serie ad
WP-AD 2011-21

A reinforcement learning approach to solving incomplete market models with aggregate uncertainty

Andrei Jirnyi and Vadym Lepetyuk

Edita / Published by: Instituto Valenciano de Investigaciones Económicas, S.A.

Depósito Legal / Legal Deposit no.: V-3399-2011

Impreso en España (septiembre 2011) / Printed in Spain (September 2011)

WP-AD 2011-21

A reinforcement learning approach to solving incomplete market models with aggregate uncertainty*

Andrei Jirnyi and Vadym Lepetyuk**

Abstract

We develop a method of solving heterogeneous agent models in which individual decisions depend on the entire cross-sectional distribution of individual state variables, such as incomplete market models with liquidity constraints. Our method is based on the principle of reinforcement learning, and does not require parametric assumptions on either the agents' information set, or on the functional form of the aggregate dynamics.

Keywords: Heterogeneous agents, macroeconomics, dynamic programming, reinforcement learning.

JEL codes: C63, C68, E20.

* Financial support from the Spanish Ministerio de Educación y Ciencia and REDEF funds under project SEJ2007-62656 is gratefully acknowledged.

** A. Jirnyi, Kellogg School of Management, Northwestern University: a-jirnyi@northwestern.edu. V. Lepetyuk, Universidad de Alicante: lepetyuk@merlin.fae.ua.es.

1 Introduction

Heterogeneous agent models with market incompleteness have been playing an increasing role in macroeconomics and finance (den Haan, Judd, and Juillard, 2010), as the assumptions behind the classic representative-agent models are too stringent to be realistic in a number of circumstances (Kirman, 1992). Explicit accounting for heterogeneity is also necessary when considering important questions concerned with differential effects of economic fluctuations, and redistributive effects of policies: as Kocherlakota (2010) notes, the models in which "the distribution of financial wealth evolves over time ... lead to a better understanding of the cost of economic downturns". However, finding equilibria of such models is often a challenge, since their state space includes cross-sectional distributions, which are typically objects of very high dimension. With most of the existing methods requiring computational time that is growing exponentially in the number of state variables, this leads to the so called "curse of dimensionality".

In this paper we propose a simulation-based nonparametric method of solving heterogeneous agent models with aggregate uncertainty. In our method agents make their decisions based on a fully-specified high-dimensional cross-sectional state distribution, and do not adopt any restrictive assumptions on its' transition law, such as separability or a specific parametric form.

This flexibility differentiates our method from others proposed in the literature, that has largely been following the approach of den Haan (1996) and Krusell and Smith (1998), which is based on restricting the agents' decision-making to a small number (often one) of aggregate statistics from the entire cross-sectional distribution, and on adopting additional assumptions (e.g. linearity) on their transition law. Our algorithm is comparable in simplicity of implementation and computational time to that of Krusell and Smith (1998). In addition, it allows to overcome naturally a number of limitations faced by the other algorithms, such as difficulties in accounting for explicit inter-dependencies between the cross-sectional distribution and individual decisions.

Our proposed approach is motivated by the substantial recent advances that have been made in the fields of machine learning and operations research, leading to a development of a class of "reinforcement learning," or "approximate dynamic programming" algorithms, which have been used to compute approximate solutions to previously intractable large dynamic optimization problems with thousands, and sometimes hundreds of thousands, of state variables (Powell, 2007). The problems that have been addressed by these methods range from large-scale industrial logistics (Simão, Day, George, Gifford, Nienow, and Powell, 2009) to optimal investment (Nascimento and Powell, 2010) to the game of backgammon

(Tesauro, 1994), but the main underlying common feature of the methods in this class is that they rely on stochastic simulations in order to both approximate the expectation of the objective in different future states of the world, and to prioritize the most important states for further analysis.

Our method consists of a combination of a nonparametric k-nearest-neighbor (k-nn) regression with stochastic simulations. On each iteration of the algorithm, we simulate a realization of aggregate uncertainty, and estimate the expected continuation value using k closest historical simulated realizations of the cross-sectional distribution in the functional space. Equipped with the estimated continuation value, we solve the optimization problem.

We illustrate the method for the classical Krusell and Smith (1998) economy. In the model, the agents face both idiosyncratic employment shocks and aggregate productivity shocks, and can choose to save only into capital, subject to a no-borrowing constraint. The recursive formulation of the agent decision problem includes the cross-sectional distribution of capital stock, which in our implementation is described by 1,000 continuously-valued state variables.

2 A heterogeneous agent model

In this section, we describe the environment of Krusell and Smith (1998) with the unemployment insurance as in den Haan, Judd, and Juillard (2010), and we define the recursive competitive equilibrium.

We select this environment as our example because it is a classic, well-studied and well-understood model. In addition, it has received extensive attention in a recent (January, 2010) issue of the Journal of Economic Dynamics and Control (see den Haan, Judd, and Juillard, 2010), where several competing methods for its' solution have been presented and compared in detail (den Haan, 2010b).

One limitation of this comparison is that, as Krusell and Smith (2006) argue, in this particular setup the agents decisions *are*, in fact, primarily driven by the mean of wealth, whose dynamics *is* close to linear, and thus the methods that explicitly make such an assumption are applicable and show good performance. Since in this paper we solve the model without making these assumptions, our results can serve as an independent confirmation of the validity of the Krusell and Smith's conjecture.

2.1 The model

The baseline model is a modified version of Krusell and Smith (1998), as described by den Haan, Judd, and Juillard (2010), whose notation we largely adopt.

There is a measure-one continuum of ex-ante identical consumers, with the preferences over the stream of consumption given by the following utility function:

$$\mathrm{E}\sum_{t=0}^{\infty}\beta^t\frac{c_{it}^{1-\gamma}-1}{1-\gamma} \tag{1}$$

Agents face two sources of uncertainty: aggregate shock to productivity a_t and individual shock to employment e_t, where $e_t = 1$ if the agent is employed and zero otherwise. Employed agents inelastically supply l units of labor on which they earn a wage w_t. The employed agents pay a labor tax at rate τ_t, while unemployed agents receive a subsidy μw_t. The agents cannot pool away the employment risk by trading any contingent bonds, thus the markets are incomplete. The agents can only save nonnegative amounts k_{it} by investing in capital, earning the net return $r_t - \delta$, where r_t is the rental price of capital and δ is the rate of capital depreciation.

