_{vg} & \mathbf{B}_{vp} & \mathbf{B}_{vZ} \\ \mathbf{B}_{gv} & \mathbf{B}_{gg} & \mathbf{B}_{gp} & \mathbf{B}_{gZ} \\ \mathbf{0} & \mathbf{B}_{pg} & \mathbf{I} & \mathbf{B}_{pZ} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\nu \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} \widehat{\mathbf{v}}_t \\ \widehat{\mathbf{g}}_t \\ \widehat{\mathbf{p}}_t \\ Z_t \end{bmatrix} dt \qquad \text{(LINEAR)}$$

The variables in the system – $\widehat{\mathbf{v}}_t$, $\widehat{\mathbf{g}}_t$, $\widehat{\mathbf{p}}_t$, and Z_t – are expressed as deviations from their steady state values, and the matrix $\mathbf{B}$ is composed of the derivatives of the equilibrium conditions evaluated at steady state. Since the pricing equations are static, the third row of this matrix equation only has positive entries on the right hand side.[11]

Step 3: Solve Linear System The final step of the method is to solve the system of stochastic linear differential equations (LINEAR). We follow standard procedures by performing a Schur decomposition of the matrix $\mathbf{B}$ to identify the stable and unstable roots of the system. If the Blanchard and Kahn (1980) condition holds – the number of stable roots equals the number of state variables – then we can form a linear solution that sets the control variables $\widehat{\mathbf{v}}_t$ as a linear function of the state variables $\widehat{\mathbf{g}}_t$ and Z_t:[12]

$$\begin{aligned} \widehat{\mathbf{v}}_t &= \mathbf{D}_{vg}\widehat{\mathbf{g}}_t + \mathbf{D}_{vZ}Z_t \\ \widehat{\mathbf{p}}_t &= \mathbf{D}_{pg}\widehat{\mathbf{g}}_t + \mathbf{D}_{pZ}Z_t. \end{aligned} \qquad \text{(SOLUTION)}$$

Again, the practical challenge in solving the system (LINEAR) is its size ($4N+2$ equations in $4N+2$ unknowns). In this simple Krusell and Smith model, standard linear solvers based on matrix decompositions work fine for grid sizes N necessary to achieve acceptable accuracy.

increment is zero $\mathbb{E}_t[dW_t] = 0$, to write (NON-LINEAR) compactly without the stochastic term as

$$\mathbb{E}_t \begin{bmatrix} d\mathbf{v}_t \\ d\mathbf{g}_t \\ \mathbf{0} \\ dZ_t \end{bmatrix} = \begin{bmatrix} \mathbf{u}(\mathbf{v}_t;\mathbf{p}_t) + \mathbf{A}(\mathbf{v}_t;\mathbf{p}_t)\mathbf{v}_t - \rho\mathbf{v}_t \\ \mathbf{A}(\mathbf{v}_t;\mathbf{p}_t)'\mathbf{g}_t \\ \mathbf{F}(\mathbf{g}_t;Z_t) - \mathbf{p}_t \\ -\nu Z_t \end{bmatrix} dt.$$

Finally, we linearize this system to arrive at (LINEAR).

[11] The fact that we can express prices as a static function of $\widehat{\mathbf{g}}_t$ and Z_t is a special feature of the Krusell and Smith model. More generally, the market clearing conditions involve integrals of decisions against the distribution, and only define prices implicitly. In those cases, one must approximate the time derivative of the market clearing condition (rather than the static market clearing condition) in order for the resulting the linear system to be of full rank.

[12] We have written the price vector $\widehat{\mathbf{p}}_t$ as a linear function of the state vector to easily exposit our methodology in a way that directly extends to more general models. However, in this Krusell and Smith model this is not actually necessary because we can simply substitute the expression for prices directly into the households' budget constraint, and in practice that is what we do.

However, for larger models with more state variables (for example the two-asset model of Section 3), the required matrix decomposition becomes prohibitively expensive; for large enough systems, the memory requirements render the problem impossible to solve. In these cases we must first reduce the size of the linear system before solving it.

We approach the problem of dimensionality reduction in three separate steps: (1) solving the static conditions, (2) reducing the size of the distribution $\widehat{\mathbf{g}}_t$, and (3) reducing the size of the value function $\widehat{\mathbf{v}}_t$.

