On the Solution of High-Dimensional Macro Models with Distributional Channels *

Luca Mazzone†

September 2021

Abstract

I propose a global solution method for the computation of infinite-horizon, heterogeneous agent macroeconomic models with aggregate uncertainty. Details of the algorithm are illustrated by presenting its application to a an example model of firm dynamics. In the model, aggregate dynamics depends explicitly on firm entry and exit, and individual choices are often constrained by a form of market incompleteness. Existing computational strategies are either unfeasible or provide inaccurate solutions to this class of models. Moreover, global solutions are computationally expensive because the minimal representation of the aggregate state space - and thus the aggregate law of motion - faces the curse of dimensionality. The proposed strategy combines adaptive sparse grids with a cross-sectional density approximation, and introduces a framework for solving the more general class of dynamic models with firm or household heterogeneity accurately. JEL codes: C630, E320

† : *European Department, International Monetary Fund*

*I am grateful to Felix Kübler, and Simon Scheidegger for continuous support and guidance. I also thank Marlon Azinovic, Chris Carroll, Phillip Renner, Jean-Charles Rochet, Paymon Khorrami, Karl Schmedders, Daniel Harenberg, Ken Judd, and participants at seminars at UZH, at the SFI Research Days and the CEF conference for helpful comments.

1 Introduction

In this paper I introduce a global solution method for solving infinite-horizon, discrete time, heterogeneous agent models with aggregate uncertainty. I illustrate how it works by presenting a model economy which, in its relative simplicity, contains numerous computational challenges that make most existing methods inapplicable or unfeasible.

Despite considerable amount of effort aimed at incorporating agent heterogeneity and distributional effects into macroeconomics, their inclusion in the standard policy toolbox is far from widespread. A relevant obstacle is the computation of equilibria in models where heterogeneity plays a crucial role. Such economies typically contain two sources of risk: agents are subject to idiosyncratic shocks as well as some form of aggregate uncertainty, and risk can be non-diversifiable. It is known that economies where all idiosyncratic uncertainty cancels out in the aggregate are characterized by a steady state cross-sectional distribution which can be computed fairly easily. With aggregate uncertainty, however, a steady state wealth distribution will not exist in general. Also, as the effects do not cancel out in the aggregate, so the cross-sectional distribution changes with the stochastic aggregate shock. As a consequence, the set of state variables has to include the time-varying cross-sectional distribution of agents' characteristics. In models with a continuum of agents, the object that characterizes the cross-sectional distribution becomes infinite-dimensional, and thus intractable.

The development of computational methods in economics of the 1990s (see Judd (1998) and Ríos-Rull (1997)) has led to an increased interest in the computation of equilibrium. Contemporaneous increased availability of computational tools and resources allowed solution procedures to exploit parallelization - see Dongarra and van der Steen (2012). A seminal contribution on the solution of distributional models with aggregate uncertainty came by Krusell and Smith (1998), who proposed a simple and versatile solution method. Their model is based on Aiyagari (1994): ex-ante identical agents face non-insurable unemployment risk and have to decide over consumption and savings. The presence of non-insurable risk implies individual histories of shocks of all agents are relevant for determining current choices: their distribution affects prices, hence the cross-section of choices enters the individual decision problem.

In practice, however, in the model economy of Aiyagari (1994) only a rather small fraction of agents is constrained, and this small fraction of agents hold an even smaller fraction of aggregate wealth. Therefore, the aggregate behavior of the economy depends almost entirely on the actions of non-constrained agents, and it can be summarized by the first moment of the distribution - a property known as approximate aggregation[1].

[1] A note of caution: approximate aggregation does not imply that the economy can be equivalently described by a representative agent model, nor that there will be zero dispersion among agents.

This feature drives the main result of their paper, that is the possibility to use the mean of individual choices as a sufficient statistic to predict future aggregates. When approximate aggregation holds, the aggregate law of motion influences the distribution of individual choices, but the solution of each individual problem can rely on aggregates as a set of sufficient statistics for the whole distribution. As Krusell and Smith themselves argue in their article, a more involved interplay between the cross-sectional distribution of agents' choices and individual policies requires different solution methods.
Algan et al. (2008) propose an algorithm that is aimed specifically at solving models where distributional channels have an important role in individual choices: their global solution method uses a smooth cross-sectional density approximation to compute agents' expectation of future aggregates. However, this approach has seen limited use, mostly because it requires a full discretization of the aggregate state space that becomes rapidly unfeasible as the dimensionality of the state space increases - a classic issue known as the curse of dimensionality[2].

To address these issues with the help of a practical example, I introduce a model economy where aggregate dynamics depends explicitly on firm entry and exit, and individual choices are often constrained by a form of market incompleteness. A simpler version of the model has been proposed by Arellano et al. (2016) [3] in order to study the effects of uncertainty shocks on the business cycle. The baseline model includes heterogeneous firms whose choices are characterized by a timing structure that forces them to choose labor with a period delay before its use in production. Timing structure makes uncertainty on individual productivity relevant for profit maximization[4]. Although market incompleteness does distort firm choices, entry and exit play no explicit role in the model, as firms default exogenously and the measure of active firms is restricted to be constant over time. In my own extension, idiosyncratic risk generates revenue risk that cannot be hedged and might force the firm to default depending on their past choices, so that exit as well as entry are endogenous, and they are not restricted to cancel each other. Time-varying net entry means the measure of active firms is itself time-varying. Default risk, in turn, implies that firms will become more cautious when uncertainty increases, generating a potential aggregate first-order effect from shocks to the second moment of the distribution of productivity.

In this framework, a departure from approximate aggregation is more likely, as every agent is affected by default risk with the intensity of this distortion depending on its past choices, so that higher order moments of the distribution of individual choices will influence individual policies. In presence of such micro-level dynamics, the strategy

[2] see Bellman (1961)

[3] Later circulated as Arellano et al. (2019). I explicitly refer to the older version of the model in the paper, because it presents unique computational challenges.

[4] A similar timing structure can be traced back to the capital investment technology in Kydland and Prescott (1982)

in Krusell and Smith (1998) is expected to fail in approximating the aggregate law of motion accurately, and global solution methods can be required; however, given the multidimensional nature of the aggregate state space, those methods would be computationally burdensome. The computational challenge is twofold: the aggregate law of motion has to be determined at many points in the state space, and each point involves an expensive computation.

The problem of approximating an aggregate law of motion can be seen from the perspective of the interpolation of a high-dimensional function: if the function is sufficiently smooth, sparse grids perform this task with small losses in terms of accuracy (see Bungartz and Griebel (2004)); if this is not the case, sparse grids are of little help. Brumm and Scheidegger (2017) is the first application of using sparse grids with another layer of sparsity to economic problems: their solution algorithm "learns" which areas of state space need a finer grid by relying on so-called adaptive sparse grids (see Pflüger (2012)). I follow their approach and show how the use of adaptive sparse grids makes it feasible to compute global solutions of heterogeneous agents models even with a reasonably high-dimensional state space.

The present paper combines adaptive sparse grids with a global solution method for economies populated by a continuum of agents, and proposes a general framework for solving dynamic models with firm or household heterogeneity. This approach manages to obtain accurate solutions in settings where common methods have limited application: non-linear or even discontinuous individual policies, big aggregate shocks, multidimensional state spaces, and relevant distributional channels. It also alleviates the curse-of-dimensionality problem discussed above. However, even if adaptive sparse grids reduce significantly the number of points to be evaluated, obtaining a solution in a reasonable time can still be an impossible task. I discuss how my approach can be implemented on a HPC cluster, and describe a suitable parallelization scheme that allows a fast computation of the solution. As the main computation tasks in my algorithm are fully independent of each other, they can be solved in parallel by distributing them via Message Passing Interface among different computer nodes (see Dalcín et al. (2005) and Dalcín et al. (2008)).

I compare the relative performance of my algorithm to an adapted version of the Krusell and Smith (1998) method, after illustrating different suitable measures of accuracy for heterogeneous agent models with aggregate uncertainty, and their application to the example economy of my paper. The proposed method is successful in solving for the model equilibrium, using various accuracy measures based on training sets over the whole state space, or on simulation outcomes. The latter case is particularly relevant, because it allows a fair comparison with the accuracy results of the competing approaches; in particular, Krusell and Smith (1998) fails in obtaining equally accurate approximations of the aggregate law of motion. Also, by analyzing the equilibrium solutions of the model, interesting dynamics that could not be observed without a global

solution method emerge. Specifically, higher uncertainty induces a misallocation of labor caused by a real option effect on firms that are constrained by default risk; at the same time, net entry can increase because of a combination of general equilibrium effects and increased dispersion of productivity which make high productivity projects even more promising. Uncertainty thus creates two "option" effects that go in different direction, and so the sign of business cycles responses to a second-moment shock can vary depending not only on being paired with other shocks, but also on their duration.

The method proposed in this paper opens the door for analyzing models where the economy is hit by shocks that combine first moment and second moment shifts, and in general for giving a more precise definition of business uncertainty, policy uncertainty, and so forth. Future work on solution methods can instead improve the speed performance of the algorithm discussed here.

2 Related Literature

The present work is mostly related to two strands of literature: models that aims at exploring connections between macroeconomic aggregates and idiosyncratic risk, and papers that aim at providing useful solution methods for models of the mentioned type.

In particular, this paper relates to the literature that analyzes the role of firm heterogeneity in the business cycle, as in Khan and Thomas (2008) and Jermann and Quadrini (2012). A more general survey of the inclusion of heterogeneity in macroeconomics is provided by Guvenen (2011) and Krueger et al. (2016). Firm heterogeneity is a natural environment for analyzing the effects of aggregate shocks, which in turn points to recent efforts in understanding the effects of changes in uncertainty. In a seminal paper, Bloom (2009) reports that uncertainty shocks produce a sharp but temporary decline in aggregate output and employment. This occurs because higher uncertainty causes firms to temporarily pause their investment and hiring . Once macroeconomic relevance of policy and business uncertainty is recognized, a natural step forward is to model explicitly the impact of shocks to higher moments of the processes driving the economy, as in Bloom et al. (2013), in order to understand which channels are more important and why.

The model in this paper includes uncertainty shocks as a source of aggregate risk, and draws mostly from the economy in Arellano et al. (2016). It differs in making default depending on the idiosyncratic state of the firm, rather than on some exogenous process, and entry to depend on current business cycle conditions, rather than calibrated in order to have constant measure of firms. Besides the methodological relevance, such an explicit dynamics is interesting in light of the role of net firm entry in the business cycle. Evidence in, among others, Davis et al. (2012), points to a greater sensitivity of small businesses to aggregate conditions and a fundamental role in entry and exit in the business cycle: in

bad times, small firms tend to contract employment more and to default more than big firms. Also, Haltiwanger (2012) shows for the US that, between 1980 and 2009, 18% of gross job creation is accounted for by new firms, while 19% of gross job creation by the opening of new establishments of existing firms. In terms of gross job destruction, 17% is accounted for by firm exit, 14% by the closing of establishments of existing firms. Those numbers, especially relative to firm exit, are obviously even higher during recessions.

The main contribution of the paper lies in the literature whose aim is to provide suitable solution methods to macroeconomic models with heterogeneous agents and aggregate uncertainty. As discussed above, this features makes it difficult to approximate equilibria, because history-dependent policies in economies with a continuum of agents generate an infinite-dimensional state space in individual agents' problem. Investigating the conditions under which equilibrium exists in these economies has been the subject of an extensive theoretical literature - see Miao (2006), Cheridito and Sagredo (2016) and Brumm et al. (2017).

Turning to the problem of computation, it is possible to identify three main methodological approaches. The seminal approach proposed by Krusell and Smith (1998) is based on a combination of projection and simulation techniques[5]. Their algorithm guesses an aggregate law of motion and simulates a time series of the economy, then re-estimates the law of motion itself from the simulated series using least squares, and proceeds until the set of coefficients has converged following some criteria. I describe the application of this strategy to my model economy in detail in Appendix B. On the same vein, an algorithm based on a combination of projection and simulation is proposed by Judd et al. (2011), while Young (2010) shows that the same approach can be used more efficiently by replacing simulation with a histogram-based non-stochastic algorithm. Another strand of literature puts together projection techniques with flexible functional forms to approximate the cross-sectional distribution of agents - see Den Haan and Rendahl (2010) and Algan et al. (2008).

Finally, a third approach relies on linear perturbations around a steady state, as it happens in most macroeconomic applications without heterogeneity - the software package DYNARE has long been an important tool for economists working on dynamic models (see Adjemian et al. (2011)). In heterogeneous agent models, solution methods on this vein have been advanced in this context by Reiter (2009) and, more recently, by Winberry et al. (2016). A related intuition is used by Boppart et al. (2018), although their method relies instead on solving non-linearly for a deterministic transition path - the impulse response to an "*MIT shock*".

A comparison of the methodologies listed above has been performed by Algan et al.

[5]By "projection", here, I refer to the notion that aggregate variables enter explicitly into agents' individual problems, which are computed on different points of the (aggregate) state space: in the case of Krusell and Smith, the points of the state space are the outcome of a simulation.