The consumption good is produced by competitive firms having Cobb-Douglas production function. The aggregate output is therefore equal to the following

$$Y_t = a_t K_t^{\alpha}(lL_t)^{1-\alpha} \tag{2}$$

where K_t is the aggregate capital, L_t is the employment rate, and lL_t is the aggregate employment.

The government pays unemployment benefits, paid for by taxing the employed. It balances its budget every period, implying a tax rate $\tau_t = \mu(1 - L_t)/(lL_t)$.

We consider a recursive competitive equilibrium, which includes a law of motion of the aggregate state of the economy. Denoting the density of cross-sectional distribution over capital and employment as λ, the aggregate state of the economy is (λ, a). The individual state (k, e, a, λ) consists of the individual holdings of capital, the employment status of the agent, and the aggregate state.

The individual maximization problem in the recursive form is

$$V(k, e, a, \lambda) = \max_{c,k'}\{u(c) - \beta\mathrm{E}\{V(k', e', a', \lambda')|e, a, \lambda\}\} \tag{3}$$

subject to the budget constraint

$$c + k' = r(K, L, a)k + \left[(1 - \tau(L))le + \mu(1 - e)\right] w(K, L, a) + (1 - \delta)k \tag{4}$$

the nonnegativity constraint on capital holdings

$$k' \geq 0 \tag{5}$$

and the transition law of λ

$$\lambda' = H(\lambda, a, a') \tag{6}$$

The policy function for the next period capital is denoted by function f as $k' = f(k, e, a, \lambda)$.

Wages and prices in this economy are competitive and given by

$$w(K, L, a) = (1 - \alpha)a \left(\frac{K}{lL}\right)^{\alpha} \tag{7}$$

$$r(K, L, a) = \alpha a \left(\frac{K}{lL}\right)^{\alpha - 1} \tag{8}$$

where the market clearing conditions require $K = \int_0^1 k_i \lambda_i di$ and $L = \int_0^1 e_i \lambda_i di$.

A recursive competitive equilibrium is the aggregate law of motion H, a pair of individual functions v and f, and the pricing system (r, w) such that (i) (v, f) solve the consumer problem, (ii) w and r are competitive, and (iii) the aggregate law of motion H is consistent with the individual policy function f.

2.2 Exogenous driving process

There are two types of shocks in the model, aggregate and individual. The exogenous shock to aggregate productivity a_t is a two-state Markov process, $a_t \in \{a^b, a^g\} \equiv \{1-\Delta_a, 1+\Delta_a\}$, with transition probability $\mathrm{P}\{a_{t+1} = a' | a_t = a\} \equiv \pi(a'|a)$. The individual shock to employment status is also a Markov process, conditional on the realization of the aggregate shock, with transition probabilities given by $\mathrm{P}\{e_{i,t+1} = e' | e_{it} = e, a_{t+1} = a'\} \equiv \pi(e'|e, a')$. The joint transition probability $\mathrm{P}\{a_{t+1} = a', e_{i,t+1} = e' | a_t = a, e_t = e\}$ is denoted as $\pi(a', e'|a, e)$, and is chosen so that the aggregate employment is only a function of the aggregate state of the economy[1].

[1]While this assumption serves to simplify application of other solution methods, it is not necessary for our approach. We retain it, as well as the rest of the calibration in den Haan, Judd, and Juillard (2010), for comparison purposes.

3 Solving the model

Finding the equilibrium of the model is complicated, since a distribution function (generally, an infinite-dimensional object) enters the decision problem as a state, leading to the so called "curse of dimensionality". There are two primary aspects to this "curse". First, accurate representation of λ itself requires a large number of state variables. Second, the transition function $\lambda' = H(\lambda, a, a')$ is an unknown, potentially nonlinear and nonseparable, high-dimensional function of a high-dimensional argument, which would present complications even if it were possible to represent λ with a relatively small state vector.

An innovative algorithm has been suggested for such problems by Krusell and Smith (1998). It relies upon making two assumptions. First, Krusell and Smith assume limited rationality on part of the agents. The agents only consider a small number (commonly one) of summary statistics of the distribution λ in their decisions. Second, the aggregate law of motion for these statistics, as perceived by the agents, is restricted to a simple parametric (e.g. linear) functional form. Among the functions of this form, the authors use stochastic simulation to find a "self-confirming" rule, i.e. such a rule that, when taken by agents as given, results in simulated dynamics for the statistics of interest that are close to those implied by the rule within a given tolerance.

In addition to the classic algorithm of Krusell and Smith (1998), a number of alternative techniques have also been proposed in the literature. For instance, "projection methods" still rely on parametrization, but avoid the simulation step by embedding the aggregation of the individual policy functions into the individual problem explicitly. The cross-sectional distribution is either parametrized independently from the individual decision rules (Algan, Allais, and den Haan, 2010, Reiter, 2010), or follows from the parametrization of the individual policy functions (den Haan and Rendahl, 2010). On the other hand, "perturbation methods" (Kim, Kollmann, and Kim, 2010) are based on approximation of the individual policy functions and the aggregate law of motion around the steady state. These methods, however, are most suitable in environments where individual policies can be easily approximated by a few terms in a functional expansion, and thus face a difficulty with models with occasionally binding borrowing constraints, since the policy functions in such models are not differentiable.

There are two related caveats with respect to the algorithm of Krusell and Smith (1998). First, since it is by construction a limited-rationality solution, where agents are constrained in both their information set and decision-making, it does not easily allow to check how restrictive these constraints are, and how much of an impact they have on the solution. While it is possible to partially address this issue by including additional aggregate state

variables, such as higher-order cross-sectional moments, such an expansion is limited due to the second problem: it would also require additional parameters in the transition function, and estimating these parameters in a robust manner can be quite difficult, especially when non-linear interactions between the states are allowed. Moreover, due to the "curse of dimensionality", the computational cost of solving the individual problem grows exponentially with each additional state variable. These issues become even more challenging when the aggregate distribution is multidimensional.

