Deep learning: Solving HANC and HANK models in the absence of Krusell-Smith aggregation*

Lilia Maliar† Serguei Maliar‡

December 12, 2020

Abstract

Heterogeneous-agent neoclassical model (HANC) studied by Krusell and Smith (1998) has savings through capital. This model has a remarkable feature of approximate aggregation: the mean of wealth distribution can be accurately predicted with the mean of past wealth distribution. However, if savings are done through bonds, the HANC model does not have this feature (because the mean of bond holding is zero). We solve such model using deep learning solution method in which the decision function and price functions are approximated in terms of true state space of individual and aggregate state variables. The problem has high dimensionality (hundreds of state variables) and ill-conditioning but neural network reduces dimensionality and restore numerical stability. Our deep learning method delivers accurate and reliable solutions. We also show how to solve a heterogeneous-agent new Keynesian (HANK) model with savings through bonds and a zero lower bound on the nominal interest rate in the absence of Krusell and Smith (1998) aggregation.

Key Words : artificial intelligence, machine learning, deep learning, neural network, stochastic gradient, HANK, Krusell and Smith, ZLB, ELB

*Lilia Maliar and Serguei Maliar acknowledge the support from NSF grants SES-1949413 and SES-1949430, respectively. We thank Marc Maliar for his help with writing a TensorFlow code for producing numerical solutions.

†The Graduate Center, City University of New York and CEPR.

‡Santa Clara University

1 Introduction

In a seminal paper, Krusell and Smith (1998) analyze a heterogeneous-agent neoclassical model (HANC) with production in which the agents save through capital. They discovered that such a model has a remarkable feature of "approximate aggregation", namely, the mean of future capital distribution can be accurately predicted with just the mean of the current capital distribution. In a companion paper, Krusell and Smith (1997) analyzed a similar model in which agents save through both capital and bonds and they found that such a model also has the property of perfect aggregation. However, the analysis of Krusell and Smith (1997, 1998) does not contain the version of HANC model in which the agents save only through bonds. A peculiar feature of such exchange economy is that the mean of the bond distribution is zero by construction, so it cannot be used as a predictor of future variables and hence, approximate aggregation does not hold.[1]

The HANC class of models received a considerable attention in the literature, in particular, den Haan, Judd and Juliard (2008) edited a special issue of JEDC that summarize the computational approaches proposed for analyzing such models. The participants of the special issue were invited to solve two HANC models: one with savings through capital and the other with savings through bonds. There were many methods that worked accurately and reliably for the former model but the participants did not succeed in producing accurate solutions to the later model. In particular, they tried out higher moments and other statistics of the bond distribution such as histogram or fraction of agents facing the borrowing constraint but those statistics did not have sufficient power for predicting the future aggregates. Eventually, the HANC model with savings through bonds was eventually removed from the original JEDC call and it was not studied in the literature to the best of our knowledge.

In the present paper, we use deep learning to solve the model with savings through bonds. The algorithm that we use essentially relies on Krusell and Smith (1998) style simulation. The critical difference is that they approximate decision function of an agent in terms of the individual state variables of that agent and selected moments of aggregate distributions while we approximate decision functions in terms of all individual and aggregate state variables and we let the neural network to choose how the information contained in that distribution can be reduced and condensed into smaller set of artificial features. For example, in the economy with 1000 agents, the decision functions of agents depend on 2001 state variables while we reduce it to 64 composite features.

Krusell and Smith (1998) were able to guess the composite variable that contain all necessary information for describing the aggregate behavior of the economy which is the 1st moment but it is more difficult to find such a representation for the model with bonds. However, we show that if we take enough moments, our DL method can still produce an accurate solution. In turn, Krusell and Smith (1998) method will not be able to produce so accurate solution. The reason is that Krusell and Smith (1998) method constructs separately the individual and aggregate laws of motion. We show that finding the aggregate law of motion is problematic and the explanatory power of Krusell and Smith (1998) regression is relatively low. Our DL algorithm does not involve a separate approximation of the law of motion for aggregate variables and an alternation between the individual and aggregate approximations and it does not suffer from that problem. We just simulate the panel of heterogeneous agents and use the resulting distributions to infer the aggregate quantities and prices as the economy evolves over time.

Our DLC method is related to recent papers on deep learning, including Duarte (2018), Villa and Valaitis (2019), Fernández-Villaverde, Hurtado, and Nuño (2019), Azinović, Luca and Scheidegger

[1] Related HANC models were also studied in den Haan (1997) and Gaspar and Judd (1997) but they also do not consider the version of the model with savigs through bonds alone.