Solve out Static Conditions When prices are defined as explicit functions of the state variables, the corresponding static conditions can be solved out prior to reducing the dimensionality of the problem. This can be done simply by solving the third line of (LINEAR) system to get

$$\widehat{\mathbf{p}}_t = -\mathbf{B}_{pg}\widehat{\mathbf{g}}_t - \mathbf{B}_{pZ}Z_t.$$

Substituting $\widehat{\mathbf{p}}_t$ back into (LINEAR), we get

$$\begin{aligned}\mathbb{E}_t\begin{bmatrix} d\widehat{\mathbf{v}}_t \\ d\widehat{\mathbf{g}}_t \\ dZ_t \end{bmatrix} &= \begin{bmatrix} \mathbf{B}_{vv} & \mathbf{B}_{vg} - \mathbf{B}_{vp}\mathbf{B}_{pg} & \mathbf{B}_{vZ} - \mathbf{B}_{vp}\mathbf{B}_{pZ} \\ \mathbf{B}_{gv} & \mathbf{B}_{gg} - \mathbf{B}_{gp}\mathbf{B}_{pg} & \mathbf{B}_{gZ} - \mathbf{B}_{gp}\mathbf{B}_{pZ} \\ \mathbf{0} & \mathbf{0} & -\nu \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{v}}_t \\ \widehat{\mathbf{g}}_t \\ Z_t \end{bmatrix} dt \\ &:= \begin{bmatrix} \mathbf{C}_{vv} & \mathbf{C}_{vg} & \mathbf{C}_{vZ} \\ \mathbf{C}_{gv} & \mathbf{C}_{gg} & \mathbf{C}_{gZ} \\ \mathbf{0} & \mathbf{0} & -\nu \end{bmatrix}\begin{bmatrix} \widehat{\mathbf{v}}_t \\ \widehat{\mathbf{g}}_t \\ Z_t \end{bmatrix} dt\end{aligned}$$

With N wealth grid points (and after dropping one element of $\widehat{\mathbf{g}}_t$ to satisfy the normalization condition (AUX 4), these are now exactly $4N$ equations in $4N$ unknowns.

Reducing the Distribution The distribution enters the system for two separate reasons: first, it is needed to compute current decisions (i.e., it enters the HJB equation through the matrix $\mathbf{C}_{vg}$) and second, it is needed to forecast future values of the distribution (i.e., it enters the KFE through the matrix $\mathbf{C}_{gg}$). Economic intuition suggests that including the entire distribution for these two purposes is an overkill. In fact, for this model, Krusell and Smith (1998) find that the aggregate capital stock contains almost all the relevant features of the distribution. However, directly generalizing Krusell and Smith's insight to other models is potentially problematic because it requires the user to guess which features of the distribution are important and a method for checking *ex-post* that the guess is correct. Instead, we draw on tools from systems and control theory in Reiter (2010) to

reduce the distribution in an automatic, model-free way.

The first step is to compute which features of the distribution are necessary to compute current decisions, that is we find the directions of $\widehat{\mathbf{g}}_t$ that have the biggest impact on the value function as captured by $\mathbf{C}_{vg}\widehat{\mathbf{g}}_t$. To do that, we compute a singular value decomposition (SVD) of the matrix $\mathbf{C}_{vg} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$, where $\mathbf{\Sigma}$ is a $2N \times 2N$ diagonal matrix with the singular values $\sigma_1, ..., \sigma_{2N}$ in descending order and $\mathbf{U}$ and $\mathbf{V}$ are $2N \times 2N$ unitary matrices which contain the left and right singular vectors $\mathbf{u}_{\sigma_j}$ and $\mathbf{v}_{\sigma_j}, j = 1, ..., 2N$ (both of which form an orthonormal basis for $\mathbb{R}^{2N}$, i.e. the space in which $\widehat{\mathbf{g}}_t$ lives). If a singular value, σ_j, is small, then the norm $\|\mathbf{C}_{vg}\mathbf{v}_{\sigma_j}\| = \sigma_j$ is small. Therefore, variations in $\widehat{\mathbf{g}}_t$ in the direction of $\mathbf{v}_{\sigma_j}$ have a small impact on the value function. For singular values that are smaller than machine accuracy, there is no improvement in accuracy from including the corresponding singular vectors. In practice, singular values decay fast, and only a small fraction of singular vectors are required to achieve machine-level accuracy. Denoting the number of the important singular vectors by k, we eliminate the unimportant ones and only keep the remaining k-dimensional subspace of $\mathbf{V}$.