In contrast, our proposed method does not require these assumptions (although, it still allows to make use of them, if warranted by the theory). It is based on the "reinforcement learning" approach[2] to solving high-dimensional dynamic optimization problems. As noted by Powell (2007), there are two critical components that any stochastic algorithm facing the curse of dimensionality requires in order to be effective: (i) a way to infer approximate objective values in the states (e.g. realizations of the cross-sectional distribution) that have not yet been investigated from those that are already known, and (ii) a way to focus attention on the more likely states of the world (the so-called "ergodic set"). Importance of focusing the procedure on the ergodic set has been recently highlighted by Maliar, Maliar, and Judd (2010). In our example, we combine both approaches to find solution to a model with a distribution that is at all times fully described by a 1000-dimensional state vector.

The basic intuition for the proposed method is that it splits the value function the decision-maker maximizes into the current utility and the conditional expectation of the continuation value, and uses stochastic simulation to approximate the latter.

3.1 Continuation values and their estimation

We can rewrite the individual maximization problem (3) as

$$V(k, e, a, \lambda) = \max_{c,k'} \{u(c) + \beta\psi(k', e, a, \lambda)\} \tag{9}$$

subject to the conditions (4)-(8), where ψ is the continuation value of picking capital k', conditional on current shocks (e, a) and the current cross-sectional distribution λ:

$$\psi(k', e, a, \lambda) = \mathrm{E}_{e',a'|e,a} \left\{V\left(k', e', a', H(\lambda, a, a')\right) | e, a\right\} \tag{10}$$

Note that ψ is a scalar-valued, non-stochastic function that encompasses both the transitional dynamics $H(\lambda)$ and the dependency between λ and V, and maximization (9) no

[2]Sometimes also referred to as "approximate", "asynchronous", or "adaptive" dynamic programming; see Sutton and Barto (1998) for an excellent introduction.

longer involves computing an explicit expectation. Nevertheless, it requires finding ψ, which is a function of λ and therefore a high-dimensional object.

If there were no heterogeneity and no aggregate uncertainty in the model, ψ could be found by value function iteration: first, given a value of $V^{(j)}$ in an iteration j, find $\psi^{(j)}$ by evaluating the expectation in (10). Next, compute the value $V^{(j+1)}$ at the next iteration by maximizing (9), and repeat the procedure until convergence. In this way, the classic value function iteration algorithm can be visualized as proceeding backward through time, with iteration-$(j+1)$ value function computed as the previous-period expectation of the iteration-(j) value.

Aggregate uncertainty and distribution-dependency greatly complicate things. First, the expectation in (10) can no longer be evaluated directly, since the exact form of H is not generally available. Second, even if it were available, evaluating (9) and (10) would require to cover all possible values of the distribution λ, leading to the "curse of dimensionality" issue.

Instead of solving for the continuation value ψ explicitly, we propose using stochastic simulation to approximate ψ with a sequence of easy to compute random functions $\hat{\psi}_t$, such that $\hat{\psi}_t \to \psi$ as $t \to \infty$.

Imagine that at time t we know the true continuation value $\psi(\cdot, \cdot, a, \lambda)$ in N points $\left\{(\tilde{a}_\tau, \tilde{\lambda}_\tau)\right\}_{\tau=1}^{N}$, i.e. we have a set of triplets $\Psi = \left\{\tilde{a}_\tau, \tilde{\lambda}_\tau, \psi_\tau(\cdot, \cdot, \tilde{a}_\tau, \tilde{\lambda}_\tau)\right\}_{\tau=1}^{N}$, and observe a_t and λ_t. Finding an approximation $\hat{\psi}_t(\cdot, \cdot, a_t, \lambda_t)$ can then be interpreted as a problem of statistical estimation, which can be addressed nonparametrically.

One method of such estimation is a k-nearest-neighbor regression, which is the most popular nonparametric method that dates back to Fix and Hodges (1951). It involves first finding M nearest realizations, i.e. such $(\tau_1, ..., \tau_M)$ that $\tilde{a}_{\tau_1} = ... = \tilde{a}_{\tau_M} = a_t$ and $d(\lambda_t, \tilde{\lambda}_{\tau_1}) \leq ... \leq d(\lambda_t, \tilde{\lambda}_{\tau_M}) \leq d(\lambda_t, \tilde{\lambda}_\tau)\ \forall \tau \notin \{\tau_1, ..., \tau_M\}$, where $d(\lambda_t, \tilde{\lambda}_\tau)$ is some distance metric between two distributions λ_t and $\tilde{\lambda}_\tau$. Then, the k-nn estimator of the continuation value ψ is computed as a simple average:

$$\hat{\psi}_t(k', e, a_t, \lambda_t) = \frac{1}{M} \sum_{j=1}^{M} \psi_{\tau_j}(k', e, \tilde{a}_{\tau_j}, \tilde{\lambda}_{\tau_j})\ \forall (k', e) \tag{11}$$

Since the k-nn regression estimator in a separable metric space is asymptotically consistent (Cover and Hart, 1967), $\hat{\psi}_t$ converges to ψ with probability one as the sample size N increases and thus allows to approximate functions of arbitrary complexity.

Consider now a sample realization of this economy, driven by a shock sequence $\{\tilde{a}_\tau\}$, with a corresponding sample path of cross-sectional distributions $\{\tilde{\lambda}_\tau\}$, observed up to time $(t-1)$.

Given this history, in time t there are only two possible combinations of (a_t, λ_t), namely $\left(a^g, H(\tilde{\lambda}_{t-1}, \tilde{a}_{t-1}, a^g)\right)$ and $\left(a^b, H(\tilde{\lambda}_{t-1}, \tilde{a}_{t-1}, a^b)\right)$, with known probabilities $\pi(a^g|\tilde{a}_{t-1})$ and $\pi(a^b|\tilde{a}_{t-1})$, and with the two values of transition function H that are implied by the individual policy functions at time $t-1$. Therefore, if we knew the value function for these two values of (a_t, λ_t) and for all possible (k_t, e_t), then computing the period-$(t-1)$ continuation value $\psi(\cdot, \cdot, \tilde{a}_{t-1}, \tilde{\lambda}_{t-1})$ in (10) could be done for all k_t and e_{t-1} by simply applying the Markov transition matrix for (a, e):

$$\psi(k_t, e_{t-1}, \tilde{a}_{t-1}, \tilde{\lambda}_{t-1}) = \mathrm{E}_{a_t, e_t}\left\{V\left(k_t, e_t, a_t, H(\tilde{\lambda}_{t-1}, \tilde{a}_{t-1}, a_t)\right)|e_{t-1}, \tilde{a}_{t-1}\right\} \tag{12}$$

Thus, observing a time evolution of such an economy up to time $t-1$ provides a set of values $\{\psi(\cdot, \cdot, a_\tau, \lambda_\tau)\}_{\tau=1}^{t-1}$.