(2013), and by Terry (2015) using respectively the Aiyagari economy and the Kahn and Thomas economy as benchmark models. Testing the performance of competing methodologies in terms of accuracy and speed, Terry finds the algorithm by Krusell and Smith to be superior to its competitors. The result is to be expected, given that in both models aggregate dynamics is driven at large by unconstrained agents. More recently, Pröhl (2017) shows this class of models can be solved with a solution method that obtains the fully rational equilibrium depending on the whole cross-sectional distribution, using polynomial chaos expansions. A relevant finding that follows from this approach is that idiosyncratic risk does not aggregate in equilibrium, and approximate aggregation does not hold anymore.

Finally, the method proposed in this paper relies on the use of adaptive sparse grids. Krueger and Kubler (2004) noticed that, in problems that are sufficiently well behaved, one can address the curse of dimensionality by using sparse grids. For mappings that can be represented as vector functions with bounded second order derivatives, i.e. for models that imply a sufficiently smooth law of motion, classical sparse grids define an optimal (a priori) selection of grid points. However, many economic applications have features such that these prerequisites are not met, either at the individual level or at the aggregate.

Brumm and Scheidegger (2017) show that classical sparse grids are not sufficient to obtain an accurate approximation of the state space using two examples (an international real business cycle model with irreversible investment and a firm price-setting problem with menu costs), and propose to use adaptive sparse grids instead. In their algorithm, the dynamic programming problem of the agent is affected by the *curse of dimensionality* because it has to take into account of the action of each other agent in equilibrium, so that adaptive sparse grids are used as a tool to perform higher-dimensional dynamic programming.

3 Example Economy

In order to illustrate the algorithm, I introduce a model whose solution poses a series of computational challenges. For the sake of clarity, the explanation of the algorithm will often refer to components of this economy. Still, the method has broader validity, and I chose the model to which it is applied only because, despite its small size, it contains a good number of interesting methodological features.

Time is indexed by $t \in \mathbb{N}$. The economy is populated by a unit mass of identical households, by a unit mass of identical final firms, and by a unit mass of intermediate firms subject to idiosyncratic demand shocks.

Market incompleteness in the form of non-insurable idiosyncratic risk and aggregate uncertainty in the form of time-varying distribution of idiosyncratic shocks imply that intermediate firms are subject to default (and exit) risk; at the same time, households can finance the creation of new firms.

3.1 Exogenous Processes

The idiosyncratic and aggregate shocks follow two autoregressive processes, given respectively by:

$$\log(z_t) = \mu_t + \rho_z \log(z_{t-1}) + \sigma_{t-1}\varepsilon_t \tag{1}$$
$$\log(\sigma_t) = (1-\rho_\sigma)\log(\mu_\sigma) + \rho_\sigma \log(\sigma_{t-1}) + v_t \tag{2}$$

where $\varepsilon_t \sim N(0,1)$, and $v_t \sim N(0,\varphi^2)$.

In this application, the two auto-regressive processes are approximated, using the method developed by Tauchen (1986), with Markov chains. The discretization generates n_s states for the aggregate shock ($n_s = 2$ implies there is a high and a low uncertainty state), and $n_z \cdot n_s$ states for the idiosyncratic shocks. Transition probabilities among idiosyncratic states, i.e. probability of z' given z, will be denoted as $\pi_z(z'|z)$.

3.2 Households

A unit mass continuum of identical households supplies labor to firms, consumes non-durables produced by final good firms, owns a unit mass distribution of firms. There is a representative household, and we will refer to it as a single agent, and it has flow utility given by:

$$U(C_t, N_t) = \frac{C_t^{1-\eta}}{1-\eta} - \theta \frac{N_t^{1+\chi}}{1+\chi} \tag{3}$$

where C_t and N_t are, respectively, aggregate consumption and hours. Define the aggregate state variable S_t as $S_t = (\sigma_{t-1}, \sigma_t, N_t, C_t, Y_t)$ at time t, where σ_t is the aggregate shock known at time t [6].

Households have access to a saving and borrowing technology, which can be interpreted as a riskless bond, so that the representative household's budget constraint can be written as:

$$C_t + Q(\sigma_t|S_t)B_{t+1} = w_t(S_t)N_t + B_t(\sigma_{t-1}) + D_t(S_t) - T_t(S_t) \tag{4}$$

where $Q(\sigma_t|S_t)$ is the security price, B_t is aggregate borrowing, $w_t(S_t)$ is the hourly wage, $D_t(S_t)$ are dividends accruing to households from firms, and $T_t(S_t)$ is a lump-sum tax levied by the government to finance a subsidy handed to workers of defaulting firms. Solving the household problem yields the usual stochastic discount factor Q, and the

[6] Timing structure is an important feature of this model: the index t refers to the moment in which agents learn about the realization of σ, but idiosyncratic shock at t depends on the previous, and not the current, realization of it: this way, agents know that whether next periods' idiosyncratic shocks will be more, less or equally volatile than current shocks

wage w_t. Finally, consumption is determined by market clearing, using the condition $Y = C + I$, where I denotes the resources used to activate new firms, following the dynamics described below.

3.3 Firms

Perfectly competitive final good firms purchase inputs y_t at prices $p_t(x_t)$ from intermediate good firms, then produce Y_t according to the technology:

$$Y_t = \left(\int z_t y_t(n_t, b_t, z_t)^{\frac{\gamma-1}{\gamma}} d\Upsilon_t(n_t, b_t, z_t) \right)^{\frac{\gamma}{\gamma-1}} \tag{5}$$

where Υ_t is the measure of active intermediate firms, indexed by their characteristics: z_t is the idiosyncratic shock to demand of their good, n_t the choice of hours for period t, and b_t the amount of debt due in period t. Good y_t is produced by monopolistic intermediate firms using technology $y_t = n_t^{\alpha}$ and sold to final good firms at price $p_t(n_t, b_t, z_t)$, so that demand for intermediate goods is given by:

$$y_t(n_t, b_t, z_t) = \left(\frac{z_t}{p_t(n_t, b_t, z_t)} \right)^{\gamma} Y_t \tag{6}$$

Intermediate goods firms are subject to idiosyncratic demand shocks that can generate temporary losses, which can be covered with non-contingent debt issuance. I abstract from agency problems and assume the manager maximizes the expected value for shareholders. The timing of their problem can be described as follows: at the beginning of time t, each firm observes current prices Q, p and w, the current uncertainty state σ_t, its past choices n_t, b_t, and its demand shock z_t. Each firm then has to make its choices on how much labor to purchase (it will be employed and paid in the following period), and how much to borrow, depending on their forecasted prices p and w. Risk-neutral financial intermediaries observe firm choices, assess the probability of default of each firm and set the price of credit accordingly.

The main friction comes into play here: firms may - and often will - be subject to a form of debt overhang, as described by Myers (1977), in that past debt choices affects current choices of labor. This, in turn, will make the cross-sectional distribution relevant for equilibrium prices.

The problem of incumbent firms can thus be written as:

$$V(n_t, b_t, z_t, S_t) = \max_{n_{t+1}, b_{t+1}} \kappa d_t + (1-\kappa)\beta \mathbb{E}\left[Q(\sigma_t | S_t) V(n_{t+1}, b_{t+1}, z_{t+1}, S_{t+1}) \right] \tag{7}$$

subject to

$$d_t = p_t n_t^{\alpha} - w_t n_t - b_t + q_t b_{t+1} \geq 0 \tag{8}$$

Since firms have a credit line consisting of the (endogenously determined) maximum amount of resources that can be borrowed in each period, they can be forced to default when a very low realization of z_t occurs and there is no policy $\{n_{t+1}, b_{t+1}\}$ such that $d_t \geq 0$ [7].

If funds generated and borrowed in the current period are sufficient to cover existing debt, the firm continues; otherwise, it defaults and pays wages first, then the rest is claimed by debtholders, who get $y_t p_t - w_t n_t$. If wages cannot be paid in full, the proceeds of lump-sum taxes levied on households are used to pay workers. The defaulting firm then exits.
The model also has a continuum of potential entrants, characterized by the same demand shock process as active firms. Each firm can be "activated" by paying a fixed cost ξ (the cost is sustained by households), and enters the economy with zero debt. The potential entrant problem is thus:

$$V^e(z_t, S_t) = \max_{n_{t+1}} \kappa\xi + (1-\kappa)\beta\mathbb{E}\left[Q(\sigma_t|S_t)V(n_{t+1}, 0, z_{t+1}, S_{t+1})\right] \tag{9}$$

Clearly, a firm will be started only if $V^e(z_t, S_t) > 0$.

Modeling choices in the firm problem may seem a bit ad hoc, but they reflect an important feature of business cycles: the constraint on dividends can be seen as if the firm is only able to access debt to finance its obligations, whereas other sources of funding (most importantly, equity) are unavailable. Also, as discussed above, small firms tend to increase their default rate more than big firms during recessions. They are, in general, more sensitive to shocks during bad times, consistently with evidence summarized in Fort et al. (2013). In sum, intermediate good producers in this model can be seen as small firms, subject to a series of frictions in their access to capital and risk-sharing[8].

3.4 Financial Intermediaries

Financial intermediaries operate in a perfectly competitive market, borrowing funds from households and lending to firms with an endogenous discount schedule á la Eaton and Gersovitz (1981).

Since competition prevents intermediaries from achieving positive profits, discount rates $q_t(S_t, z_t, n_{t+1}, b_{t+1})$ will simply equal the expected value of the amount that each firm pledges to repay to the intermediary. Borrowers will obtain the discounted amount of the face value of the bond, multiplied for the probability of the firm being active next period, plus the expected value of what is left in the firm in case of default.
Bond rates will then be denoted as:

[7] Notice that the most relevant friction in this case is the absence of state-contingent debt, as firms cannot insure against a specific realization of the idiosyncratic shock z

[8] More in general, the combination of endogenous entry dynamics, financial frictions that force firms to rely on debt only and production where labor is the most important input point to a model that can be interpreted as an economy of heterogeneous *small* firms.

$$q_t(S_t, z_t, n_{t+1}, b_{t+1}) b_{t+1} = \mathbb{E}\left[\phi_{t+1} Q(\cdot) b_{t+1} + (1-\phi_{t+1}) Q(\cdot) \max\{p_{t+1} y_{t+1} - w_{t+1} n_{t+1}, 0\}\right] \tag{10}$$

where $\phi_{t+1} = 1$ indicates the firm is continuing, and $\phi = 0$ indicates the firms is defaulting. Risk-neutral bond pricing from intermediaries implies a time-varying, endogenous borrowing limit for firms, which becomes more stringent in bad times, inducing the typical pro-cyclical borrowing behavior that is seen in models of strategic default. Even if default is not strategic here, it still happens that, in good times, relatively less profitable firms are kept alive by buoyant financial markets; conversely, some relatively more productive firms find themselves *underwater* during downturns.

3.5 Government

The government has a limited role in this economy: it runs a balanced budget, levying lump-sum taxes on all household to finance the difference between payments that defaulting firms owe to workers and the remaining resources within those firms.

3.6 Parameters

Most calibration choices are reflecting commonly used values. Table 2 reports parameter values in three blocks: households, firms and exogenous processes. Labor share α does target the actual labor share in the US, which averages around 60% in US postwar data. Discount rate β targets a (riskless) rate of return close to $1,5\%$, risk aversion of 1 means household has log-preferences, and the choice of χ implies a labor elasticity of 2, which is a standard calibration.

The parameter κ is introduced in the firm problem to capture a tension between shareholders and managers as in Jensen (1986): incomplete markets and non-negativity of dividends would give firms an incentive to build up savings, the parameter is instead chosen as summarizing a contractual agreement that induces managers to borrow. Elasticity of substitution between goods γ is parametrized in order to give intermediate firms a markup $\frac{\gamma}{\gamma-1}$ (see equation (11)) that is around 15%. Entry costs target U.S. capacity utilization in manifacturing, which fluctuates around 75%.

Calibration of the exogenous shocks reflects estimates in Stanfield et al. (2008), that are used in Arellano et al. (2016) to target moments in the distribution of firms based on Compustat data. Finally, the two autoregressive processes are discretized into a Markov chain with n_z states using the technique proposed by Tauchen (1986) and setting $n_z = 10$.

Parameter	Symbol	Value
Discount factor	β	0.985
Risk aversion	η	1
(Inverse) labor elasticity	χ	0.5
Labor share	α	0.6
Jensen effect	κ	0.4
Entry cost	ξ	1.0
Elasticity of substitution	γ	7.7
persistence of idiosyncratic shock	ρ_z	0.7
persistence of aggregate shock	ρ_σ	0.75
std of of aggregate shock	φ	0.1
mean of uncertainty process	μ_σ	0.09
number of idiosyncratic states	n_z	10

Table 1: Parameter calibration

4 Solving the Model

The present model, in its relative simplicity, contains a collection of methodological difficulties. Starting from the firm problem, notice how the presence of default risk induces a discontinuity of the value function in (8), as some combination of choices for hours and debt imply $d_t < 0$. The non-negativity constraint then binds, so the firm exits and thus its value function *jumps* to zero. Moreover, solving the firm problem requires knowledge of some mapping between current state space and future aggregates and prices. Obtaining this forecasts is the actual challenge of this model. The firm problem is not standard, but a simple approach for solving it, i.e. iterating on the value function, is still an option once future prices are at hand. Other options, however, are on the table: in Appendix A I discuss how to implement a version of the Howard acceleration based on Ljungqvist and Sargent (2012), that I use in the paper. The next subsection discusses the role of market incompleteness in making it hard to obtain an accurate forecast of future prices.