(2019), Lepetyuk, Maliar and Maliar (2020) and especially, Maliar, Maliar and Winant (2018, 2019), see the latter paper for a discussion of the literature. Recent numerical methods for solving Krusell and Smith's (1998) model include Ahn et al. (2017), Bayer and Luetticke (2018), Boppart et al. (2018), and Fernández-Villaverde et al. (2018). These new developments in solution techniques are primarily motivated by recent interest in modeling the effects of fiscal and monetary policies on distributions. Our DL method differs from the literature in that it requires essentially no simplifying assumptions and can solve the studied class of models both accurately and reliably.

The rest of the paper is as follows: In Section 2, we analyze the HANC model with savings through bonds; in Section 3, we solve the HANK model; and finally, in Section 3, we conclude.

2 Heterogenous agent neoclassical model (HANC) with savings through bonds

Ex ante agents are exactly the same, but ex post they differ due to the presence of idiosyncratic shocks. In particular, the endowment of the agent is equal to $y_t^i z_t$ where is the idiosyncratic shock and at is an aggregate shock that is the same for each agent. Both processes are assumed to be first-order Markov processes. Markets are incomplete but the agents can smooth their consumption by trading in a risk-free one-period bond.

Agent i's maximization problem is as follows:

$$\max_{\left\{c_t^i, b_{t+1}^i\right\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t^i\right) \right] \tag{1}$$

$$\text{s.t. } c_t^i + q_t b_{t+1}^i = b_t^i + y_t^i z_t, \tag{2}$$

$$b_{t+1}^i \geq \overline{b}, \tag{3}$$

where c_t^i is the amount of consumption of agent i in period t, b_{t+1}^i is the demand for one-period bonds that pay one unit of the consumption commodity in the next period, q_t is the price of this one-period bond; $\overline{b}$ is the borrowing limit.

Idiosyncratic and aggregate shocks follow identical AR(1) processes normalized for convenience to

$$\begin{aligned} \ln\left(y_t^i\right) &= -\frac{0.5\left(1-\rho_y\right)\sigma_y^2}{1-\rho_y^2} + \rho_y \ln\left(y_{t-1}^i\right) + \sigma_y \varepsilon_{y,t}^i, \\ \ln\left(z_t\right) &= -\frac{0.5\left(1-\rho_z\right)\sigma_z^2}{1-\rho_z^2} + \rho_z \ln\left(z_{t-1}\right) + \sigma_z \varepsilon_{z,t}, \end{aligned}$$

where $\varepsilon_{y,t}^i$ and $\varepsilon_{z,t}$ are i.i.d. random variables with a standard Normal distribution. The normalizing constant guarantees that the mean of both y_t^i and z_t is equal to one. The initial condition z_0 and (b_0^i, y_0^i) for all i is given.

Equilibrium conditions of the individual problem (1)-(3) is given by Kuhn Tucker condition

$$\eta_t h_t = 0,$$

where $\eta_t \equiv b_{t+1}^i + \overline{b} \geq 0$ and $h_t \equiv q_t u'\left(c_t^i\right) - \beta E_t\left[u'\left(c_{t+1}^i\right)\right]$ are Lagrange multipliers associated with the first order condition and borrowing constraint respectively. The individual problem is affected by variables of other agents exclusively through the price of bonds q_t. The aggregate equilibrium condition is that the market clearing in the bond market, namely, the supply and demand of bonds are equal meaning that the aggregate bond holding is zero.

$$\int_{\overline{b}}^{\infty} b_{t+1}^{i}\left(b_t^i, y_t^i, z_t\right) dF_t\left(b_t^i, y_t^i\right) = 0$$

where $F_t\left(b_t^i, y_t^i\right)$ is the cross sectional cumulative joint distribution function of beginning-of-period bond holdings and income.

State variables The model has $2n+1$ state variables where ℓ is the number of heterogeneous agents, including ℓ endogenous state variables $\{b_t^i\}$, $i = 1, ..., \ell$ and ℓ idiosyncratic productivity shocks $\{y_t^i\}$ and and one aggregate shock z_t.

Market clearing condition The individual problem is affected by variables of other agents exclusively through the price of bonds. The aggregate equilibrium condition is that the market clearing in the bond market, namely, the supply and demand of bonds are equal meaning that the aggregate bond holding is zero.

$$\int_{\overline{b}}^{\infty} b_{t+1}^{i}\left(b_t^i, y_t^i, z_t\right) dF_t\left(b_t^i, y_t^i\right) = 0$$

where $F_t\left(b_t^i, y_t^i\right)$ is the cross sectional cumulative joint distribution function of beginning-of-period bond holdings and income.