The main idea of our method is to simulate a sequence of shocks; at each time t to approximate $\psi(\cdot, \cdot, a_t, \lambda_t)$ by the k-nn regression estimator $\hat{\psi}_t$ (11) using the simulated values Ψ_{t-1}; and substitute this approximation into the time t optimization problem (9) in order to find the *approximate* value function $\tilde{V}_t$:

$$\tilde{V}_t(k_t, e_t, a_t, \lambda_t) = \max_{c, k'}\left\{u(c) + \beta\hat{\psi}_t(k', e_t, a_t, \lambda_t)\right\} \tag{13}$$

subject to (4)-(8). This newly-obtained estimate $\tilde{V}_t$ is then used to compute an approximate continuation value at time $(t-1)$:

$$\tilde{\psi}_{t-1}(k_t, e_{t-1}, \tilde{a}_{t-1}, \tilde{\lambda}_{t-1}) = \mathrm{E}_{a_t, e_t}\left\{\tilde{V}_t\left(k_t, e_t, a_t, H(\tilde{\lambda}_{t-1}, \tilde{a}_{t-1}, a_t)\right)|e_{t-1}, \tilde{a}_{t-1}\right\} \tag{14}$$

which, in its' turn, is then added to the set of observations:

$$\Psi_t = \Psi_{t-1} \cup \left(\tilde{a}_{t-1}, \tilde{\lambda}_{t-1}, \tilde{\psi}_{t-1}(\cdot, \cdot, \tilde{a}_{t-1}, \tilde{\lambda}_{t-1})\right) \tag{15}$$

This way, the problem is solved iteratively *forward* in time, with each new iteration corresponding to the next simulated time period, as opposed to the backward direction (with each new iteration corresponding to the previous time period) in the standard value function iteration algorithm.

One complication that arises in this approach is that it results in a data-generating process which is not stationary due to the ongoing learning by the agents. For example, as the simulation progresses and more observations become available, the agents learn their continuation values better, and their value function approximations and policy decisions improve. In order to mitigate the effect of non-stationarity, we only use the most recent $m(t)$ real-

izations: $\Psi_t = \left\{a_j, \lambda_j, \tilde{\psi}_j(\cdot,\cdot,a_j,\lambda_j)\right\}_{j=t-m(t)-1}^{t-1}$, where $m(t)$ is an unbounded, monotonically increasing function defined on natural numbers such that $2 \leq m(t) \leq t-1$. For example if $m(t) = \max(2, \min(t-1, 0.1t))$, the look-back period only includes the most recent 10% of the sample.

The complete procedure is summarized in Algorithm 1. This procedure can be further refined, but in some cases, especially in models calibrated to annual data and thus having lower values of the discount factor β, it may already be capable of producing an acceptable solution in reasonable time.

3.2 Distance measurement

The algorithm described in the previous section requires a distance metric between two probability distributions. In our implementation, the distribution is defined on a grid of values of $k \in \mathcal{K}, e \in \{0,1\}$. The possible metrics thus include those induced by L_1, L_2, or L_∞ norms in the space of empirical distribution functions.

Recently, Sriperumbudur, Gretton, Fukumizu, Schölkopf, and Lanckriet (2010) have proposed a group of kernel-based distance metrics, and shown that they have attractive properties for learning applications. For a given kernel function $\kappa(x,y)$, an induced distance metric between two conditional empirical densities, $\lambda_{k|e}(\cdot|e)$ and $\lambda'_{k|e}(\cdot|e)$ takes the form

$$d_\kappa(\lambda,\lambda') = \sum_{k\in\mathcal{K}}\sum_{k'\in\mathcal{K}} \kappa(k,k')\left[\lambda(k)-\lambda'(k)\right]\left[\lambda(k')-\lambda'(k')\right] \tag{16}$$

In this paper, we use a distance metric based on a Gaussian kernel, $\kappa(x,y) = e^{-(x-y)^2/\sigma^2}$ to compare the conditional distributions across capital $k \in \mathcal{K}$, for each value of employment status e. The two conditional distributions are then weighted with the probabilities of each employment status e:

$$d(\lambda,\lambda') = \sum_{e\in\{0,1\}} d_\kappa(\lambda(k|e),\lambda'(k|e))\pi(e) \tag{17}$$

Choice of the distance metric is important for convergence. Since different metrics emphasize different divergent features of distributions, metrics that focus on features of low relevance do poorly at signal extraction and matching neighbors, resulting in noisier k-nn estimates $\hat{\psi}$, leading to noisier policy functions, and poor (or no) convergence.