4.1 Addressing the importance of distributional channels

In principle, firms should also be aware of the full cross-sectional distribution of other firms' choices, which is an infinite-dimensional object, and which determines future aggregates. In fact, to obtain the expected values for N, Y, C, i.e. $\hat{C}_{t+1} = E_t(C_{t+1}|\mathscr{F}_t)$, $\hat{Y}_{t+1} = E_t(Y_{t+1}|\mathscr{F}_t)$ and $\hat{N}_{t+1} = E_t(N_{t+1}|\mathscr{F}_t)$, where $\mathscr{F}_t$ is the information set of the firm at time t, one has to calculate an expectation over policy functions that takes into account the whole history of shocks occurred to each agent, in an economy where the number of agent is infinite. Such integral is clearly intractable, representing an obstacle to a solution that

involves rational expectations. Different solution strategies are typically characterized by how they deal with this specific issue.

Most of the algorithms in the literature, as discussed above, are extremely successful in obtaining an accurate solution (in some cases, even for models with high-dimensional state spaces, large shocks and discontinuities in the agents' value and policy functions) under particular conditions. The most important is that aggregate dynamics are mainly driven by the fraction of agents whose choice are unconstrained - and thus independent of their own idiosyncratic state or history. In other circumstances, departures from approximate aggregation can be quite severe. To see this in my model economy, take labor choice by firms. In an unconstrained environment firm problem has a closed form solution, and optimality implies that the marginal product of labor equates a constant markup over wage, i.e.:

$$\mathbb{E}\left(p(z_t)\right)\alpha n_t^{\alpha-1,*} = \frac{\gamma}{\gamma-1}w_t \tag{11}$$

As pointed out by Arellano et al. (2016), when financial markets are incomplete and there is aggregate risk, the firm problem changes. Default risk implies there is some threshold value for the idiosyncratic shock, call it $\hat{z}$, below which the firm with past choices n_t, b_t will always default. So the optimality condition becomes:

$$\mathbb{E}\left(p(z_t)\right)z_t\alpha n_t^{\alpha-1,*} = \frac{\gamma}{\gamma-1}\left[w_t + V\frac{\pi_z(\hat{z})}{1-\Pi(\hat{z})}\frac{d\hat{z}}{dn_t^*}\right] \tag{12}$$

where $\Pi(\hat{z})$ is the cumulative distribution function associated with the density π_z. Labor choices impact the threshold value $\hat{z}$, and through it, the hazard rate $\frac{\pi_z(\hat{z})}{1-\Pi(\hat{z})}$: therefore, all firms' choices are to some extent distorted by market incompleteness, and distributional channels become relevant. Notice that the effect of this distortion of the individual firm does not have necessarily the same sign for all levels of the idiosyncratic shock: indeed, as we will see, firms with high z_t will produce more in a high-uncertainty environment, because of general equilibrium effects. At the same time, inactive firms that draw a higher z_t enter more often in high uncertainty times, counterbalancing reduced output by existing firms whose shock is low-to-medium.

To address this additional complexity, a strand of literature starting from Den Haan and Marcet (1990), argues for approximating expectations with the use of polynomials; for heterogeneous agents' models, Algan et al. (2008) propose a polynomial approach for approximating cross-sectional distribution of agents in a global solution scheme with a flexible functional form. A similar strategy would directly take into account the importance of higher order moments, and thus take into account distributional channels directly, by using an approximation of the cross-sectional density in the forecasting of aggregates.

4.2 Cross-Sectional Density Approximation

Global solution algorithms typically use quadrature methods to integrate individual policies into aggregates. I build on the contribution of Algan et al. (2008) to the latter approach, which is to use two sets of moments in the solution algorithm: one characterizes the aggregate state space, and determines the grid over which we have to approximate an aggregate law of motion; the other is composed of "reference" moments, which pin down the cross-sectional distribution through a flexible functional form, but are not used as state variables. The separation into variables that enter directly into agents' problems and reference moments is what makes the approach feasible: if one had to take higher moments into account explicitly, the dimensionality of the problem would quickly explode.

The calculation of reference moments, as the formulation of the functional form for approximated density, implies choices on the solution strategy which depend on the characteristics of the model. In this case, I make the simplifying assumption that the distributions of individual choices in debt and labor are independent of each other. Independence allows to characterize the relevant distribution for obtaining aggregate hours and production in one dimension, which greatly improves the stability and convergence properties of the algorithm. Reference moments and coefficients can be updated in the solution by using simulation outcomes: as in Den Haan and Rendahl (2010), the algorithm starts from the computation of reference moments from a steady state solution with no aggregate uncertainty. Simulation then supplies information on the behavior of the model in the ergodic set. Another option is to adjust the first reference moments using information from the approximate law of motion.

A classical choice for the flexible form is the following:

$$P(n,\rho^{z_i}) = \rho_0^{z_i} exp\left\{\rho_1^{z_i}\left(n-m_1^{z_i}\right)+\rho_2^{z_i}\left[\left(n-m_1^{z_i}\right)^2-m_2^{z_i}\right]\right\}\times\cdots$$
$$\cdots\times exp\left\{\rho_{n_M}^{z_i}\left[\left(n-m_1^{z_i}\right)^{n_M}-m_{n_M}^{z_i}\right]\right\}$$

where $\left\{m_1^{z_i},...,m_{n_M}^{z_i}\right\}$ are the $n_z n_M$ reference moments, initially calculated from a steady state solution without aggregate uncertainty[9], n_z is the number of points in which the idiosyncratic process is discretized, and n_M is the polynomial degree of the functional form. Once the reference moments are calculated for each aggregate state, one obtains a convex minimizazion problem, in that coefficients $\rho_1^{z_i},...,\rho_{n_M}^{z_i}$ solve

$$\min_{\rho_j^{z_i}, j=1,...,n_M} \sum_{i=1}^{n_z}\int_{\underline{n}}^{\bar{n}} P(n,\rho^{z_i})dn \tag{13}$$

[9] the steady state solution yields reference moments following the histogram approach of Young (2010)

and the solution immediately follows from finding the zeros of the system of first order conditions of (13). Notice that $\rho_0^{z_i}$ is calibrated in order to have $\int_{\underline{n}}^{\bar{n}} P(n, \rho^{z_i}) dn = 1$.

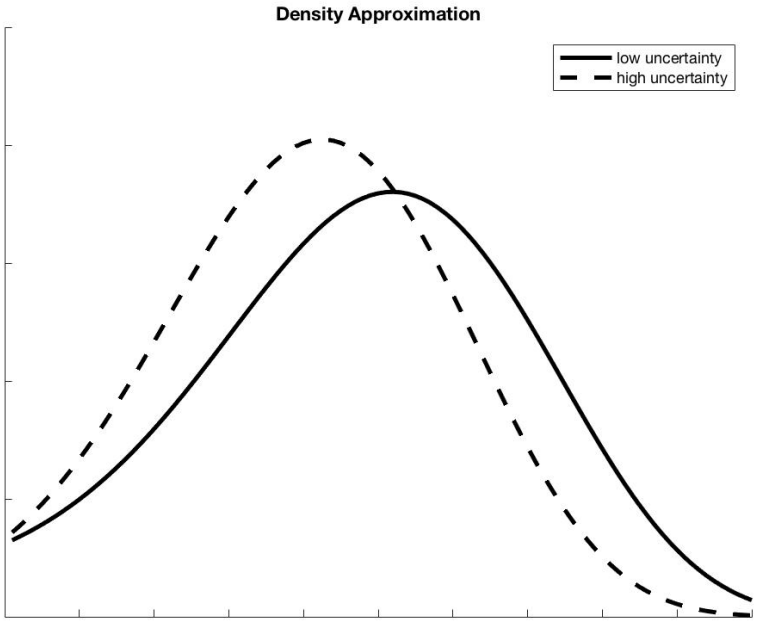


Figure 1: $P(n, \rho^z)$, with reference moments and coefficients from steady state

The integral is computed on a given number of Simpson nodes, where the values of individual policies are obtained through interpolation, using splines. Once reference moments and coefficients are at hand, one can obtain aggregate variables using the measure obtained from the flexible functional form on the Simpson integral. This allows to compute aggregate hours and production, as:

$$N_t = \sum_{i=1}^{n_z} \tilde{\pi}_i^z \int n_{t+1}^*(n_t, b_t, z_t, S_t) P(n, \rho_\sigma^{z_i}) dn \tag{14}$$

$$Y_t = \left[\sum_{i=1}^{n_z} \tilde{\pi}_i^z \int z_t y_t(n_t, b_t, z_t)^{\frac{\gamma-1}{\gamma}} P(n, \rho_\sigma^{z_i}) dn \right]^{\frac{\gamma-1}{\gamma}} \tag{15}$$

where $\tilde{\pi}_i^z$ is the asymptotic distribution of the Markov chain that discretizes the exogenous process for idiosyncratic shocks. Using flexible functional forms allows another use of simulated data: the one-step-ahead forecast can be used to adjust the first reference moments of the parametrized distributions, in order to follow the time-varying

cross section of agents' choices. This, however, doesn't come at no cost: to get the one-step ahead forecast of higher moments, one has to include them explicitly in the law of motion to keep track of.

4.3 Using Adaptive Sparse Grids

Appropriately interpolating on a discretized grid over the aggregate state space would, in principle, yield a sufficiently accurate approximation of the law of motion of the economy. However, a discretization of the full aggregate state space is computationally burdensome even in a relatively small model like the one presented in this work, where the importance local dynamics requires a fine grid in each dimension. As discretization proceeds with adding points, finer Cartesian grids see dimensionality growing exponentially, in a classic display of the *curse of dimensionality* problem.

To understand why this matters here, notice that elements that enter directly into the firm problem at t are three future prices: equilibrium wage, prices of intermediate goods, and the stochastic discount factor. Those are, in turn, determined in equilibrium by three continuous variables (consumption, hours, output) and one discrete variable (aggregate uncertainty). Hence, a minimal representation of the the economy at time t has to include at least these three aggregates. In solving the benchmark model of Aiyagari (1994), where heterogeneous agents face non-insurable unemployment risk, Algan et al. (2013) include the measure of unemployed agents in the set of state space variables. This is done in order to better keep track of changes in the cross-section of agents. Similarily, I include the measure of active firms μ_t. This brings the dimension of the state space to four (continuous) plus two (discrete) states - already problematic if one wants to use Cartesian grids.
The point is illustrated in Table 2.

d	$\lvert V_4\rvert$	$\lvert V_4^L\rvert$	$\lvert V_4^G\rvert$
1	17	17	9
2	289	65	28
3	4913	177	84
4	83521	400	210
5	$1419 \cdot 10^3$	801	462
10	$2015 \cdot 10^9$	8801	8008

Table 2: Number of gridpoints for different grid types of level 4. First column reports the dimension; second column reports the number of points for a full Cartesian grid; third column reports local polynomial grids that constitute the starting point of adaptive sparse grids before refinement; fourth column reports classical (global) sparse grid induced by Lagrange polynomials

In the model economy presented in this work, the main obstacle to an accurate

approximation of the law of motion is the relevance of important local sharp behavior in given areas of the aggregate state space. In particular, the behavior of the solution depends crucially on the extent to which there is firm entry or not in that particular point[10] . The algorithm will adaptively add points in regions of the state space with more curvature - this implies the need to choose a refinement threshold below which the approximation is considered sufficient and points are not added. However, as argued above, different refinement strategies are not equivalent, and choosing one against the other is a possible source of stark differences in algorithm performance.
In implementing the solution algorithm, I postulate the existence of a box that contains the endogenous state; sparse grids will be defined in this box $\mathbf{B} \subset \mathbb{R}^{\mathbf{N}}$. However, there is no guarantee that resources in approximating the law of motion will be spent in the most relevant areas of the state space - i.e. at values visited by the solution. Solutions will all be inside the box, if it is suitably large, but they could lie in a rather small fraction of its total area. Moreover, even if information about the ergodic set can be learned from simulating the model, the use of this information then depends on the algorithm chosen. In the case of my method, the researcher can update the boundaries of set $\mathbf{B}$, and the flexible form for the cross-sectional density approximation can be updated with new reference moments and coefficients. However, unfortunately, the method cannot yet account for irregular geometries. For a discussion on why the geometry of state spaces is relevant, and on how to deal with possible irregular shapes, see Scheidegger and Bilionis (2017).

4.4 From Classic to Adaptive Sparse Grids

Obtaining the solution to a dynamic programming problem, the aggregate law of motion of a model economy, or even the coefficients of a statistical model over some training set, can be seen as performing the interpolation of a function, say $f : \Omega \to \mathbb{R}$ that is known only algorithmically. In those cases, even if f is not known, it can be evaluated at points in its domain using a numerical procedure. A discretization of the domain of interest Ω to obtain those points is trivial and obviously not problematic for lower dimensional problems.