Although $\mathbf{V}'\widehat{\mathbf{g}}_t$ is sufficient for computing current decisions, it may not be sufficient for computing the dynamics of $\widehat{\mathbf{g}}_t$ since $\mathbf{V}'\widehat{\mathbf{g}}_t$ may itself depend on other distributional features. Hence the second step is to compute which features of the distribution are necessary to forecast the reduced state vector $\mathbf{V}'\widehat{\mathbf{g}}_t$. From left-multiplying the linearized KFE, i.e. the second equation of (LINEAR), by $\mathbf{V}'$, the dynamics of $\mathbf{V}'\widehat{\mathbf{g}}_t$ is given by

$$d\mathbf{V}'\widehat{\mathbf{g}}_t = \mathbf{V}'\mathbf{C}_{gv}\widehat{\mathbf{v}}_t dt + \mathbf{V}'\mathbf{C}_{gg}\widehat{\mathbf{g}}_t dt + \mathbf{V}'\mathbf{C}_{gZ}Z_t dt,$$

which in principle depends on the entire distribution through the term $\mathbf{V}'\mathbf{C}_{gg}\widehat{\mathbf{g}}_t$ (as opposed to depending only $\mathbf{V}'\widehat{\mathbf{g}}_t$). However, we can compute another singular value decomposition of the matrix $\mathbf{V}'\mathbf{C}_{gg}$ to find which subspace of $\widehat{\mathbf{g}}_t$ matters for the dynamics of $\mathbf{V}'\widehat{\mathbf{g}}_t$ up to machine accuracy. If this decomposition returns the same singular vectors as $\mathbf{V}'$, then the reduced state vector forecasts itself, and we are done. However, if the decomposition returns new singular vectors, then we must iterate on this procedure to compute the features of the distribution necessary to forecast the new singular vectors. In practice, this iteration converges quickly. Denoting by k_g the number of important singular vectors that are needed to forecast, let the $(2N-1) \times k_g$ matrix $\mathbf{V}_g$ be the orthogonal basis corresponding to the subspace of the dimensionality reduction (note that, in general, $\mathbf{V}_g$ differs from $\mathbf{V}$ above).[13]

[13]We have described the state reduction procedure as characterizing the features of the distribution necessary to compute decisions, but we can easily augment this procedure to characterize other features of the distribution that a researcher is interested in. This is particularly useful for moment-based estimation.

	Grid Size N = 200		Grid Size N = 600	
	No dim reduction	*Dim reduction*	*No dim reduction*	*Dim reduction*
Steady State	0.141 sec	0.113 sec	0.264 sec	0.257 sec
Derivatives	0.073 sec	0.070 sec	0.228 sec	0.219 sec
Dim reduction	×	0.539 sec	×	6.880 sec
Linear system	1.67 sec	0.189 sec	42.1 sec	3.21 sec
Total	**1.884 sec**	**0.911 sec**	**42.592 sec**	**10.566 sec**

Table 1: Seconds to solve Krusell and Smith model

Notes: Time to solve model in one run in Matlab R2014b on a MacBook Pro laptop computer. "Grid size" refers to number of points in asset space; total size of the system is 4 times grid size.

It is then sufficient to keep track of the k_g dimensional vector $\widetilde{\mathbf{g}}_t := \mathbf{V}_g' \widehat{\mathbf{g}}_t$ with law of motion

$$d\widetilde{\mathbf{g}}_t = \mathbf{V}_g' \mathbf{C}_{gv} \widehat{\mathbf{v}}_t dt + \mathbf{V}_g' \mathbf{C}_{gg} \mathbf{V}_g \widetilde{\mathbf{g}}_t dt + \mathbf{V}_g' \mathbf{C}_{gZ} Z_t dt,$$

which importantly depends only on the reduced distribution $\widetilde{\mathbf{g}}_t$ itself rather than the full distribution $\widehat{\mathbf{g}}_t$. After reducing the dimensionality of $\widehat{\mathbf{g}}_t$ in this way, we are left with a system of dimension $2N + k_g$ which is in practice considerably smaller than $4N$. That being said, the dimensionality of the linear system is still large because it contains $2N$ equations for the value function at each point on the grid for the idiosyncratic state variables. Depending on the model, even after reducing the dimensionality of the state space, the dimensionality of the linear system may be too large for matrix decomposition methods to be feasible. One way to get around this issue is to use sparse matrix methods to find only the smaller number of eigenvectors associated with the stable eigenvalues corresponding to the reduced state space. This is much faster than computing the full matrix decomposition required to obtain the full set of eigenvectors, whenever the number of stable eigenvalues are much smaller than the number of unstable eigenvalues.[14]