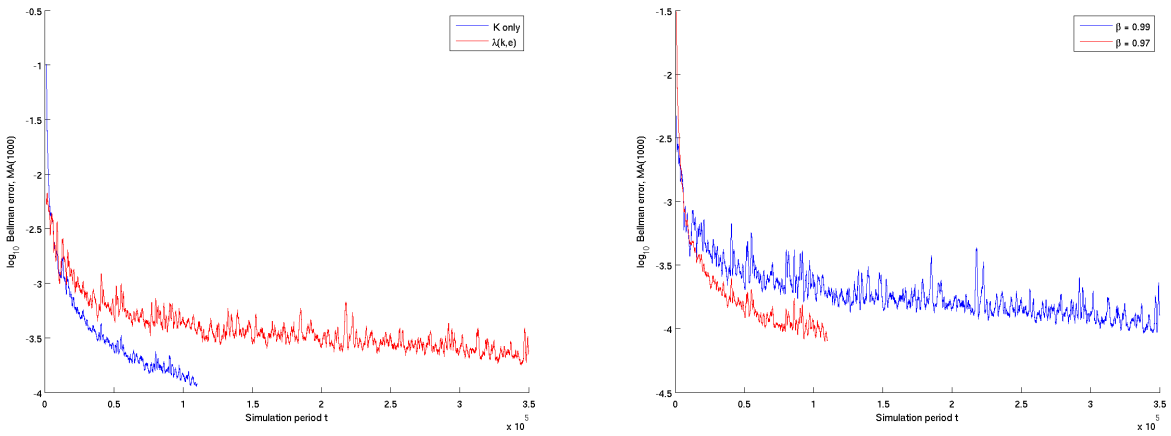
It is necessary to note, however, that even though in this paper we, for demonstration purposes, explicitly compute the distance between the distribution functions at their highest level of disaggregation, in some cases there may be prior economic reasons why a smaller number of specific aggregate statistics would be expected to matter. Krusell and Smith

(2006) argue that the model considered in this paper is in fact such a case, and the agents' decisions in it are primarily governed by the mean capital stock. In order to take such information into account, one could, for example, measure distances between distributions by distances between the corresponding aggregates, thus reducing the estimation noise and improving the continuation value estimates, while still retaining full flexibility in terms of their transitional dynamics (i.e. avoiding the linearity assumption). Not surprisingly, in this example explicit accounting for such information improves the speed of convergence (see Figure 1, Panel A).

Figure 1: Convergence of NPRL

$\log_{10}$ of the mean absolute Bellman equation error $\varepsilon_t = |\hat{\psi}_t - \tilde{\psi}_t|$, 1000-period moving average. Left: Convergence under different metrics: mean-K only (blue line), fully disaggregated with d_κ (red line). $\nu = 0.05$. Right: Convergence for different values of discount factor β. $\nu = 0.10$, distance on $\lambda(k)$.

A: Convergence and distance metrics **B: Convergence and β**



3.3 Improving convergence

Convergence of algorithm 1 ensures that the resulting solution is optimal in the limited-rationality sense, that is, the agents are content with their decision rule to the extent that value forecasts they are making are off by no more than $\bar{\varepsilon}$. Unfortunately, in practice convergence of reinforcement learning algorithms can be slow, especially in models where the discount factor β is close to unity (see Figure 1, Panel B).

We propose a refined based on the principle of "temporal difference" (TD) learning (Sutton, 1988). The intuition for the adjustment is as follows. Note that if the continuation

value estimator $\hat{\psi}$ is unbiased, at time $t-1$ we have[3]:

$$\mathrm{E}_{t-1}\tilde{\psi}_{t-1} = \hat{\psi}_{t-1} \equiv \frac{1}{M}\sum_{j=1}^{M}\tilde{\psi}_{\tau_j}$$

and the expected one-step discrepancy is equal to zero:

$$0 = \mathrm{E}_{t-1}\varepsilon_{t-1} \equiv \mathrm{E}_{t-1}(\hat{\psi}_{t-1} - \tilde{\psi}_{t-1}) = \mathrm{E}_{t-1}\left(\tilde{\psi}_{t-1} - \frac{1}{M}\sum_{j=1}^{M}\tilde{\psi}_{\tau_j}\right) = \frac{1}{M}\mathrm{E}_{t-1}\sum_{j}\left(\tilde{\psi}_{t-1} - \tilde{\psi}_{\tau_j}\right)$$

In case there is a bias, $\mathrm{E}\varepsilon_{t-1}$ is no longer zero, and the sample realization of ε_{t-1} (observed at time t) is an estimate of this bias. Since the predicted continuation value $\hat{\psi}_{t-1}$ was computed as a mean of past continuation values, a bias in $\hat{\psi}_{t-1}$ implies that, on average, there is also a bias in $\{\tilde{\psi}_{\tau_j}\}$. Adjusting $\hat{\psi}_{t-1}$ by a fraction α_{TD} of the bias, where $0 \le \alpha_{\mathrm{TD}} \le 1$, we have

$$\hat{\hat{\psi}}_{t-1} \equiv \hat{\psi}_{t-1} - \alpha_{\mathrm{TD}}(\hat{\psi}_{t-1} - \tilde{\psi}_{t-1}) = \frac{1}{M}\sum_{j=1}^{M}\left[(1-\alpha_{\mathrm{TD}})\tilde{\psi}_{\tau_j} + \alpha_{\mathrm{TD}}\tilde{\psi}_{t-1}\right] \tag{18}$$

i.e., such an adjustment can be done at time t by "nudging" each of the neighbors in step $t-1$ towards $\tilde{\psi}_{t-1}$, thus affecting future estimates that would depend on these neighbors. The entire modified procedure is summarized as Algorithm 2.

4 Results

The solution proceeds as follows: first, a sequence of $T = 350,000$ random aggregate shocks is generated, and Algorithm 2 is applied to find a solution, which in this case is represented by a lookup table Ψ_T. This lookup table is subsequently used to simulate an economy that is subjected to a predefined sequence of 10,000 aggregate shocks to productivity, and a single agent within this economy, who has her starting value of individual capital equal to 43, and is subjected to another predefined sequence of 10,000 individual shocks to employment. Both sequences are as specified in den Haan, Judd, and Juillard (2010).

As a benchmark to compare our nonparametric reinforcement learning (NPRL) algorithm against, we select the implementation of the KS algorithm by Maliar, Maliar, and Valli (2010) (henceforth KS-sim), subject to the same test shock sequence.

The baseline parameters of the NPRL algorithm are:

[3]Note that $\tilde{\psi}_{t-1}$ here is an object from the time t information set.