In the simplest case, the first step for building an interpolant u of f is to construct a Cartesian grid G over Ω - i.e., a grid constituted of equidistant grid points with mesh width $h_n = 2^{-n}$ - then evaluate f at G. A piece-wise linear interpolant can then be obtained as the weighted sum of piecewise linear basis functions $\varphi_i(x)$:

[10]The reason why entry matters has to do with the presence of mass points induced by the discretization of the idiosyncratic shock process. It is useful to consider the entrant problem in my example economy: optimal hours chosen by entrants depend only on z_t and on forecasted aggregates. Therefore, if z_t can take only finite many values, at every discretized value of the shock corresponds a positive mass point of firms choosing the same employment level, making the cross-sectional approximation less accurate.

$$f(x) \approx u(x) = \sum_i \alpha_i \varphi_i(x)$$

The coefficients α_i will be obtained as solutions to a (linear) system of equations generated by function evaluations at G. It is easy to see that this approach will face the curse of dimensionality rather soon. Moving to the d-dimensional case, basis functions can be obtained via tensor products, so that:

$$\varphi_i(x) = \Pi_{j=1}^{d} \varphi_{i_j}(x_j)$$

with i being the index for each dimension. Define V_n as the space of piecewise d-linear functions with mesh width h_n, so that $V_n = \bigoplus W$, with $W = \text{span}\left\{\varphi_i(x)\right\}$.
The asymptotic interpolation error decay is:

$$\left\| f(x) - u(x) \right\| \in \mathscr{O}\left(h_n^2\right) \tag{16}$$

The function can thus be satisfactorily approximated if it is sufficiently well behaved, but this requires $\mathscr{O}\left(h_n^{-d}\right) = \mathscr{O}\left(2^{nd}\right)$ function evaluations - and here lies the curse of dimensionality.

A well-known strategy to overcome this issue relies on the hierarchical decomposition of the approximation spaces which allows to build sparse grids. I will illustrate the the univariate case for illustrative purposes, as the same reasoning carries over to the multidimensional case with just heavier notation. Consider a standard hat function:

$$\varphi(x) = \max\left\{1 - |x|, 0\right\} \tag{17}$$

from which one-dimensional hat basis functions at level l can be derived as

$$\varphi_{l,i}(x) = \varphi\left(2^l x - i\right) \tag{18}$$

with $0 < i < 2^l$. The set of hierarchical subspaces is obtained as:

$$W_l = \text{span}\left\{\varphi_{l,i}(x) : i \in I_l\right\} \tag{19}$$

The space of piecewise linear functions for a given level l can then be expressed as a direct sum of the hierarchical subspaces, i.e. $V_n = \bigoplus_{l \leq n} W_l$. The basic intuition behind sparse grids is to exclude ex ante those subspaces that contribute little to the interpolant as they have many basis functions and small support. The procedure yields the sparse grids space

$$V_n^{(1)} = \bigoplus_{|l|_1 \leq n+d-1} W_l \tag{20}$$

Under some regularity conditions (namely, boundedness of mixed second derivatives),

the resulting sparse grid is optimal with respect to the L_2-norm and the maximum-norm. Different choices of error measurement would lead to different optimal grids. The advantage of sparse grids is evident: while asymptotic error decay is now

$$\left\| f(x) - u(x) \right\| \in \mathscr{O}\left(h_n^2 (\log h_n^{-1})^{d-1} \right) \tag{21}$$

the number of function evaluations decreases sharply to $\mathscr{O}\left(h_n^{-1} (\log h_n^{-1})^{d-1} \right)$.

If the function of interest exhibits sharp local behavior - discontinuities, kinks, or even steep regions - classic sparse grids will not provide an accurate approximation. Those limitations, however, can be addressed if adaptivity is used: in this case, refinement depends on local error estimation, and suits the problem at hand. In adaptive sparse grids implied by local rules, interpolation is based again on hierarchical piecewise polynomials with local support and varying order. In contrast to classical sparse grids, however, adaptive sparse grids use functions with support restricted to a neighborhood of each point and adopt a iterative refinement strategy. This implies that, as the hierarchy of points in the grid advances to lower levels, points are added based on function evaluations and error estimation. In the context of my example economy, this implies concentrating in areas of the aggregate state space S_t where the aggregate law of motion, call it Γ, has more curvature or sharper behavior. The solution is then approximated as

$$\Gamma(S_t) \approx \sum_{k=1}^{l} \sum_{i \in I_l} \alpha_{k,i} \varphi_{k,i}(S_t) \tag{22}$$

where φ are basis functions of arbitrary order evaluated at grid points on the aggregate, high-dimensional, state space. What is important to highlight, here, is that coefficients $\alpha_{k,i}$, called *refinement surpluses*, are generated directly from the refinement of the grid. Different refinement strategies are discussed in the next subsection.

4.5 On refinement strategies

Consider the single-dimensional case, first. The hierarchical structure in sparse grids can be highlighted by ordering all points in the grid G according to an index-to-level map $g(j) : \mathbb{N}_0 \to \mathbb{N}_0$, so that point x_j is associated with level $l = g(j)$. Define level index sets D^l as the indexes of points in level l. Cumulative level sets V^l are defined as all points at levels less than or equal to l, which implies $V^0 = D^0$, and $V^l = V^{l-1} \bigcup D^l$. Each point x_j is associated to a hierarchical family structure, so that:

$$P_j = \left\{ i \in \mathbb{N}_0 : x_i \text{ is a parent of } x_j \right\} \subset D^{g(j)-1} \tag{23}$$

$$O_j = \left\{ i \in \mathbb{N}_0 : x_i \text{ is a child of } x_j \right\} \subset D^{g(j)+1} \tag{24}$$

A graphical illustration of the parent-children structure is offered in Figure 1, taken

from (Ma and Zabaras, 2010) :

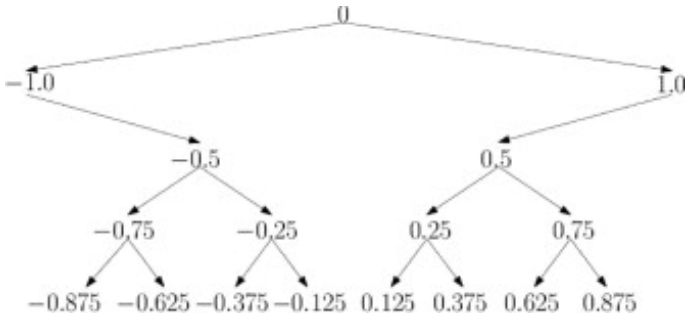


Figure 2: Data structure

The first point, at 0, belongs to D^0 and obviously has no parents. Point -0.5 is in D^2, has -1.0 as a parent (with $-1.0 \in D^1$) and $\{-0.75, -0.25\}$ as children. Refining the grid under this framework imposes no methodological issues: given basis functions $\left\{\varphi_j\right\}_{j\in \mathtt{N}_0}$, coefficients α for the interpolant in (22) can be obtained by matching the value of the function at grid points. This already suggests the refinement strategy: one includes points from the next level of the hierarchy only around points associated with large surplus coefficients. Formally, define the set of large surpluses as:

$$B^l = \left\{ j \in V^l : \frac{|\alpha_j|}{f_{\max}} > \varepsilon \right\} \tag{25}$$

for desired tolerance $\varepsilon > 0$. The set of large surpluses yields the indexes of points that are candidates for belonging to the refinement set R_j^l. A way to determine the convergence of the interpolation algorithm is to reach a level for which $V^{l+1} = V^l$, or $R_j^{l+1} = \emptyset$. A graphical illustration of the construction of adaptive grids in this context can be found in Appendix C.

In the multidimensional case, however, the situation is not so simple. As (Bungartz and Dirnstorfer, 2003) point out, refinement in multiple dimensions might lead to unbalanced grids and children nodes being generated with an empty parent sets (so called "orphans"), a feature that can undermine accuracy, or lead to unstable results. Formally, define the set of large surpluses in the multidimensional context as:

$$\mathbf{B}^l = \left\{ \mathbf{j} \in \mathbf{V}^l : \frac{|\alpha_\mathbf{j}|}{f_{\max}} > \varepsilon \right\} \tag{26}$$

The refinement set $\mathbf{R}_j^l$ is not determined mechanically now. For instance, it is possible that some point x_j is associated with a large coefficient, but that some parents of x_j are not included in the interpolant. Also, the researcher has to decide whether to refine isotropically (i.e., in all directions) or anisotropically (i.e., in a selected set of directions).

The first approach to refinement of adaptive sparse grids (hence the label "classic"

refinement) is isotropic, and adds only the children of points with large coefficients. Refinement sets is then defined as:

$$\mathbf{R}_j^l = \bigcup_{\alpha=1}^{d} O_j^\alpha \tag{27}$$

The classic refinement approach is typically working fine for most problems. However, adding points in every direction can be inefficient in higher-dimensional problems, increasing the computational burden. In other cases, failure to include the full family of a point can lead to instability around *orphans*.

To deal with the latter issue, the family-selective refinement approach checks whether parents of every point in $\mathbf{B}^l$. If they are not part of the grid, it adds them first, otherwise it adds children. The refinement set is thus:

$$\mathbf{R}_j^l = \left(\bigcup_{\alpha \in \Delta_j^l} P_j^\alpha \right) \bigcup \left(\bigcup_{\alpha \notin \Delta_j^l} O_j^\alpha \right) \tag{28}$$

where Δ_j^l is the set of *orphan* points at level l.

On the other hand, the problem with the classic approach can be that too many points per level are added on straight lines like the main axis, the boundaries, and so on. The direction-selective refinement approach aims at limiting this tendency by adding points only on Λ_j^l, defined as the set of directions associated with a large one-directional coefficient. Hence the refinement set becomes:

$$\mathbf{R}_j^l = \bigcup_{\alpha \in \Lambda_j^l} O_j^\alpha \tag{29}$$

As for the classic approach, this refinement strategy can be fragile when failure to add parents leads to instability.

For problems where instability can be a problem but it is limited to a small portion of the function domain, a combination of the advantages of the direction-selective and the family-selective refinement can be desirable. This is the case of the family-direction-selective refinement approach: each direction is considered separately, but children are added only if parents are not missing. Finally, the refinement set results in:

$$\mathbf{R}_j^l = \left(\bigcup_{\alpha \in \Lambda_j^l \cap \Delta_j^l} P_j^\alpha \right) \bigcup \left(\bigcup_{\alpha \notin \Lambda_j^l \setminus \Delta_j^l} O_j^\alpha \right) \tag{30}$$

To illustrate how different refinement schemes work, take the function:

$$f(x,y) = \frac{1}{|0.5 - x^4 - y^4| + 0.1} \tag{31}$$

with $x, y \in [0,1] \times [0,1]$. Brumm and Scheidegger (2017) show adaptive sparse grids are a suitable interpolation algorithm for f, while classic grids and sparse grids end up performing considerable worse[11]. The function is plotted in Figure 2 using points from the last level of the adaptive sparse grid. As refinement proceeds, some areas of the state space become "denser", while other are covered by very few points: computation resources are thus invested where they are more needed to obtain an accurate solution. The final grid, after just seven refinements, is at the bottom right of Figure 3. Tolerance parameter, which determines refinement, is constant and set at $1.E-4$ (more efficient interpolation would obtain allowing the tolerance parameter to adjust for different levels). The grid interpolates the function with a maximum absolute error of $3.63e-03$ on a training set of 1000 points scattered over the state space.