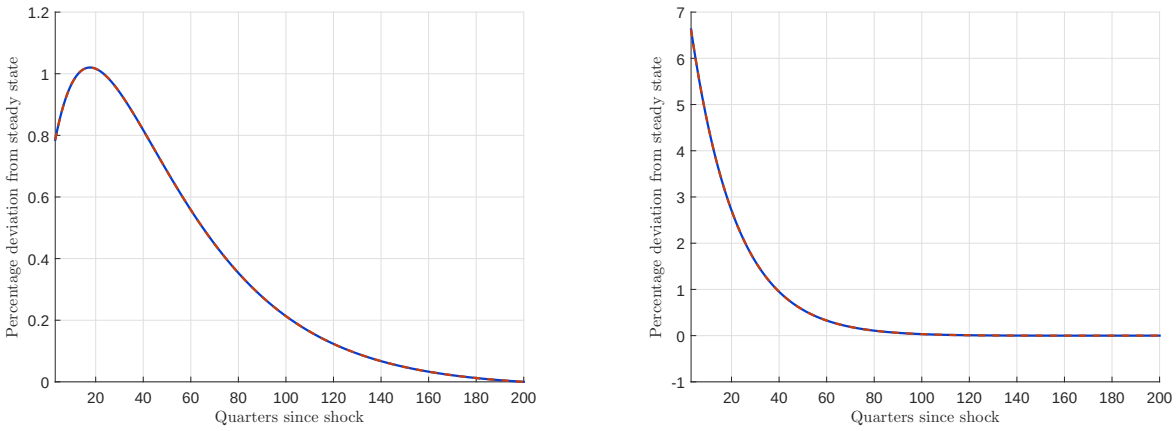
Table 1 shows that the distribution reduction step can significantly speed up the algorithm, particularly when the dimensionality of the system is large.[15] The table computes the

[14] A program using this method to solve the linear problem will be provided for general use

[15] We use the following illustrative parameterization to solve the model. We set the coefficient of relative risk aversion $\gamma = 2$ and the discount rate $\rho = .01$ quarterly. We set capital depreciation $\delta = .025$ quarterly and the capital share $\alpha = \frac{1}{3}$. We choose the levels and transition probabilities of the idiosyncratic shocks following Den Haan, Judd and Julliard (2010). Finally, we set the aggregate productivity shock process to yield a quarterly persistence of .95 and innovation volatility of .007.

Figure 1: Impulse Response to Positive Productivity Shock, Full vs. Reduced Model

(a) Consumption (b) Investment



Notes: Impulse responses to one quarter of positive aggregate productivity shocks. Solid blue line is without dimensionality reduction, using 400 points for distribution. Dashed red line with with dimensionality reduction, using 95 points for distribution.

total time to solve the model, and decomposes it into the various components: computing the steady state, computing the first-order Taylor expansion, reducing the distribution, and solving the linear system. For an asset grid size of $N = 200$ (and therefore total system size of $4N = 800$), our method without reducing the distribution solves the less than model in two seconds; our method with dimensionality reduction solves the model in less than one second. For a larger asset grid size of $N = 600$ (and a total system size of $4N = 2400$), reducing the distribution becomes more advantageous; without it, our method solves the model in 42 seconds, but with it, our method solves the model in 10 seconds.

It is important to note, however, that the dimensionality reduction step is not only about speed. For larger models, such as in Section 3, without some form of dimensionality reduction, the linear system is simply too large to be solved.

Figure 1 shows that the aggregate dynamics of the model are unaffected by the state reduction step. The figure plots the impulse responses of aggregate consumption and investment to one quarter of aggregate productivity shocks for both the full model (with $N = 200$) and the reduced model.[16] The responses under the two cases are indistinguishable.

[16]That is, we feed in positive one standard deviation innovations to the aggregate productivity process for one quarter.

Reducing the Value Function As just discussed, even after reducing the dimensionality of the distribution, the dimensionality of the linear system is still large because it contains $2N$ equations for the value function at each point on the grid for the idiosyncratic state variables. Economic intuition, however, suggests that for the purposes of linearizing with respect to the aggregate states, using a fine grid to approximate the value functions is overkill. In most models, the value function is smooth, and over much of the state space is almost linear. Capturing this behavior does not require the value function at each of the points along a finely spaced grid.

Any good spline that takes advantage of smoothness can be used to reduce the dimensionality of the problem. We have found that a non-uniformly spaced quadratic spline works well in most commonly used economic models, for the following reasons: (i) splines exploit smoothness; (ii) non-uniform spacing reduces required grid points by taking advantage of systemic differences in curvatures of value functions; (iii) the local nature of quadratic spline does not create spurious oscillations from boundary conditions as often occurs with approximations based on global functions, such as Chebychev polynomials; (iv) quadratic splines preserve monotonicity between knot points unlike cubic or higher order polynomial splines; and (v) the resulting relation between the original approximation on the fine grid points and the spline knot points that is linear, yielding a simple subspace defined by a $2N \times k_v$ matrix $\mathbf{W}_v$, where k_v is the number of knot points.[17]