- Window width coefficient: $\nu = 0.05$
- Kernel bandwidth: $\sigma^2 = 30,000$
- Grid for capital: 501 points uniformly spaced on $[0, 100]$
- "Temporal difference" adjustment factor: $\alpha_{\text{TD}} = 0.5$
- Number of neighbors $\overline{M} = 4$

The remaining parameters of the model correspond to den Haan, Judd, and Juillard (2010):

- Time-discount factor (quarterly): $\beta = 0.99$
- Coefficient of relative risk aversion: $\gamma = 1$
- Capital share of total output: $\alpha = 0.36$
- Capital depreciation rate: $\delta = 0.025$
- Labor endowment: $\bar{l} = 1/0.9$
- Unemployment benefit: $\mu = 0.15$
- Standard deviation of aggregate productivity shocks $\Delta_a = 0.01$

4.1 Aggregate law of motion

Figure 2 shows part of the sample path of the aggregate law of motion of capital, according to the solution by the KS-sim and NPRL algorithms. Table 1 presents summary statistics of the capital stock per capita[4].

Clearly, the resulting aggregate dynamics are very similar, both across employment status and across productivity states, indicating that the KS algorithm is indeed well-suited for its' namesake application, and that its' assumptions are not overly restrictive in this case.

4.2 Accuracy evaluation

We measure the accuracy of solution by Euler equation errors (den Haan, 2010b, Judd, 1992). For a given agent, the Euler equation error at time t is defined as the percentage difference between the computed period-t consumption and that implied by the Euler equation:

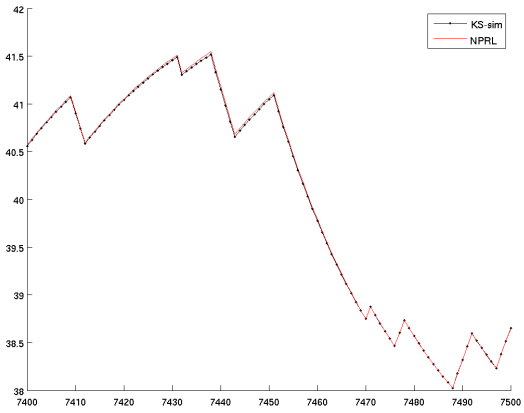
$$\mathrm{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} (1 + r_{t+1} - \delta) \right] = 1 \tag{19}$$

[4]Since the initial state of the economy at the beginning of the simulation would be generally different between the two methods (in case of the NPRL model, determined by the realizations of the random shocks at the end of the training sample), we drop first 1,000 periods of the test sequence, and compute the statistics over the remaining 9,000 periods.

Figure 2: Dynamics of Aggregate Capital

Aggregate capital dynamics in two solutions: Maliar, Maliar, and Valli (2010) (KS-sim, black lines) and nonparametric reinforcement learning (NPRL, red lines, 350,000 periods simulated, TD with $\alpha = 0.5$, kernel distance with $\sigma^2 = 30,000$, $\nu = 0.1$.) Time is from the beginning of the aggregate shock sequence in den Haan, Judd, and Juillard (2010). Panel A: aggregate capital; Panel B: aggregate capital by employment status.

A: Aggregate capital, all agents



B: Aggregate capital, by employment status

Table 1: Aggregate Capital Stock

Time-series means and standard deviations of capital stock per capita. Last 9,000 periods of the sample sequence in den Haan, Judd, and Juillard (2010). Recessions and expansions defined as periods of low and high aggregate productivity, respectively. KS-sim is the solution in Maliar, Maliar, and Valli (2010), NPRL is the solution produced by the nonparametric reinforcement-learning algorithm.

	Full sample		Recessions		Expansions	
	KS-sim	NPRL	KS-sim	NPRL	KS-sim	NPRL
Means:						
Total	39.333	39.321	39.040	39.027	39.645	39.635
Employed	37.697	37.684	36.974	36.959	38.467	38.456
Unemployed	39.475	39.463	39.269	39.257	39.694	39.684
Standard Deviations:						
Total	1.026	1.031	0.988	0.992	0.971	0.977
Employed	1.456	1.460	1.274	1.278	1.223	1.228
Unemployed	0.989	0.994	0.968	0.972	0.964	0.970

We exclude the periods in which the agent is binding by the liquidity constraint (by setting the Euler equation error to zero), since the Euler equation does not hold in those periods.

Euler equation errors primarily reflect the accuracy of the solution of the individual problem. In this paper, we used a piecewise-linear approximation of the value function on the grid of capital $\mathcal{K}$. In addition to a uniform equally-spaced grid, we evaluated a polynomial grid suggested in Maliar, Maliar, and Valli (2010), with grid points defined as

$$k_j = \bar{k} \left(j/N\right)^{\theta} \tag{20}$$

where N is the grid size and $\bar{k} = 100$ is the upper bound for capital; we consider a number of values for the power parameter θ. In Table 2 we report mean and maximum errors for different grid sizes and types. Clearly, an irregular grid is superior to the uniformly-spaced one for solving the individual problem. However, grid choice had very little effect on the dynamics of the *aggregate* capital: for example, while individual Euler equation errors were at 9% in case of a uniformly-spaced 501-point grid, the largest difference between the two resulting series of aggregate capital was only 0.07%. The reason for this is that the largest Euler equation errors are observed for agents with very low level of capital, whose decisions have a very small effect on the aggregate investment.

Table 2: Euler equation errors and prediction R^2.

Euler equation errors (19). 350,000 simulations; kernel distance with $\sigma^2 = 30,000$. Column 1: number of grid points for capital on $[0, 100]$, for each level of individual shock. $p(n)$ denotes a polynomially-spaced grid (20) with power factor n. Euler equation errors are in percent of current consumption. Presented are time-series mean and maximum of the Euler equation errors for a single simulated agent, using the individual and aggregate shock sequences as per den Haan, Judd, and Juillard (2010).