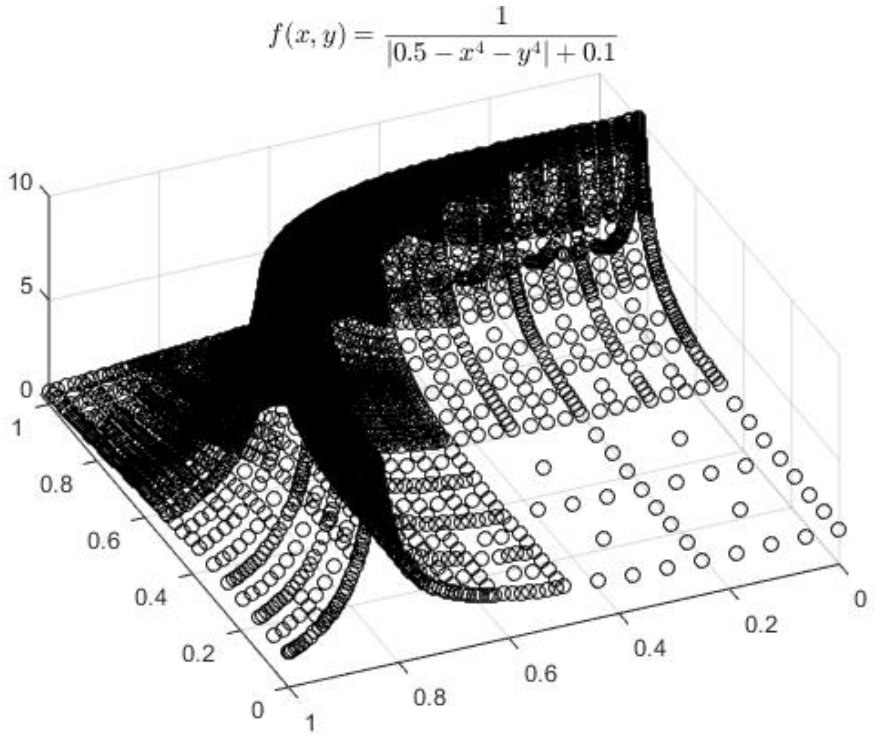


Figure 3: The example function

Since refinement is an iterative process (and points are added in "cohorts", from parents to children), the threshold need not be constant across refinement levels. Instead,

[11] Regarding classic sparse grids, notice that the regularity conditions are not satisfied by f: namely, the function is non-differentiable in a segment of its domain

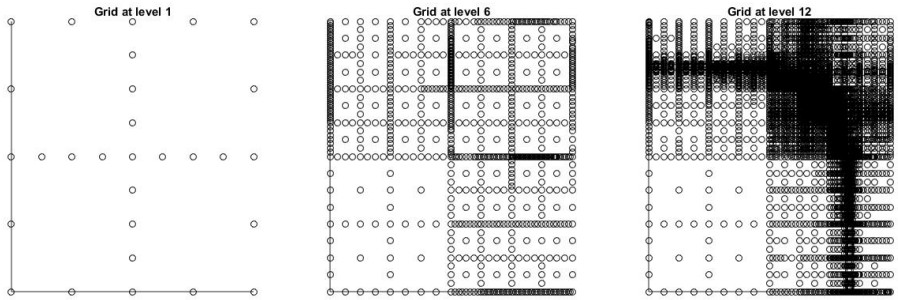


Figure 4: Refinement on the example function using FDS

there is room for adapting the criterion by taking into consideration the specificity of the problem. Clearly, this implies there is a tension between the degree of accuracy one can obtain: too high refinement threshold at the first stages, i.e. a coarse grid at the beginning, could miss important action in some regions of the state space; conversely, a too low refinement threshold, hence a finer grid, could invest too many resources in areas of low curvature and dampen the benefits of the whole strategy. As discussed above, some refinement strategies can loosen the efficiency-accuracy trade-off: the desirability of each strategy depends on the specificity of the problem, with parents-first being preferable for functions with local sharp behavior and family-direction-search over-performing alternatives as dimensionality of the problem grows.

Figure 4 compares the adaptive grid after the sixth refinement level for the same function. Points that would be present both under classic and FDS refinement are represented by empty dots, whereas I use full dots to highlight the position of points that are added only by the classic refinement strategy. Classic refinement starts adding more points, especially at the boundaries, already at such lower refinement levels. Notice that this happens even if, in principle, FDS could end up adding more points because of the "parents-first" component of the strategy.

5 The Algorithm

Before running the main iteration, one needs to obtain the Markov process - again using Tauchen (1986) - to obtain the chain for aggregate shocks to be used for the simulation later.

The algorithm will then need to use steady state reference moments and distributions; these can be obtained by solving the model without aggregate uncertainty. I then obtain coefficients by solving (13); notice that this can be done by taking the first order condition, and then using a Broyden–Fletcher–Goldfarb–Shanno method. With everything at hand, one can finally start the solution algorithm.

After the algorithm has reached convergence, the model is solved, it can be simulated and

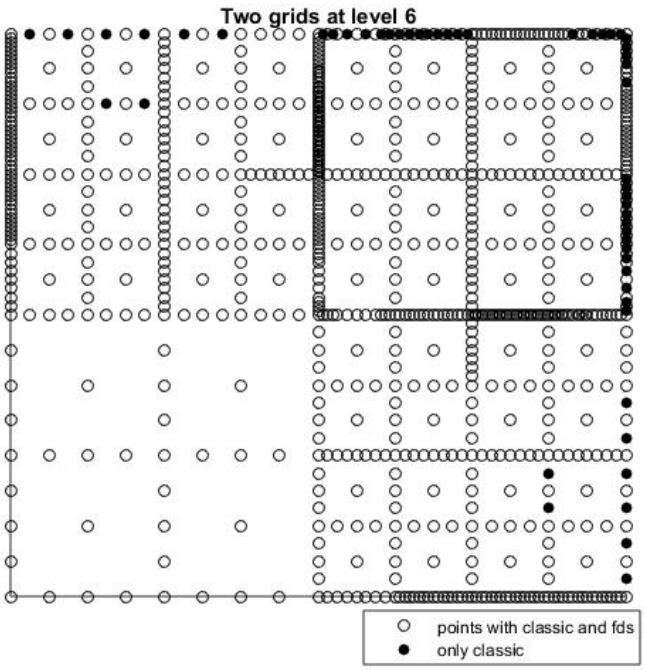


Figure 5: Refinement on the example function

used to perform impulse response analysis. However, as discussed above, other useful information on the behavior of the model can be obtained via simulation. In particular, new reference moments and coefficients can be obtained from simulated series. If one wants to follow this path, the solution algorithm adds another (outer) loop on reference moments and coefficients, which stops once it reaches some convergence criterion.
As discussed above, choice of threshold value $\bar{\varepsilon}$ is crucial: it turns out that, in this model, choosing "wrong" values either worsens accuracy or ends up producing unnecessarily dense grids in the first stages of iteration. However, as the starting value is not too low from the beginning, it is possible calibrate it dynamically, by having it adjusting to lower values with higher levels of refinement. Regarding the refinement itself, a reasonable prior would be to assume that family-direction search is the strategy that delivers the best performance. Relevant non-linearities around the endogenous threshold that separates areas of the state space where there is no firm entry and areas with positive firm entry, in fact, could prevent the algorithm from converging because of the instability around *orphan* points discussed in Section 2. I solve the model using classical refinement for robustness, and results confirm this intuition.

Data: Level of initial (classical, unrefined) sparse grid, $\tilde{l}$. Initial guess $\Gamma_0(\cdot)$ for the law of motion at every point of initial grid $G_{\tilde{l},0}$. Maximum number of refinement steps $l_{\max}$. Approximation accuracy ε for obtaining guess mapping and threshold value $\bar{\varepsilon}$ for hierarchical surplus in adaptive sparse grids interpolation, with $\varepsilon > \bar{\varepsilon}$. Coefficients $\rho_{z,m}$ and reference moments $\mathbf{m_z}$.

Result: The (approximate) equilibrium aggregate law of motion $\Gamma_{\tilde{l}}$, the corresponding policy functions $\sigma^*(\cdot,\cdot)$ for every firm type indexed by $\{n_t, b_t, z_t\}$, and entry decisions $I_t^*(z_t)$.

while $e > \varepsilon$ **do**

  for $g \in G_{\tilde{l},0}$ **do**

    With $\Gamma_0(g)$ at hand:
    obtain policy functions $\sigma^*(\cdot,\cdot)$ for active firms, and
    determine entry choices $I_t^*(z_t)$ of potential entrants.
    Clear markets and obtain aggregates, hence obtain $\hat{\Gamma}_0(g)$.

  end

  Obtain approximation error $e = \left\| \Gamma_0(G_{\tilde{l},0}) - \hat{\Gamma}_0(G_{\tilde{l},0}) \right\|$.

  Update $\Gamma_0(G_{\tilde{l},0})$ using $\hat{\Gamma}_0(G_{\tilde{l},0})$ with dampened fixed point iteration.

end

Now, randomly generate n training inputs $\mathbf{X} = \{x_i\}_{i=1}^n \subset \mathbf{B}$.

while $\bar{e} > \bar{\varepsilon}$ *and* $\tilde{l} \leq \tilde{l}_{max}$ **do**

  Update $l = l + 1$

  Use $\hat{\Gamma}_{l-1}(\cdot)$ to refine the grid, obtain $G_{\tilde{l},l}$

  Interpolate to get law of motion $\Gamma_l(G_{\tilde{l},l})$

  for $g \in G_{\tilde{l},l} \setminus G_{\tilde{l},l-1}$ **do**

    With $\Gamma_l(g)$ at hand:
    obtain policy functions $\sigma^*(\cdot,\cdot)$ for active firms, and
    determine entry choices $I_t^*(z_t)$ of potential entrants.
    Clear markets and obtain aggregates, hence obtain $\hat{\Gamma}_l(g)$.

  end

  Obtain approximation error $\bar{e} = \left\| \Gamma_l(\mathbf{X}) - \hat{\Gamma}_l(\mathbf{X}) \right\|$.

  Update $\Gamma_l(G_{\tilde{l},l})$ using $\hat{\Gamma}_l(G_{\tilde{l},l})$ with dampened fixed point iteration.

end

Algorithm 1: Overview of the critical steps of the solution algorithm.

The key of the present solution method lies in the way individual policy functions are aggregated across the state space. I obtain firms' policy functions using an adapted version of time iteration based on Ljungqvist and Sargent (2012). The algorithm for solving the firm problem is illustrated in the Appendix. Value functions of incumbent firms at $b = 0$ are then used to evaluate entry decisions. Consumption is obtained via market clearing equation $C_t = Y_t - I_t$ - this implies that general equilibrium effects come from changes in value functions. Notice that current and forecasted consumption determine the stochastic discount factor, then firms' value functions and hence decision by potential entrants. Since Y_t is a state, the inner loop starts by guessing an equilibrium value of $\tilde{C}_t$ determined by entry decisions, and iterates using bisection until guessed consumption converges to consumption implied by entry choices.

5.1 Parallelization

A useful feature of the solution method presented above is the possibility to exploit high-performance computing systems. The algorithm is based on Message Passing Interface (MPI), a communication protocol for parallel computers, based on a distributed shared memory concept. A detailed outline of MPI is offered in Gropp et al. (1999).

To measure of how the solution method benefits from parallelization, I show the algorithm exhibits strong scaling properties. To do so, I report on the single node performance[12] on different refinement levels (i.e. the computation of the whole grid $G_{\bar{l},l}$ at levels $l = 1,2,3...$) with increasingly larger numbers of MPI ranks. The first level of the adaptive grid in the low uncertainty state has 114 points - since this is a relatively small number of jobs, the maximum efficiency is reached at an already relatively low number of ranks. In general, parallel efficiency improves in the size of the grid, as the workload would be better distributed among different MPI processes. Since adaptive sparse grids do not increase monotonically in size, parallel performance doesn't necessarily increase as one proceeds to higher levels. Still, it is possible to say that scaling efficiencies will be even stronger in higher-dimensional models.

6 Results

6.1 Measuring Accuracy

Once the model is solved, evaluating the performance of the algorithm depends on its ability to obtain accurate approximations. In most dynamic models with heterogeneous agents, heterogeneity takes place at the household level. Hence a straightforward measure

[12] In this section I always refer to the architecture of the "Piz Daint" system installed at Swiss National Supercomputing Centre (CSCS).

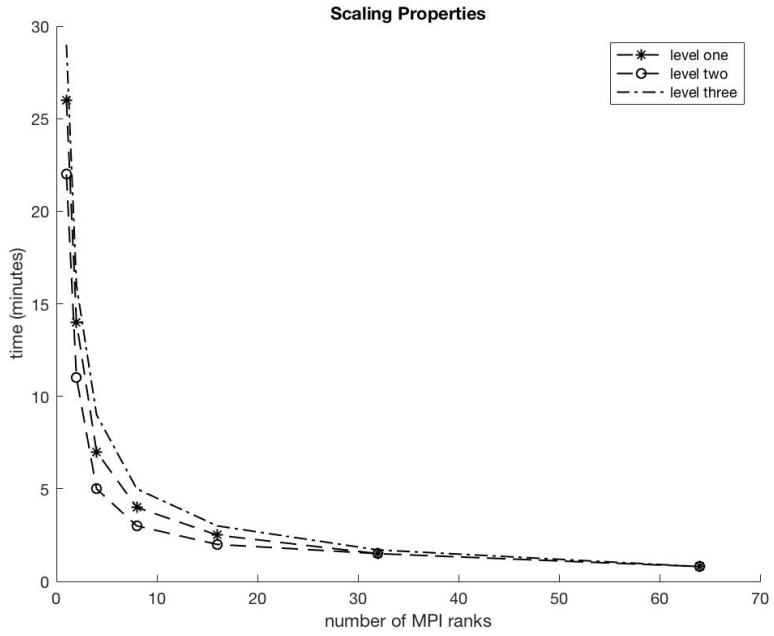


Figure 6: Performance for different levels of refinement, as number of MPI processes increase

of accuracy is given by some function of errors in the approximation of Euler equations over a set of points. In the model of this paper, the crucial task is the approximation of the aggregates law of motion, that firms use to forecast prices. It is thus tempting, when comparing the proposed method to more established solution strategies like Krusell and Smith, to report the R-squared for evaluating the accuracy of the latter. The analysis of this chapter will show that it would be misleading, consistently with issues highlighted by, among others, Den Haan (2010). Table 3 reports the maximum and mean percentage absolute differences between realized data and static forecasts for all of the relevant variables as accuracy statistics. Both are relevant measures: an approximation that tracks the dynamics nicely most of the time but is subject to enormous, albeit rare, deviations will not do a good job; an approximation that always stays close to realized data but noisily tracks the process will not be ideal either.