Hence, the (LINEAR) system is reduced to

$$\mathbb{E}_t \begin{bmatrix} d\widetilde{\mathbf{v}}_t \\ d\widetilde{\mathbf{g}}_t \\ dZ_t \end{bmatrix} = \begin{bmatrix} \mathbf{W}_v' \mathbf{C}_{vv} \mathbf{W}_v & \mathbf{W}_v' \mathbf{C}_{vg} \mathbf{V}_g & \mathbf{W}_v' \mathbf{C}_{vZ} \\ \mathbf{V}_g' \mathbf{C}_{gv} \mathbf{W}_v & \mathbf{V}_g' \mathbf{C}_{gg} \mathbf{V}_g & \mathbf{V}_g' \mathbf{C}_{gZ} \\ \mathbf{0} & \mathbf{0} & -\nu \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{v}}_t \\ \widetilde{\mathbf{g}}_t \\ Z_t \end{bmatrix} dt,$$

where $\widetilde{\mathbf{v}}_t$ and $\widetilde{\mathbf{g}}_t$ are the reduced vectors that approximate the full control and state variables

$$\mathbf{v}_t \approx \mathbf{W}_v \widetilde{\mathbf{v}}_t$$
$$\mathbf{g}_t \approx \mathbf{V}_g \widetilde{\mathbf{g}}_t$$

Depending on the reducibility of the problem, the problem can be solved by standard solution methods in fractions of a second. For the example case with $N = 600$ wealth grid points,

[17] It is crucial to note that we are not suggesting to exploit these properties and use a low-dimensional approximation when solving for the steady-state value functions. A sufficiently rich approximation to the steady-state value functions is necessary to capture the non-linear behavior at the individual level. We are instead proposing that once the steady-state solution has been found, it can be accurately represented in a lower-dimensional subspace for the purposes of linearization with respect to the aggregate states.

the model can be reduced to $k_{\text{state}} = 6$ and k_{value}=24 with the resulting system size of 30 compared to that of 2400.

2.4 Micro Heterogeneity and Macro Nonlinearities

An important motivation for including heterogeneity in macroeconomic models is to study how micro heterogeneity can potentially generate nonlinear dynamics in aggregate variables. For example, researchers are often interested in how the response of the economy to an aggregate shock depends on the initial distribution of agents, the size of the shock, or the sign of the shock because of the nonlinearity in decision rules at the individual level. Even though our methodology relies on linear approximations, it preserves much of these aggregate nonlinearities in simulations of the model.

Our procedure preserves nonlinearities for two reasons. First, it preserves "state dependence" of aggregates, i.e. dependence on the initial distribution, because the linear approximation involves different derivatives for each point in the state space so that the aggregate derivative depends on the distribution over these points. Second, it preserves "size dependence", i.e. dependence on the size of the shock, because we only use the linear approximation to households' value function when simulating models. Given the value function, we then derive all other variables fully nonlinearly.

To fix ideas, consider the impulse response of aggregate consumption C_t to an aggregate productivity shock Z_t, starting from some initial distribution $g_0(a, z)$. To compute the effect on the shock upon impact, we first use the linear approximation of the value function $v_0(a, z) = v(a, z) + \widehat{v}_0(a, z)$ from (SOLUTION) to compute individual consumption from the nonlinear first order condition $c_0(a, z) = (\partial_a v_0(a, z))^{-\frac{1}{\gamma}}$. We then aggregate this over households to arrive at aggregate consumption

$$C_0 = \int c_0(a, z) g_0(a, z) da dz.$$

This procedure preserves state dependence with respect to the initial distribution $g_0(a, z)$ because this distribution enters the aggregation directly.[18]

[18]To see this even more clearly, note that one can write the deviation of initial consumption from steady state in the form $\widehat{c}_0(a, z) \approx D_{cZ}(a, z)Z_0$ where the coefficients $D_{cZ}(a, z)$ are found as follows:

$$\widehat{c}_0(a, z) = (\partial_a v(a, z))^{-\frac{1}{\gamma}-1} \partial_a \widehat{v}_0(a, z) = (\partial_a v(a, z))^{-\frac{1}{\gamma}-1} \partial_a D_{vZ}(a, z) Z_0 := D_{cZ}(a, z) Z_0.$$

and where $D_{vZ}(a, z)$ are the elements of $\mathbf{D}_{vZ}$ in (SOLUTION), i.e. the different derivatives for each point in the state space. Therefore the aggregate derivative depends on the distribution over these points, i.e. state

To then compute the dynamics of the shock after impact, we update the distribution using the nonlinear form of the KFE, i.e. the second equation of (NON-LINEAR) rather than the linear counterpart in (LINEAR). We first compute prices from the nonlinear pricing equation $\mathbf{p}_0 = F(\mathbf{g}_0, Z_0)$ and use the policy function from the household optimzation problem to get the transition matrix $\mathbf{A}(\mathbf{v}_0;\mathbf{p}_0)$. We then update the distribution according to

$$\frac{d\mathbf{g}_0}{dt} = \mathbf{A}(\mathbf{v}_0;\mathbf{p}_0)'\,\mathbf{g}_0$$