2.1 Deep learning solution algorithm

Our analysis builds on deep learning solution algorithm introduced by Maliar, Maliar and Winant (2018, 2019). Following that paper, we cast the equations of the studied heterogenous agent models into objective functions of the DL framework, we use the stochastic gradient descent method for training the machine, we use one or few (batch) random grid points on each iteration, instead of conventional fixed grid with a large number of grid points and we approximate the kink due to the borrowing constraint by using the Fischer-Burmeister function. Importantly, we use all-in-one expectation method for approximating high-dimensional integrals with just two random draws. This technique enables us the use DL-style Monte Carlo simulation for evaluating expectation with respect to both state variables and future shocks simultaneously, merging in one step the approximation of decision functions and the integration with respect to future shocks. Finally, we write the code in TensorFlow in a way that makes it ubiquitous and portable to a variety of economic models and applications; see Maliar, Maliar and Winant (2019) for discussion and computational details.

Let $s_t \equiv S\left(\{b_t^i, y_t^i\}_{i=1}^{\ell}\right)$ be a set of individual characteristics in the economy. A distinctive feature of our DL framework is that it enables to solve the HANC model by working directly with the actual state space. Let us introduce variable disposable wealth $w_i = b_t^i + y_t^i z_t - q_t \overline{b}$ in which we include the possibility of borrowing. We specifically parameterize the consumption share of agent i by $\frac{c_t^i}{w_t^i} \equiv \zeta_t^i = \sigma\left(\zeta_0 + \varphi\left(b_t^i, y_t^i, \{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t; \theta\right)\right)$, where $\varphi(\cdot)$ is a neural network and $\sigma(x) = \frac{1}{1+e^{-x}}$ is a sigmoid function, which ensures that ζ_t^i is bounded to be in interval $[0, 1]$. Furthermore, we parameterize the Lagrange multiplier $\eta_t^i = \exp\left(\varphi\left(b_t^i, y_t^i, \{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t; \theta\right)\right)$. If agents are heterogenous in fundamentals, we need to approximate 2ℓ different individual decision functions, each of which has $2\ell+1$ dimensions. With symmetric agents, as in Krusell and Smith (1998), we need just 2 decision functions with $2\ell+1$–dimensions to characterize the choices of all ℓ agents.

In the approximating function of the consumption share $\varphi\left(b_t^i, y_t^i, \{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t; \theta\right)$, we include the state variables of agent i twice, namely, they enter both as variables of agent i and as an element of the distribution. This repetition implies perfect collinearity in explanatory variables, so that the inverse problem is not well defined. Such a multicollinearity would break down a conventional numerical method which solves the inverse problem but neural networks can learn to ignore redundant colinear variables, as shown earlier. Thus, even though it is possible to design a transformation that avoids a repetition of variables, it would require cumbersome and costly permutations. Thus, it is easier to keep the repeated variables.

Our method differs from MMW (2018, 2019) in two important respects: one is that their analysis focuses on the model with Krusell and Smith (1998) aggregation and here, we solve model for which such aggregation is not feasible. Second, we have to approximate two neural networks with two different topologies because we need to approximate numerically the function q (in the model with savins through capital, the prices of capital and labor follow analytically). In the approximating function of bond price, we treat variables symmetrically,

$$q_t = Q\left(\left\{b_t^i, y_t^i\right\}_{i=1}^{\ell}, z_t; \theta\right),$$

The model with savings through bonds represent important computational challenge relative to the Krusell and Smith (1998) model with savings through capital. The equilibrium prices there are not possible to obtain from the individual data (like interest rate and wage) but we need to solve for the bond price that makes supply and demand coincide in terms of state variables. Furthermore, when we simulate the model with ℓ agents, we have ℓ data points for these agents but we have just one point for the price. So, we use batches of the size N, in other words, we repeat the simulation $N = 100$ times to get sufficient data for approximating q. Batches play a very important rule in our analysis. We use batches which replicate the simulation N times, so we have $N \times \ell$ individual data but just N data points for the prices.

Our DL algorithm constructs the solution by simulating a panel of heterogenous agents as follows:

DL algorithm for HANC model.

Step 1. Draw future idiosyncratic and aggregate shocks $\Sigma_1 = \left\{\varepsilon_{y,1}^i\right\}, \Sigma_2 = \left\{\varepsilon_{y,1}^i\right\}, \epsilon_{1,}, \epsilon_2$ to compute next period realization z_{t+1} and the distributions $\left\{y_{t+1}^i\right\}_{i=1}^{\ell}$ to find the next period economy state $s_{t+1} \equiv S\left(\left\{b_{t+1}^i, y_{t+1}^i\right\}_{i=1}^{\ell}\right)$.