	Hours	Output	Consumption
Max	2.77	2.99	2.75
Mean	1.00	0.76	0.63

Table 3: Maximum and mean percentage differences between realized data and forecasts in the training set

It is an established tenet of the literature that such measures are not informative enough: they do not concentrate on the ergodic set, but consider all of the state space equally, and do not take into account of the extent to which forecast errors can cumulate over time if the actual law of motion is never observed by agents. While the second issue depends on specific model assumptions, the first is extremely relevant in a multi-dimensional setting, where areas of the state space are included in the state space hypercube but are very unlikely to be ever visited by the solution. In turn, this implies that using accuracy measures based on dynamic forecasts can in general show an improved performance in multi-dimensional settings when compared to static forecast accuracy measures. A first diagnostic can be made by looking at how forecasting accuracy is distributed over the training set $\mathbf{X} \subset \mathbf{B}$, as in Figure 4. Errors are concentrated in high-output, low-hours areas, which suggests that they might not affect heavily accuracy in simulations.

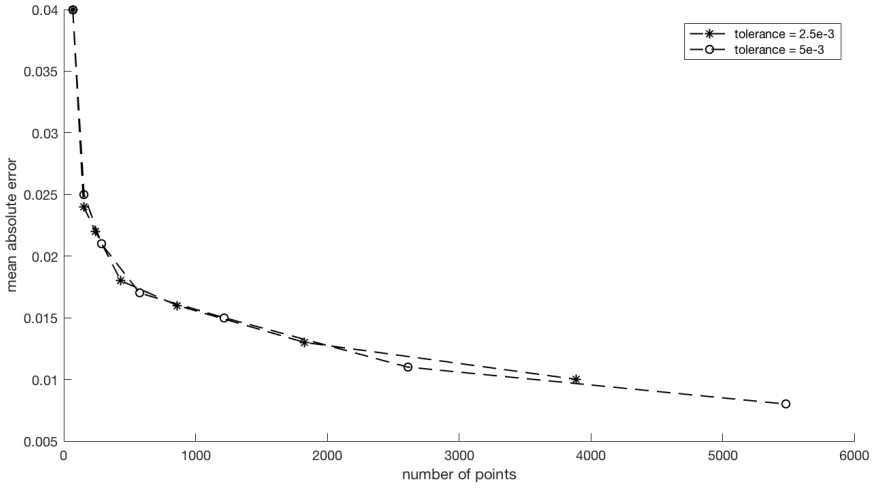


Figure 7: Accuracy over $\mathbf{B}$ for different refinement levels at $\sigma = \sigma_H$

The set of statistics from Table 4 provide an account of how accurate the solution is on the ergodic set: dynamic forecasts are computed from the absolute percentage differences of one-step and two-steps-ahead forecasts versus realized data for the relevant variables. Properties of approximation seem to be preserved, if not improved, in the ergodic set. A comparison with one-step errors from the Krusell and Smith simulation shows the improvement in accuracy, that is massive for hours, and still quite relevant for consumption. Figure 9 gives an example of simulated series and their approximation, which show that the law of motion is followed rather closely.

Another issue arises when it is to be established whether approximation errors are systematic: in other words, if the approximate law of motion is constantly leading to higher or lower expectations of realized values. A way to address this issue has been

proposed by Den Haan (2010): if the law of motion generates systematic errors, then n-steps ahead forecasts ($n > 1$) will progressively perform worse.

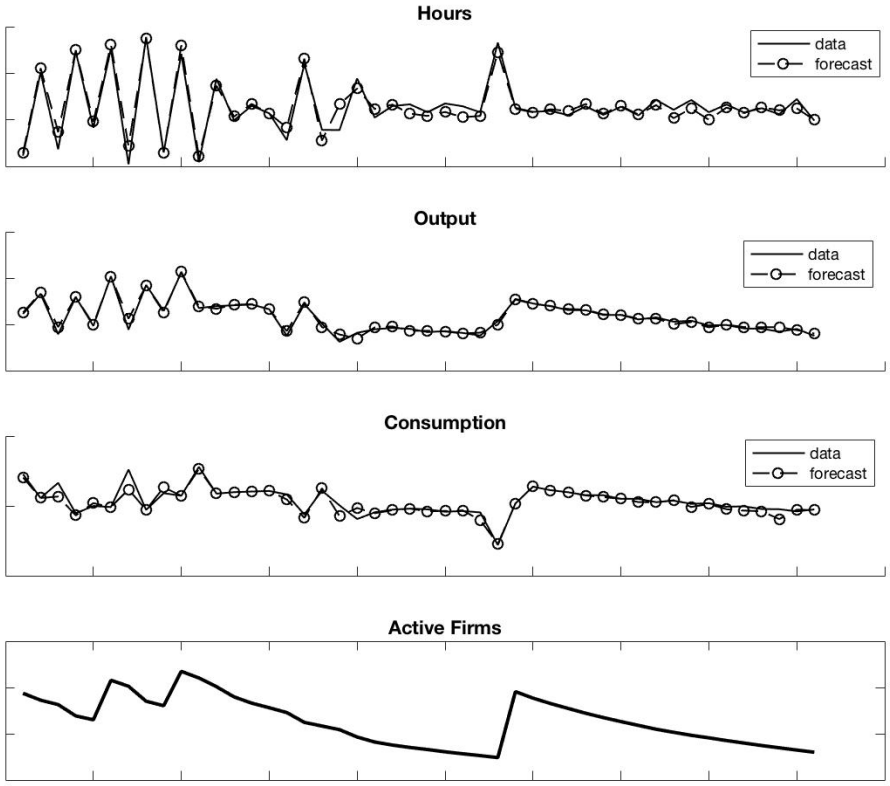
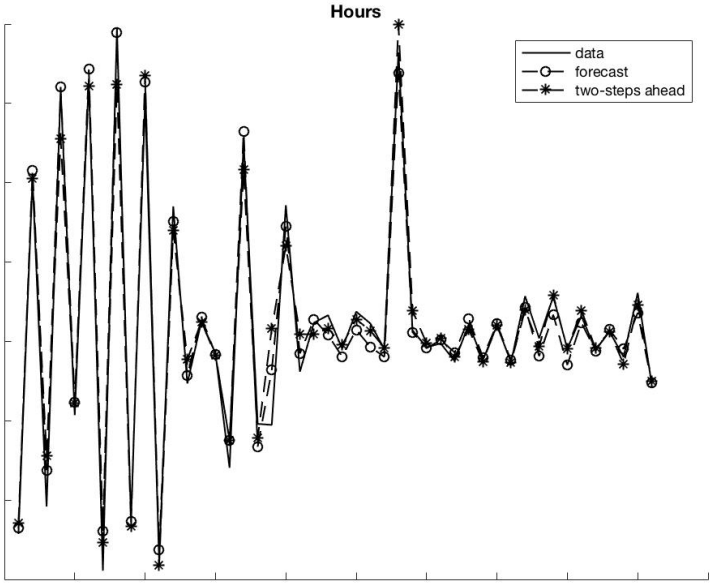
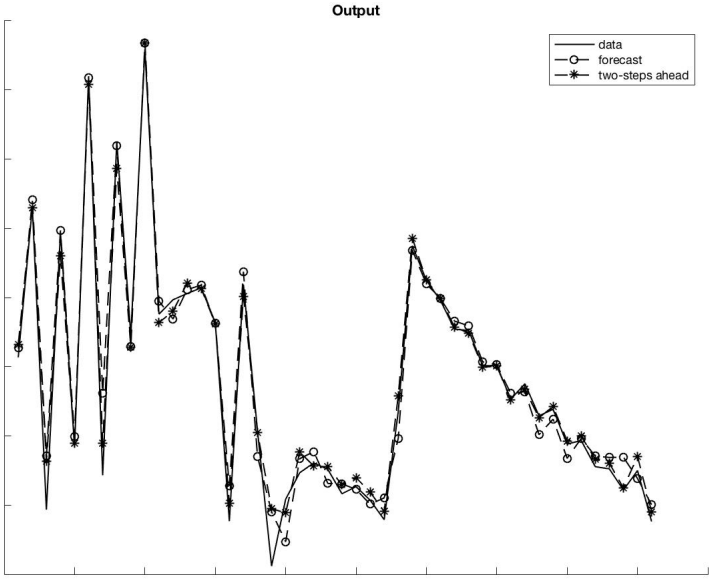


Figure 8: A series of 50 periods of simulated data

		Hours	Output	Consumption
One-step ahead	Mean	0.62	0.42	0.4
	Max	2.77	2.36	2.84
Two-steps ahead	Mean	0.78	0.29	0.68
	Max	2.43	1.66	3.65
One-step in KS	Mean	6.19	2.6	1.2
	Max	13.4	8.15	4.79

Table 4: Maximum and mean percentage differences between realized data and forecasts for the indicated variable. The third line reports, for the sake of comparison, one-step ahead forecast errors from the Krusell and Smith solution


Hours
data
forecast
two-steps ahead


Output
data
forecast
two-steps ahead

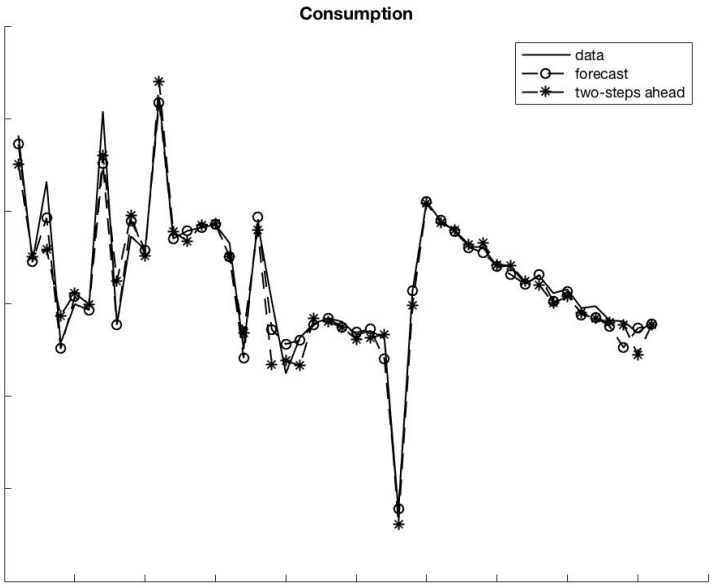


Figure 9: One-step vs. two-steps ahead forecasts

A formal testing procedure for numerical solutions of dynamic models has been advanced by Den Haan and Marcet (1994). They observe that any solution method implies that approximation error $e_t = \hat{\Gamma}_t(S_t) - \Gamma_t(S_t)$ should satisfy

$$\mathbb{E}_t \left[e_{t+1} \otimes h(S_t) \right] = 0 \tag{32}$$

for any function $h : \mathbb{R}^k \to \mathbb{R}^k$. It is then possible to construct a test statistic, $C = TB_T^{'} A_T^{-1} B_T$ such that $C \longrightarrow \chi_k^2$, where:

$$B_T \equiv \sum_{t=1}^{T} \frac{e_{t+1} \otimes h(x_T)}{T} \quad \text{and} \quad A_T \equiv \sum_{t=1}^{T} \frac{e_{t+1}^2 h(x_T) h^{'}(x_T)}{T} \tag{33}$$

It is known that any given numerical solution will fail the accuracy test for sufficiently large T. However, for reasonable number of simulated periods, the solution passes the test: critical values for a χ_k^2 with $k = 3$ are 0.115 at 1% and 0.352 at 5% .