This procedure allows us to preserve "size dependence" because larger shocks induce potentially non-proportional movements around the individual state space in a given interval of time, and therefore different distributional dynamics going forward. The only time at which this is not true is the impact period $t = 0$, since at $t = 0$ the value function is itself linear in the aggregate shock (see footnote 18). However, that period is arbitrarily small in continuous time. After $t = 0$, we simply repeat this procedure throughout the course of the simulation.[19]

Figure 2 shows that our solution method preserves state dependent impulse responses in the Krusell and Smith model. We compute the impulse response to a positive one standard deviation productivity shock starting from two different initial distributions: the steady state distribution and the distribution starting from one quarter of negative shocks (a recession). Qualitatively, aggregate consumption responds by more starting from the recession, because households are on average closer to their borrowing constraints and therefore have higher MPCs. Quantitatively, however, this effect is small. This is due to the well-known "approximate aggregation" property of the Krusell and Smith model that renders the distribution quantitatively unimportant. The figure shows though that even in this environment our procedure captures what small nonlinear dynamics do exist.

dependence:

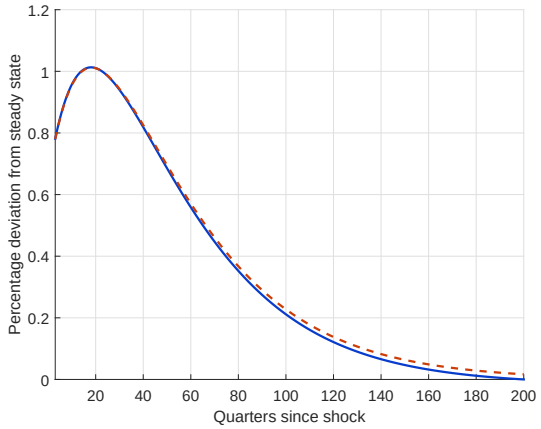
$$\widehat{C}_0 \approx \int D_{cZ}(a,z)g_0(a,z)dadz \times Z_0.$$

At the same time, the initial impulse response $\widehat{C}_0$ does *not* feature size dependence with respect to the initial impulse Z_0 which enters linearly.

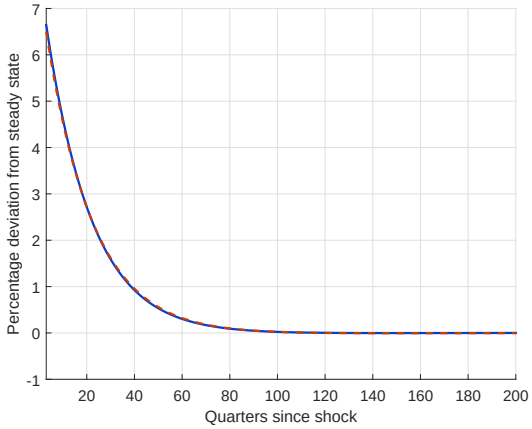
[19] Again, a special feature of the Krusell and Smith model is that prices are directly defined by the distribution, but in more general models prices may only be defined implicitly through market clearing conditions. In those cases we use the linearized pricing equation from the linearized solution of the model. Although this imposes that prices are a linear function of the distribution, it will still allow for nonlinear dynamics of the prices, since the distribution evolves fully nonlinearly.

Figure 2: Impulse Response to Positive Productivity Shock, Starting From Steady State vs. Recession

(a) Consumption



(b) Investment



Notes: Impulse responses to one quarter of positive aggregate productivity shocks. Solid blue line is starting from steady state. Dashed red line is starting after one quarter of negative productivity shocks.

2.5 `Matlab` Code Package

Although our methodology is straightforward on a conceptual level – extending linearization to heterogeneous agent models – we had to deal with many practical challenges inherent in working with such high-dimensional systems. Fortunately, our solutions to these challenges are general purpose and can therefore be automated for future use. In order to encourage other researchers to use our methods, we have developed a `Matlab` code package which implements each step in a model-free way. The user simply provides two inputs: (i) a function that defines their model's equilibrium conditions; and (ii) the value of the model's variables in steady state. Our package then (1) differentiates the equilibrium conditions, using our automatic differentiation software that exploits sparsity; (2) reduces the distribution using the singular value decomposition approach; (3) reduces the value function using quadratic splines; (4) solves the linear system, using a solver that exploits sparsity; and (5) simulates the model in a way that preserves aggregate nonlinearities due to micro heterogeneity.