Grid points	Spacing	EE error, %: mean	EE error, %: max
1001	Uniform	0.0937	8.5426
	p(2)	0.0923	0.4302
	p(4)	0.0933	0.2981
	p(7)	0.0968	0.3499
501	Uniform	0.0976	9.2257
	p(2)	0.0953	0.8661
	p(4)	0.1005	0.6313
	p(7)	0.1125	0.4063
KS-sim		0.0930	0.4360

Accuracy of the KS method is often measured by the R^2 statistic. While, as den Haan (2010a) notes, this measure does not reflect the accuracy of the solution well, it is nevertheless an indicator that reflects how well the actual aggregate law of motion corresponds to the one perceived by the agents, and serves to evaluate whether the "limited-rationality" assumption of a linear aggregate law of motion is realistic. Since our method relies on approximation of the continuation values for each agent, rather than of the aggregate law of motion for capital, a similar statistic is the R^2 of the one-step-ahead k-nn predictor $\tilde{\psi}$ (11), which can be computed for any (k, e) as follows:

$$R^2 = 1 - \sum_{t=1}^{T} \frac{(\tilde{\psi}_t - \hat{\psi}_t)^2}{(\tilde{\psi}_t - \overline{\psi})^2} \tag{21}$$

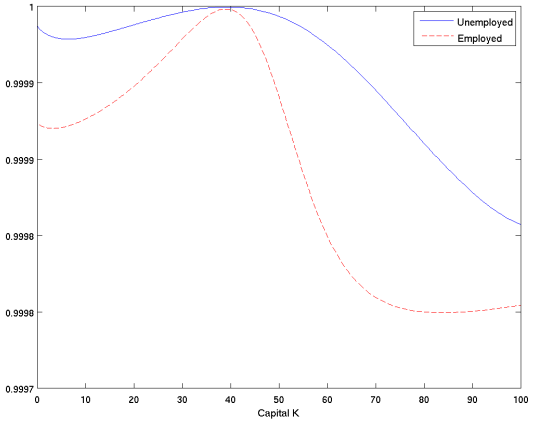
where $\overline{\psi}$ is the sample mean of $\tilde{\psi}_t$. Similarly to the R^2 statistic in the KS method, a high R^2 implies that an agent finds her estimates of the continuation value to be sufficiently precise.

Figure 3 shows R^2 of the nearest-neighbor estimator for different values of k and e, Panel A corresponding to a simulation with distance measured between the full λ distributions, and Panel B to one with distance between mean capital only. Both methods result in high values of the R^2 statistic for all agents, although slightly higher in the capital-only case (equally-weighted mean $R^2 = 0.999988$) than in the full-distribution case (mean $R^2 = 0.999948$).

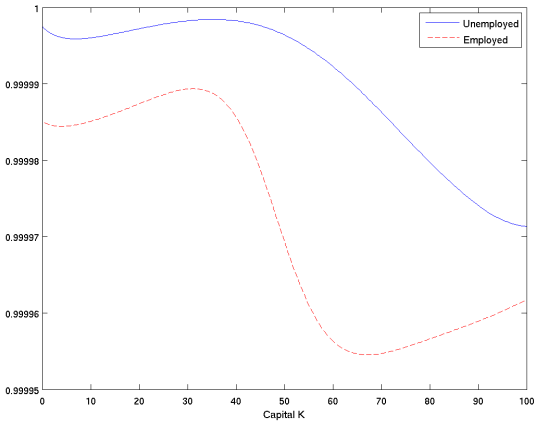
Figure 3: R^2 of the k-nn estimator

One-step-ahead R^2 of the k-nn continuation-value estimator for each value of individual capital and employment status. 350,000 simulations; 500-point polynomial grid (20) with $\theta = 4$. A: kernel distance metric with $\sigma^2 = 30,000$. B: distance between means of capital.

A: Full λ



B: mean K only



5 Conclusion

In this paper, we have develop a method of solving heterogeneous agent models in which individual decisions depend on the entire cross-sectional distribution of individual state variables, that does not require parametric assumptions on either the agents' information set, or on the functional form of the aggregate dynamics.

As an illustration, we apply it to the classic Krusell and Smith economy, as described in den Haan, Judd, and Juillard (2010). Our unconstrained solution of this model is very close to the limited-rationality solution of the original Krusell and Smith algorithm.

Even though in this paper we focus on a heterogeneous-agent setting with aggregate uncertainty, we believe that related approximate optimization methods could prove useful in other large economic models, such as multi-country growth models, as well.

References

Algan, Y., O. Allais, and W. J. den Haan (2010): "Solving the incomplete markets model with aggregate uncertainty using parametrized cross-sectional distributions," *Journal of Economic Dynamics and Control*, 34, 59–68.

Cover, T., and P. Hart (1967): "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on Information Theory*, 13(1), 21–27.

den Haan, W. J. (1996): "Heterogeneity, aggregate uncertainty, and the short-term interest rate," *Journal of Business & Economic Statistics*, 14(4), 399–411.

——— (2010a): "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents," *Journal of Economic Dynamics and Control*, 34(1), 79–99.

——— (2010b): "Comparison of solutions to the incomplete markets model with aggregate uncertainty," *Journal of Economic Dynamics and Control*, 34(1), 4–27.

den Haan, W. J., K. Judd, and M. Juillard (2010): "Computational suite of models with heterogeneous agents: multi-country real business cycle models," *Journal of Economic Dynamics and Control*, 34.

den Haan, W. J., and P. Rendahl (2010): "Solving the incomplete markets model with aggregate uncertainty using Krusell-Smith algorithm and non-stochastic simulations," *Journal of Economic Dynamics and Control*, 34, 36–41.

Fix, E., and J. Hodges (1951): "Discriminatory analysis. Nonparametric discrimination: Consistency properties," Technical report 4, project number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

Judd, K. L. (1992): "Projection methods for solving aggregate growth models," *Journal of Economic Theory*, 58(2), 410–452.

Kim, S. H., R. Kollmann, and J. Kim (2010): "Solving the incomplete markets model with aggregate uncertainty using a perturbation method," *Journal of Economic Dynamics and Control*, 34, 50–58.

Kirman, A. (1992): "Whom or what does the representative individual represent?," *The Journal of Economic Perspectives*, 6(2), 117–136.

Kocherlakota, N. (2010): "Modern macroeconomic models as tools for economic policy," *The Region*, pp. 5–21.

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

——— (2006): "Quantitative macroeconomic models with heterogeneous agents," in *Advances in Economics and Econometrics: theory and Applications, ninth World congress*, vol. 1, pp. 298–340.

Maliar, L., S. Maliar, and F. Valli (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm," *Journal of Economic Dynamics and Control*, 34(1), 42–49.

Maliar, S., L. Maliar, and K. Judd (2010): "Solving the multi-country real business cycle model using ergodic set methods," Discussion paper, National Bureau of Economic Research.