T	200	250	500	1000
Test statistic	0.0257	0.0276	0.0317	0.0621

Table 5: Application of the test by Den Haan and Marcet (1994): the simulation passes the test at the 1% significance level for all lengths

6.2 Impulse Responses

The model economy used in this paper is constructed to analyze business cycle effects of uncertainty shocks; however, lacking a linear representation $X_t = AX_{t-1} + B\varepsilon_{\sigma,t}$ classical impulse response functions $\hat{x} = A^{t-1}B$ are not available. Following Koop et al. (1996), I adopt the following procedure to compute impulse responses:

1. Fix large N simulation samples of T_{IRF} periods, and fix shock period T_{shock}, then simulate the Markov chain of the aggregate shock for the T_{IRF} periods of each sample;

2. For each sample $i \in N$, draw $u_i \sim U(0,1)$; if $u_i > \bar{u}$, then label the series i as belonging to the "shock" group, and impose high uncertainty in T_{shock}, then let the process transition as usual[13]. Otherwise, simply label the series as "no shock";

3. Simulate the series, so to obtain X^{shock} and $X^{\text{no shock}}$

4. Obtain the Impulse Response Function as:

$$IRF = 100\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{X_{i,t}^{\text{shock}}}{X_{i,t}^{\text{no shock}}}\right) \tag{34}$$

The impulse responses are represented in Figure 11 for the most relevant variables of the model economy. They highlight an interesting and perhaps counter-intuitive effect of the interaction between endogenous firm entry and uncertainty shocks: constrained incumbent firms immediately respond by producing less output, as in the partial equilibrium. A general equilibrium effect, operating via decreasing wages, makes firms with a high idiosyncratic shock willing to use more labor, thus dampening the impact on aggregate hours. On impact, then, the effect of uncertainty operates primarily via a misallocation of labor: firms that are more constrained by default risk will reduce employment, while others will take advantage of falling wages and increase their production. As a result of the two effects, aggregate output is lower. However, the real option effect on (some) incumbent firms is not the only relevant dynamics induced by increased uncertainty: as future productivity can take higher as well as lower values, the value of more promising projects increases, and this implies an increase in firm entry. This increase more than compensates any increase in firm exit, and thus boosts output out of the small uncertainty-induced contraction. Longer periods of uncertainty can sustain longer (and deeper) contractions; however, the interpretation of such long periods of high uncertainty is less immediate.
Segal et al. (2015) discuss the impact of uncertainty shocks that are not mean-preserving,

[13]The number of periods after T_{shock} can change: in this paper, I model the shock as lasting two periods, so that in practice the economy starts its transition back to "normal" times in $T_{\text{shock}} + 2$

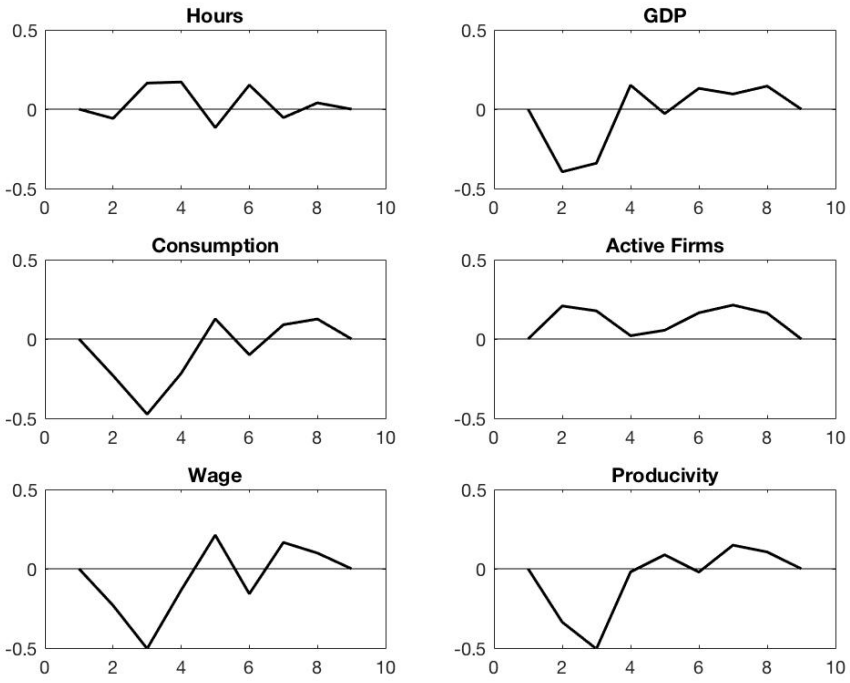


Figure 10: One standard deviation shock to uncertainty

allowing to identify and treat separately increased policy uncertainty (bad shocks) or increased uncertainty induced by a flow of innovation (good shocks). Such a framework can contribute with important insights for business cycle analysis - and it implies an extension of the presented model that would be simple to model under the solution method proposed in this paper.

7 Conclusions

In this paper I propose a global solution method for solving infinite-horizon, discrete time, heterogeneous agent models with aggregate uncertainty. I illustrate how it works by presenting a model economy which, in its relative simplicity, contains numerous computational challenges that make most existing methods inapplicable or unfeasible.
The approach is successful in solving the model with considerable accuracy; by treating distributional channels explicitly, it is able to analyze models where cross-sectional dynamics play a major role. The algorithm is designed for efficient implementation on HPC clusters, a feature that will become increasingly important in quantitative economics. Future research will be needed to deal with an even larger state space, where adaptive sparse grids become unfeasible. Another promising avenue is the combination of adaptive approaches with strategies that concentrate computational efforts on areas of the state space that are part of the ergodic set.

A Solving the Firm Problem using Howard Acceleration

The (intermediate) firm is defined by: past choices over labor and hours, l_t and b_t, and current level of idiosyncratic shock to demand, z_t. These define the idiosyncratic state x_t. The economy is characterized by the aggregate hours and debt, N_t and B_t, and the shock to volatility σ_t. These define the aggregate state as S_t. Moreover, aggregate production Y_t is obtained via Dixit-Stiglitz aggregation of each firm's production $y_t = l_t^{\alpha}$
The firm problem for the choice of hours and debt is then:

$$V(x_t, S_t) = \max_{l,b} \left\{ \kappa d_t + \beta(1-\kappa)\Sigma_{z_{t+1}} \pi(z_{t+1}|z_t) Q(\sigma_{t+1}|\sigma_t) V(x_{t+1}, S_{t+1}) \right\}$$
$$\text{sub } d_t = p_t(x_t) l_t^{\alpha} - w_t l_t - b_t + q_t(x_t, S_t) b_{t+1}$$

As an additional constraint, dividends are restricted to be non-negative: when a firm can do nothing but distribute negative dividends, it defaults. The firm takes all aggregates in S_t and wages w_t as given. The price of the intermediate good is

$$p_t = z_t \left(\frac{Y_t}{y_t}\right)^{\frac{1}{\gamma}}$$

whereas intermediaries fix the price on the bond q_t in a way such that $q_t(x_t, S_t) b_{t+1}$ equals the expected payment.

1. **Starting values**
   Since S_t is exogenous for the firm, values of Y_t, N_t, as well as the law of motion that determines $\hat{C}_{t+1} = E_t(C_{t+1}|\mathscr{F}_t)$, $\hat{Y}_{t+1} = E_t(Y_{t+1}|\mathscr{F}_t)$ and $\hat{N}_{t+1} = E_t(N_{t+1}|\mathscr{F}_t)$, where $\mathscr{F}_t$ is the information set of the firm at time t, are determined outside of this problem.
   Given these values, one chooses a $n \times n$ grid in choice variables l_{t+1}, b_{t+1}, as well as a starting guess g_0 for the policy function $(l^*_{t+1}, n^*_{t+1}) = g_0(x_t, S_t)$.

2. **Obtain V from g**
   Consider that, if $g_0(x_t, S_t)$ is optimal, then

$$V_{i+1} = k d_t(g_0(x_t)) + \beta(1-\kappa)\Sigma_{z_{t+1}} \pi(z_{t+1}|z_t) Q(\sigma_{t+1}|\sigma_t) V_i \tag{35}$$

   with $V_i = V_{i+1}$. Then (35) can be rearranged as

$$V = (I - \beta(1-\kappa) Q(\sigma_{t+1}|\sigma_t) T)^{-1} \kappa d_t(g_0(x_t))$$

   where

$$T = \begin{pmatrix} \pi_{1,1}J_1 & \pi_{1,2}J_1 & \cdots & \pi_{1,5}J_1 \\ \pi_{2,1}J_2 & \pi_{2,2}J_2 & \cdots & \pi_{2,5}J_2 \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{5,1}J_5 & \pi_{5,2}J_5 & \cdots & \pi_{5,5}J_5 \end{pmatrix}$$

and J_i, i=1,...,5 is a $n^2 \times n^2$ sparse matrix whose j,k-th entry equals 1 if $\text{pos}(l_t, b_t) = j$, $\text{pos}(g(x_t)) = k$ [14].

3. **Obtain the new policy function**
   With V in hand, one can compute the expected continuation value $C(g_i(x_t)) = E\left[V(g_i(x_t), S_{t+1})\right]$, so to get the objective function as

$$OBJ(g_i(x_t)) = \kappa d_t + \beta(1-\kappa)C(g_i(x_t)) \tag{36}$$

   Clearly, $g_{i+1}(x_t) = \text{argmax}_{l_{t+1}, b_{t+1}} OBJ(g_i(x_t))$.

4. **Check convergence**
   Obtain V_{i+1} as in **1**. Stop if $\left\| \frac{g_{i+1} - g_i}{g_i} \right\| <$ some criterion, otherwise go back to **2**.

B Applying Krusell and Smith Solution Method

First, choose a functional form for calculating conditional expectations. This implies a forecasting system:

$$log(\hat{X}_{t+1}) = \alpha\delta_\sigma + \beta log(X_t) + \gamma\Upsilon_t \tag{37}$$

where $X_t = [N_t, Y_t, C_t]$, and δ_σ is an aggregate state dummy, which can take on four values, depending on present and future uncertainty state. Guess coefficients $\Gamma_0 = [\alpha, \beta, \gamma]$, draw a T- period long time series for idiosyncratic and aggregate shocks from the discretized Markov Chain of $\{z_t\}_{t=0}^T$ and $\{\sigma_t\}_{t=0}^T$. Choose a tolerance value ε.

In each iteration s, and period t:

1. Given Γ_{s-1}, and existing aggregate state S_t, guess a consumption level $\tilde{C}$. For aggregate state S_t and law of motion Γ_{s-1}, solve for individual policy functions $n_{t+1}(x_t, S_t)$, $b_{t+1}(x_t, S_t)$ and obtain the expected value of entrants: with this and the free-entry condition $V^e(x_t, S_t, \xi_t) > 0$, one obtains $\int \xi d\mu I_t$;

2. Update $\tilde{C}$ using bisection;

[14] pos refers to the correspondent entry in the (l_t, b_t) grid

Table 6: Krusell and Smith Regression

	Hours		Output		Consumption	
Constant*	-1.373		-1.1912		-1.2416	
Hours*	0.1029		-0.1030		-0.248	
Output*	-0.779		0.0181		0.158	
Consumption*	-0.915		-0.4421		-0.3861	
		-1.386		-1.1125		-1.1663
		-0.232		-0.3496		-0.465
		-0.2806		0.3438		0.6433
		-0.8982		-0.503		-0.4898
Aggregate Shock Dummy	-0.0009		0.0071		-0.0083	
Utilization	0.1006		0.2857		0.4048	
R^2	0.999		0.989		0.999	

* = Left Column is for $\sigma = \sigma_L$, Right Column for $\sigma = \sigma_H$

3. If guessed $\tilde{C}$ equals consumption implied by $C = Y - I$, market clears. Take next period's idiosyncratic shocks to obtain aggregate variables S_{t+1};

4. After having obtained a full time series from the T iterations between 1 and 3, update the forecasting system using least squares, and obtain Γ_s. If $|\Gamma_s - \Gamma_{s-1}| < \varepsilon$, for some tolerance value ε, the algorithm stops; otherwise, go back to point 1 with the new forecasting system at hand;

C Results from Krusell and Smith solution

The following table reports estimation results from the application of the Krusell and Smith algorithm to the model economy of the paper.

D Building Adaptive Sparse Grids

Be $\Omega = [0,1]^d$ the domain of the problem, where d is its dimensionality[15]. Let $\vec{x}_{\vec{l},\vec{i}}$ be a d-dimensional gridpoint in our discretized state space: $\vec{l} = (l_1, ..., l_d) \in \mathbb{N}^d$ are the indices representing the refinement level and $\vec{i} = (i_1, ..., i_d) \in \mathbb{N}^d$ represents the spatial position of the gridpoint. A full grid Ω_l on Ω will then have mesh size defined as:

$$h_{\vec{l}} = (h_{l_1}, ..., h_{l_d}) = 2^{-\vec{l}} \tag{38}$$

[15]The domain can clearly be adapted to the nature of the problem: in applications it is a common choice to tailor it around the ergodic set

The grid is such that, along each dimension, gridpoints are equidistant. However, the mesh size h_l may differ along dimensions.
Let us now assume we want to interpolate a function $f : \Omega \to \mathbb{R}$; for the sake of simplicity we will illustrate the one-dimensional case, as the multimensional extension is straightforward[16]. The set of points on the grid is characterized by a hierarchy, and at each refinement level we can associate functions of arbitrary order p:

$$\varphi_{l,j}^{(p)}(x) = \begin{cases} \Pi_{i \in F_l} \frac{x - x_i}{x_j - x_i}, & \text{if } x \in \left\{ x_{l,j} - h_l, x_{l,j} + h_l \right\} \\ 0, & \text{otherwise} \end{cases}$$

Notice that a function can have order p only if there are l level of *ancestor* points of x_j; otherwise, the largest possible p is taken (eventually, the function will be linear). To order basis functions, define increment spaces as

$$W_l := \text{span} \left\{ \varphi_{l,i:i \in I_l}^{(p)} \right\}, \; I_l = \left\{ i \in \mathbb{N}, 1 \leq i \leq 2^l - 1, i \text{ odd} \right\} \tag{39}$$

This implies that different basis functions have mutually disjoint support within a level and that the support of each function on level l is the support of two functions on the next refinement level $l + 1$. A function f is then approximated as

$$f(x) \approx \sum_{k=1}^{l} \sum_{i \in I_l} \alpha_{k,i} \varphi_{k,i}(x) \tag{40}$$

The coefficients $\alpha_{k,i}$ are called *hierarchical surpluses*. To understand how they are determined and their crucial role in the determination of grids, we will introduce an analytical example. In Figure 11, a univariate normal distribution is displayed.