Using this code package, if a researcher can compute the steady state of their model, then they compute the aggregate dynamics at a very low cost. In order to provide guidance on computing the steady state of continuous time heterogeneous agent models, we have additionally compiled codes to solve many different examples (e.g., continuous time versions of Huggett (1993), Aiyagari (1994), or Khan and Thomas (2008)). The interested reader can consult Achdou et al. (2015) for details on the theory behind the solution method for

the steady-state, and more computed examples.

3 Application to Aggregate Consumption Dynamics

In the second part of the paper we will illustrate the usefulness of these methods by applying them to a richer environment that previously was computationally infeasible. The substantive issue that we will address is the dynamics of aggregate consumption.

3.1 Background

Twenty-seven years ago, in the NBER Macroeconomics annual, Campbell and Mankiw (1989) showed that some key regularities in the aggregate time series data on consumption, income, and interest rates are inconsistent with the predictions of the representative agent Euler equation. In particular, they showed that changes in income are associated with changes in consumption, while real interest rates are not associated with changes in consumption. Campbell and Mankiw (1989) suggested a simple heterogeneous agent model (the 'spender-saver' model) in which an exogenous fraction of households are hand-to-mouth and simply consume their disposable income in every period, while the remaining households are permanent income consumers and consume according to a representative agent Euler equation. They estimate that the fraction of hand-to-mouth households should be around 50% in order to reproduce the aforementioned empirical regularities.

Due to its simplicity and ability to capture these salient features of the data, the spender-saver extension to the representative agent model has since been incorporated into most medium-scale DSGE models currently used in practice. However, there are at least two substantive issues with the spender-saver model of aggregate consumption. First, in his discussion of Campbell and Mankiw, Christiano (1989) points out that the spender-saver model generates an unconditional volatility of aggregate consumption that is too large relative to the unconditional volatility of income. The reason is that since 50% of households have a marginal propensity to consume that is always equal to one, consumption moves one-for-one with income for half of the population. This is in stark contrast to the well known empirical regularity that aggregate consumption is much smoother than aggregate income.

Second, since the publication of Campbell and Mankiw (1989), numerous empirical studies have shown that the assumption that 50% of households do not hold any wealth and consume all income changes, regardless of their size, is at odds with micro data on household portfolios and the consumption response to income shocks. This evidence suggests that

at most 10% to 15% of US households can be argued to have zero or low enough levels of net assets such that they can be classified as hand-to-mouth households - substantially below the 50% estimated by Campbell and Mankiw. On the other hand, the same empirical evidence suggests that marginal propensities to consumes are around 20% to 40% on average and have a bimodal distribution, sharply contradicting the representative agent model (Johnson, Parker and Souleles, 2006; Parker et al., 2013; Misra and Surico, 2014). In sum, the micro-evidence is broadly consistent with a more nuanced version of the spender-saver model in which at a given point in time some households have high marginal propensities to consume (albeit below unity), while other households satisfy Euler equations.

Kaplan and Violante (2014) and Kaplan, Violante and Weidner (2014) proposed a partial equilibrium model of household consumption behavior that retains the spirit of the Campbell and Mankiw (1989) insights, but is consistent with the aforementioned empirical microeconomic evidence. We propose to embed this model of household consumption in a general equilibrium framework and re-assess the Campbell and Mankiw (1989) and Christiano (1989) findings. In our model, households can save in two assets: an illiquid asset that pays a high return but is costly to adjust, and a liquid asset that pays a lower return but can be freely adjusted subject to borrowing frictions. The model endogenously generates a class of "wealthy hand-to-mouth" consumers who hold a substantial amount of illiquid assets but nevertheless have high marginal propensities to consume because they hold a small amount of liquid assets. Kaplan and Violante (2014) show that a model with these features can be made consistent with the distribution of these propensities measured in the micro data.

This framework holds great promise for being simultaneously consistent with the observations of Campbell and Mankiw (1989) and the Christiano (1989) findings, since despite generating large average marginal propensities to consume, households are nonetheless forward looking rational agents who have a long-run smoothing motive. This distinction between their short-run and long-run consumption behavior separates this class of models from the earlier spender-saver models. Moreover, Kaplan, Moll and Violante (2016) have shown that this two-asset setting generates much smaller direct responses of aggregate consumption to interest rate changes than in representative agent models. Thus the framework also has the scope to be consistent with Campbell and Mankiw's observation that expected consumption growth is not associated with expected real interest rates.

A quantitative assessment of these claims requires one to embed the two-asset precautionary savings model in a general equilibrium environment with aggregate shocks. Until now this has been computationally infeasible since the two-asset structure implies that the

rich distribution of MPCs at the micro level matters for aggregate dynamics, violating approximate aggregation. However, the methodology developed in Section 2 can accurately and efficiently solve this model.