Nascimento, J., and W. Powell (2010): "Dynamic programming models and algorithms for the mutual fund cash balance problem," *Management Science*, 56(5), 801–815.

Powell, W. B. (2007): *Approximate Dynamic Programming: Solving the curses of dimensionality*. Wiley-Interscience.

Reiter, M. (2010): "Solving the incomplete markets model with aggregate uncertainty by backward induction," *Journal of Economic Dynamics and Control*, 34, 28–35.

Simão, H., J. Day, A. George, T. Gifford, J. Nienow, and W. Powell (2009): "An approximate dynamic programming algorithm for large-scale fleet management: A case application," *Transportation Science*, 43(2), 178–197.

Sriperumbudur, B., A. Gretton, K. Fukumizu, B. Schölkopf, and G. Lanckriet (2010): "Hilbert space embeddings and metrics on probability measures," *The Journal of Machine Learning Research*, 99, 1517–1561.

Sutton, R. (1988): "Learning to predict by the methods of temporal differences," *Machine learning*, 3(1), 9–44.

Sutton, R., and A. Barto (1998): *Reinforcement learning*. MIT Press.

Tesauro, G. (1994): "TD-Gammon, a self-teaching backgammon program, achieves master-level play," *Neural computation*, 6(2), 215–219.

Algorithm 1 Stochastic simulation

1: Define a grid $\mathcal{K}$ over capital, $\mathcal{K} = (k_1, ..., k_N)$
2: Pick initial realization of the aggregate shock $\tilde{a}_0$, initial approximation $\hat{\psi}_0(k', e, a)$, and initial distribution $\lambda_1(k, e|a)$.
3: Pick tolerance $\overline{\varepsilon}$, maximum number of neighbors $\overline{M}$, and lookback window parameter $m(t)$, e.g. $m(t) = 0.1t$
4: Initialize $t \leftarrow 1$, $\Psi_0 \leftarrow \emptyset$
5: **repeat**
6: **for** all possible values of the aggregate state $a \in \{a_b, a_g\}$ and corresponding capital distribution $\lambda_t(\cdot, \cdot|a)$ **do**
7: Compute aggregate capital $K \leftarrow \sum_{k \in \mathcal{K}} [\lambda_t(k, 0|a) + \lambda_t(k, 1|a)]k$
8: Compute state-dependent wage $w(K, a)$ and interest rate $r(K, a)$
9: Search Ψ_{t-1} for M nearest realizations $\{t_1, ..., t_M\}$, i.e. find the largest $M \leq \overline{M}$ and $\{t_j\}_{j=1}^{M} \subset (t - m(t), ..., t - 2)$, such that $\forall j = 1...M$, $a_{t_j} = a$, and $d(\lambda_t, \lambda_{t_j}) \leq d(\lambda_t, \lambda_\tau)\ \forall \tau \notin \{t_1, ..., t_M\}$.
10: **for** all possible values of the individual state $e \in \{0, 1\}$ and capital $k \in \mathcal{K}$ **do**
11: Compute $\hat{\psi}_t(k', e, a, \lambda_t) \leftarrow \frac{1}{M} \sum_{j=1}^{M} \tilde{\psi}_{t_j}((k', e, \tilde{a}_{t_j}, \tilde{\lambda}_{t_j}))$, $k' \in \mathcal{K}$ if $M > 0$, or otherwise set $\hat{\psi}_t \leftarrow \hat{\psi}_0$
12: Solve the optimization problem (9) using $\hat{\psi}_t$ in place of ψ, and determine $k'_t(k, e, a)$ and $V_t(k, e, a)$
13: **end for**
14: **end for**
15: Compute $\tilde{\psi}_{t-1} \leftarrow \mathrm{E}\{V_t(k_t, e_t, a_t)|e_{t-1}, \tilde{a}_{t-1}\}$, using the Markov transition matrix $\pi(a', e'|a, e)$
16: Compute discrepancy $\varepsilon_{t-1} \leftarrow \max_{k,e} |\hat{\psi}_{t-1}(k, e, \tilde{a}_{t-1}) - \tilde{\psi}_{t-1}(k, e, \tilde{a}_{t-1})|$
17: Add an observation $(\lambda_{t-1}, \tilde{a}_{t-1}, \tilde{\psi}_{t-1})$ to the lookup table: $\Psi_t \leftarrow \Psi_{t-1} \cup (\lambda_{t-1}, \tilde{a}_{t-1}, \tilde{\psi}_{t-1})$
18: Generate the current realization of the aggregate shock $\tilde{a}_t$ according to $\pi(a_t|\tilde{a}_{t-1})$
19: Using the policy function $k'_t(k, e, a)$ found in step 12, compute the next period capital distribution $\lambda_{t+1}(k_{t+1}, e_{t+1}, a_{t+1})$ for all (e_{t+1}, a_{t+1}) and $k_{t+1} \in \mathcal{K}$
20: advance $t \leftarrow t + 1$
21: **until** $\max_{t-T_0 \leq \tau \leq t-1} \varepsilon_\tau < \overline{\varepsilon}$

Algorithm 2 Stochastic simulation with temporal-difference adjustment

1: Select the temporal difference update factor $0 \leq \alpha_{\mathrm{TD}} \leq 1$
2: Initialize $M' \leftarrow 0$
3: Proceed with steps 1 - 17 of Algorithm 1
4: if $t > 1$ and $M' > 0$, for each of the nearest neighbors of period $(t - 1)$, $\tau_1, ..., \tau_{M'}$, perform an adjustment: $\tilde{\psi}_{\tau_j} \leftarrow (1 - \alpha_{\mathrm{TD}})\tilde{\psi}_{\tau_j} + \alpha_{\mathrm{TD}}\tilde{\psi}_{t-1}(k, e, \tilde{a}_{t-1})$
5: Save $M' \leftarrow M$ and $\{\tau_j\} \leftarrow \{t_j\}$, $j = 1...M$
6: Proceed with the remaining steps of Algorithm 1 until completion

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Department of Economics
University of Alicante

Campus San Vicente del Raspeig
03071 Alicante - Spain
Phone: +34 965 903 563
Fax: +34 965 903 898

Website: www.ivie.es
E-mail: publicaciones@ivie.es