[16] Stoyanov (2015) provides with a careful explanation on how to extend this reasoning to a multidimensional environment

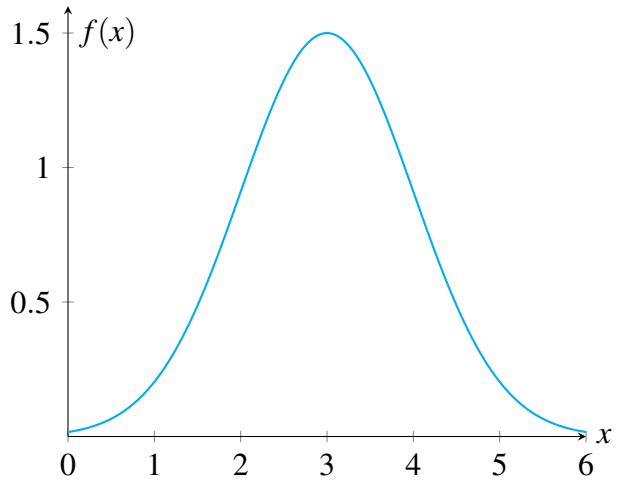


Figure 11: Normal distribution with $\mu = 2$, $\sigma^2 = 2$

To approximate it, we first may want to rescale Ω so that at least it covers the interval between 0 and 6. The first level approximation involves hierarchical function to be $\varphi_{1,1}(3) = 1$; to calculate the surplus, simply notice that $\alpha_{1,1} = f(x_1) - \varphi_{1,1}(3)$. This is highlighted in Figure 12. At this point, an evaluation should be made: if $\alpha_{1,1} > \varepsilon$, where ε is some tolerance value, two *children* points are added - and by the reasoning made above they will be 1.5 and 4.5.

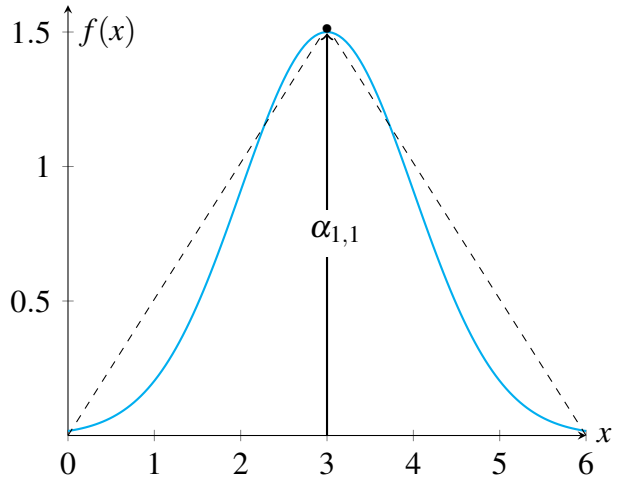


Figure 12: Level 1 hierarchical function and coefficients

In 13, the two points are added, and the relative surpluses are calculated, as above, from the difference between the approximate function at the previous level and the function evaluated in the two points. The resulting approximation for $l = 2$ is given

by the sum of the black dashed line and the red dashed line, which is precisely $\sum_{k=1}^{l} \sum_{i \in I_l} \alpha_{k,i} \varphi_{k,i}(x)$

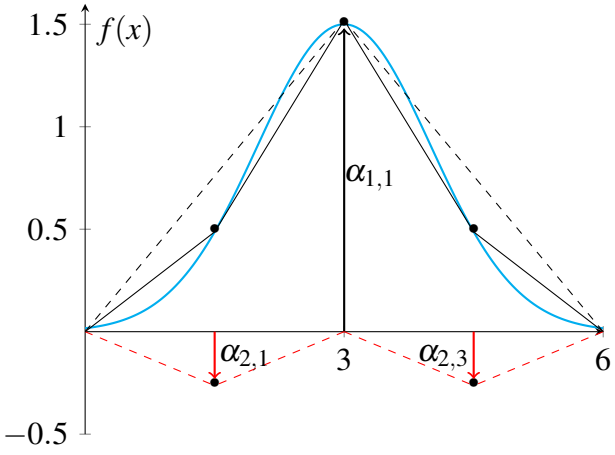


Figure 13: Level 2 hierarchical function and coefficients

Notice how the process can go on until it reaches arbitrary accuracy, limited by the tolerance value ε, by adding points in each level in the area of the domain where the function exhibits more curvature. As we will see, this will allow us to deal with the aggregate state space of the model, by approximating agents' expectations with high accuracy without spending too much time in areas of the state space where the mapping is more linear. For a discussion on properties of adaptive sparse grids in approximating functions that are known only algorithmically, see Scheidegger and Treccani (2016).

References

Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot, S. (2011), 'Dynare: Reference manual, version 4'.

Aiyagari, S. R. (1994), 'Uninsured idiosyncratic risk and aggregate saving', *The Quarterly Journal of Economics* **109**(3), 659–684.

Algan, Y., Allais, O. and Den Haan, W. J. (2008), 'Solving heterogeneous-agent models with parameterized cross-sectional distributions', *Journal of Economic Dynamics and Control* **32**(3), 875.

Algan, Y., Allais, O., Den Haan, W. J. and Rendahl, P. (2013), 'Solving and simulating models with heterogeneous agents and aggregate uncertainty', *Handbook of Computational Economics* **3**, 277.

Arellano, C., Bai, Y. and Kehoe, P. J. (2016), Financial frictions and fluctuations in volatility, Technical report, National Bureau of Economic Research.

Arellano, C., Bai, Y. and Kehoe, P. J. (2019), 'Financial frictions and fluctuations in volatility', *Journal of Political Economy* **127**(5), 2049–2103.

Bellman, R. E. (1961), *Adaptive control processes: a guided tour*, Princeton university press.

Bloom, N. (2009), 'The impact of uncertainty shocks', *Econometrica: Journal of the Econometric Society* **77**(3), 623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. and Terry, S. J. (2013), 'Really uncertain business cycles'.

Boppart, T., Krusell, P. and Mitman, K. (2018), 'Exploiting mit shocks in heterogeneous-agent economies: the impulse response as a numerical derivative', *Journal of Economic Dynamics and Control* **89**, 68–92.

Brumm, J., Kryczka, D. and Kubler, F. (2017), 'Recursive equilibria in dynamic economies with stochastic production', *Econometrica* **85**(5), 1467–1499.

Brumm, J. and Scheidegger, S. (2017), 'Using adaptive sparse grids to solve high-dimensional dynamic models', *Econometrica* **85**(5), 1575–1612.

Bungartz, H.-J. and Dirnstorfer, S. (2003), 'Multivariate quadrature on adaptive sparse grids', *Computing* **71**(1), 89–114.

Bungartz, H.-J. and Griebel, M. (2004), 'Sparse grids', *Acta numerica* **13**, 147–269.

Cheridito, P. and Sagredo, J. (2016), 'Existence of sequential competitive equilibrium in krusell–smith type economies'.

Dalcín, L., Paz, R. and Storti, M. (2005), 'Mpi for python', *Journal of Parallel and Distributed Computing* **65**(9), 1108–1115.

Dalcín, L., Paz, R., Storti, M. and D'Elía, J. (2008), 'Mpi for python: Performance improvements and mpi-2 extensions', *Journal of Parallel and Distributed Computing* **68**(5), 655–662.

Davis, S. J., Faberman, R. J. and Haltiwanger, J. C. (2012), 'Recruiting intensity during and after the great recession: National and industry evidence', *The American Economic Review* **102**(3), 584–588.

Den Haan, W. J. (2010), 'Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents', *Journal of Economic Dynamics and Control* **34**(1), 79–99.

Den Haan, W. J. and Marcet, A. (1990), 'Solving the stochastic growth model by parameterizing expectations', *Journal of Business & Economic Statistics* **8**(1), 31–34.

Den Haan, W. J. and Marcet, A. (1994), 'Accuracy in simulations', *The Review of Economic Studies* **61**(1), 3–17.

Den Haan, W. J. and Rendahl, P. (2010), 'Solving the incomplete markets model with aggregate uncertainty using explicit aggregation', *Journal of Economic Dynamics and Control* **34**(1), 69–78.

Dongarra, J. J. and van der Steen, A. J. (2012), 'High-performance computing systems: Status and outlook', *Acta Numerica* **21**, 379–474.

Eaton, J. and Gersovitz, M. (1981), 'Debt with potential repudiation: Theoretical and empirical analysis', *The Review of Economic Studies* **48**(2), 289–309.

Fort, T. C., Haltiwanger, J., Jarmin, R. S. and Miranda, J. (2013), 'How firms respond to business cycles: The role of firm age and firm size', *IMF Economic Review* **61**(3), 520–559.

Gropp, W., Lusk, E. and Skjellum, A. (1999), *Using MPI: portable parallel programming with the message-passing interface*, Vol. 1, MIT press.

Guvenen, F. (2011), 'Macroeconomics with hetereogeneity: a practical guide', *Economic Quarterly* (3Q), 255–326.

Haltiwanger, J. (2012), 'Job creation and firm dynamics in the united states', *Innovation policy and the economy* **12**(1), 17–38.

Jensen, M. C. (1986), 'Agency costs of free cash flow, corporate finance, and takeovers', *The American economic review* **76**(2), 323–329.

Jermann, U. and Quadrini, V. (2012), 'Macroeconomic effects of financial shocks', *The American Economic Review* **102**(1), 238–271.

Judd, K. L. (1998), *Numerical methods in economics*, MIT press.

Judd, K. L., Maliar, L. and Maliar, S. (2011), 'Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models', *Quantitative Economics* **2**(2), 173–210.

Khan, A. and Thomas, J. K. (2008), 'Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics', *Econometrica: Journal of the Econometric Society* **76**(2), 395–436.

Koop, G., Pesaran, M. H. and Potter, S. M. (1996), 'Impulse response analysis in nonlinear multivariate models', *Journal of econometrics* **74**(1), 119–147.

Krueger, D. and Kubler, F. (2004), 'Computing equilibrium in olg models with stochastic production', *Journal of Economic Dynamics and Control* **28**(7), 1411–1436.

Krueger, D., Mitman, K. and Perri, F. (2016), 'Macroeconomics and household heterogeneity', *Handbook of Macroeconomics* **2**, 843–921.

Krusell, P. and Smith, A. (1998), 'Income and wealth heterogeneity in the macroeconomy', *Journal of Political Economy* **106**(5), 867–896.

Kydland, F. E. and Prescott, E. C. (1982), 'Time to build and aggregate fluctuations', *Econometrica: Journal of the Econometric Society* pp. 1345–1370.

Ljungqvist, L. and Sargent, T. J. (2012), *Recursive macroeconomic theory*, MIT press.

Ma, X. and Zabaras, N. (2010), 'An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations', *Journal of Computational Physics* **229**(10), 3884–3915.

Miao, J. (2006), 'Competitive equilibria of economies with a continuum of consumers and aggregate shocks', *Journal of Economic Theory* **128**(1), 274–298.

Myers, S. C. (1977), 'Determinants of corporate borrowing', *Journal of Financial Economics* **5**(2), 147–175.

Pflüger, D. (2012), Spatially adaptive refinement, *in* 'Sparse grids and applications', Springer, pp. 243–262.

Pröhl, E. (2017), 'Computing the cross-sectional distribution to approximate stationary markov equilibria with heterogeneous agents and incomplete markets', *Working Paper* .

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

Ríos-Rull, J.-V. (1997), 'Computation of equilibria in heterogeneous agent models', *Federal Reserve Bank of Minneapolis Staff Report* .

Scheidegger, S. and Bilionis, I. (2017), 'Machine learning for high-dimensional dynamic stochastic economies'.

Scheidegger, S. and Treccani, A. (2016), 'Pricing american options under high-dimensional models with recursive adaptive sparse expectations'.

Segal, G., Shaliastovich, I. and Yaron, A. (2015), 'Good and bad uncertainty: Macroeconomic and financial market implications', *Journal of Financial Economics* **117**(2), 369–397.

Stanfield, B., Haltiwanger, J. and Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', *The American economic review* **98**(1), 394–425.

Stoyanov, M. (2015), User manual: Tasmanian sparse grids, Technical Report ORNL/TM-2015/596, Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN.

Tauchen, G. (1986), 'Finite state markov-chain approximations to univariate and vector autoregressions', *Economics letters* **20**(2), 177–181.

Terry, S. (2015), 'Alternative methods for solving heterogeneous firm models'.

Winberry, T., Kaplan, G., Moll, B., Ahn, S. et al. (2016), No more excuses! a toolbox for solving heterogeneous agent models with aggregate shocks, *in* '2016 Meeting Papers', number 1335, Society for Economic Dynamics.

Young, E. R. (2010), 'Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations', *Journal of Economic Dynamics and Control* **34**(1), 36–41.