3.2 Model Environment and Quantitative Analysis.

The following is a verbal overview of the model that we will solve. The household side of the model is a simplified version of the model in Kaplan, Moll and Violante (2016).[20] Households face idiosyncratic income shocks as well as stochastic mortality shocks. Idiosyncratic income shocks are parameterized to be consistent with the leptokurtic distribution of income changes and skewed labor income distribution evident in Social Security Administration data. Since the income process is calibrated to pre-tax income, we include a simple tax and transfer system. Labor supply is exogenous. Households can save in both a liquid asset and an illiquid asset. Moving funds between the two assets is subject to a transaction cost. The parameters of the transaction cost function are chosen to match salient features of the liquid and illiquid wealth distributions across households.

There is a representative firm with a Cobb-Douglas production function subject to stochastic total factor productivity shocks. Prices are flexible and we study the perfectly competitive equilibrium of the model. As in Aiyagari (1994) the equilibrium capital stock used in production is determined by households' savings in illiquid assets. Thus the marginal product of capital determines the equilibrium return on the illiquid asset. There is an exogenous supply curve of liquid assets with a constant elasticity with respect to the liquid return (which may be zero, i.e., liquid assets may be in fixed supply). We also allow for exogenous shocks to this supply of liquid assets, which is a simple way to generate movements in the liquid return relative to the illiquid return.[21]

We have already computed the steady state of this model and have a working calibration that matches economic aggregates as well as the micro data on household portfolios and marginal propensities to consume. The next stage involves using our `Matlab` code package to solve for the aggregate dynamics of the model, and to study the model's predictions for aggregate consumption, income, and real interest rates. In particular, we will compute both the Campbell and Mankiw (1989) and Christiano (1989) statistics using simulated data from

[20] The main simplifications relative to Kaplan, Moll and Violante (2016) are to (i) make labor supply exogenous, (ii) remove housing as an asset class, (iii) drop the intermediate goods sector with monopolistic competition and (iv) allow flexible prices. The production side of the model is then as close as possible to a standard RBC environment.

[21] A special case of this formulation is to make the supply of liquid assets inelastic.

the model, and evaluate the success of the model in in these dimensions. These results will allow us to assess the extent to which our model can be consistent with both micro data and macro consumption dynamics simultaneously.

We will also use the model to illustrate the possibilities for heterogeneous agent modeling that our computational methodology allows. First, we will compute the distributional implications of aggregate shocks, which the corresponding representative agent economy is silent on. Second, we will show how micro heterogeneity yields nonlinear aggregate dynamics, as discussed in Section 2.4.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2015): "Heterogeneous Agent Models in Continuous Time," Discussion paper, Princeton University.

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

Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden (2015): "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," Discussion paper, University of Bonn.

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–11.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. Terry (2014): "Really Uncertain Business Cycles," Discussion paper.

Campbell, J. (1998): "Entry, Exit, Embodied Technology, and Business Cycles," *Review of Economic Dynamics*, 1(2), 371–408.

Campbell, J. Y., and N. G. Mankiw (1989): "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pp. 185–216. National Bureau of Economic Research.

Christiano, L. (1989): "Comment on "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence"," in *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pp. 216–233. National Bureau of Economic Research.

Den Haan, W., K. Judd, and M. Julliard (2010): "Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty," *Journal of Economic Dynamics and Control*, 34(1), 1–3.

Guerrieri, V., and G. Lorenzoni (2011): "Credit Crises, Precautionary Savings, and the Liquidity Trap," NBER Working Papers 17583, National Bureau of Economic Research.

Huggett, M. (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17(5-6), 953–969.

Johnson, D. S., J. A. Parker, and N. S. Souleles (2006): "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 96(5), 1589–1610.

Kaplan, G., B. Moll, and G. L. Violante (2016): "Monetary Policy According to HANK," Working Papers 1602, Council on Economic Policies.

Kaplan, G., and G. L. Violante (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82(4), 1199–1239.

Kaplan, G., G. L. Violante, and J. Weidner (2014): "The Wealthy Hand-to-Mouth," *Brookings Papers on Economic Activity*, 48(1 (Spring), 77–153.

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

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

McKay, A., E. Nakamura, and J. Steinsson (2015): "The Power of Forward Guidance Revisited," NBER Working Papers 20882, National Bureau of Economic Research.

Misra, K., and P. Surico (2014): "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs," *American Economic Journal: Macroeconomics*, 6(4), 84–106.

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103(6), 2530–53.

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

——— (2010): "Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models," Economics Series 258, Institute for Advanced Studies.

Winberry, T. (2016): "Lumpy Investmet, Business Cycles, and Stimulus Policy," Discussion paper.