Step 2. Use the decision functions $b_{t+2}^i\left(b_{t+1}^i, y_{t+1}^i, s_{t+1}, z_{t+1}\right)$ and bond price $q_{t+1}\left(s_{t+1}, z_{t+1}\right)$ to find the next period state bonds $\left\{b_{t+2}^i\right\}_{i=1}^{\ell}$ using the budget constraints and restore consumption $c_{t+1}^i\left(s_{t+1}, z_{t+1}\right)$.

Step 3. Form the objective function that minimizes the Euler-residual using our technique of two uncorrelated shocks, see Maliar, Maluar and Winant (2019):

$$E_{(B,Y,z,\Sigma_1,\Sigma_2,\epsilon_1,,\epsilon_2)}\left\{\left[\Psi\left(\frac{w^i-q_t\overline{b}}{c^i}-1,\eta^i-1\right)\right]^2\right.$$
$$+v_1\left[\beta\left[\frac{\beta\left[u'\left(c^{i'}\right)\right]}{qu'\left(c^i\right)}\bigg|_{\Sigma=\Sigma_1,\epsilon=\epsilon_1}\right]-\eta^i\right]\left[\beta\left[\frac{\beta\left[u'\left(c^{i'}\right)\right]}{qu'\left(c^i\right)}\bigg|_{\Sigma=\Sigma_1,\epsilon=\epsilon_1}\right]-\eta^i\right]\left.\right\}$$
$$+\frac{v_2}{\ell}\sum_{i=1}^{\ell}b^{i\prime}, \quad (4)$$

where $B=\left(b^1,...,b^\ell\right)$, $Y=\left(y^1,...,y^\ell\right)$ and z are random draws of the economy's state; $\Sigma_1=\left(\epsilon_1^1,...,\epsilon_1^\ell\right)$, $\Sigma_2=\left(\epsilon_2^1,...,\epsilon_2^\ell\right)$ are two uncorrelated random draws of individual productivity shocks; $\epsilon_{1,}$, ϵ_2 are two uncorrelated random draws for the aggregate productivity shocks, v_1, v_2 reflect different weights on different objectives; $w^i=b_t^i+y_t^i z_t-q_t\overline{b}$ is the wealth of agent, η^i is the Lagrange multiplier and Ψ^{FB} is a Fisher-Burmiester objective function

$$\Psi^{FB}\left(a,b\right)=a+b-\sqrt{a^2+b^2}. \quad (5)$$

The solution to $\Psi^{FB}\left(a,b\right)=0$ is equivalent to the minimum function $\Psi^{\min}\left(a,b\right)=\min\left\{a,b\right\}=0$ and leads to the solution $a\geq 0$, $b\geq 0$ and $ab=0$ but it is differentiable.

The last term in the expression (4) is new: it does not appear in the Maliar, Maliar and Winant (2019) analysis: it appears because our objective function include the goal of minimizing the aggregate amount of bonds (since the net supply of bonds must be zero). In the Krusell and Smith (1998) model with savings through capital, the price of capital follows in a closed form as a marginal product.

While our DL method is related to Krusell and Smith (1998) method, there are two critical differences between our and their analysis. They construct separately the individual decision rules and the aggregate laws of motion, namely, given the aggregate law of motion (ALM), they solve the individual problem and simulate the individual solutions to update the ALM and they alternate these steps until convergence is attained. The shortcoming of this approach is that we must solve the individual problem on some grid of aggregate state variables for a tensor product grid which quickly become intractable. We merged these two steps by using the stochastic gradient descent method which trains the model only in those problems where the solution lives. We solve the individual and aggregate problems simultaneously and only on those points that happens in simulation. As the machine is trained and the panel of agents evolves, the decision functions are refined jointly with the ergodic distribution.[2] This reduces enormously the state space to be explored.

Second, to make the model tractable, Krusell and Smith (1998) introduce a numerical approach that replaces the distributions of state variables with a finite set of their moments m_t–this approach proved to work remarkably well in a variety of models and applications. In the problems they study, the first moment could explain the ALM remarkably well. But it does not need to be the case in all problems and it will not be the case here. In contrast, we parameterize the decision function using the neural networks. We can take the moments to be the arguments of such functions as Krusell and Smith (1998) do but we can take as inputs the entire joint distribution of wealth and skills.

[2] Since random variables are autocorrelated in our model, the stochastic gradient is correlated over time and hence, it is biased. To reduce the bias, we train the model on cross-sections which are sufficiently separated in time instead of using all consecutive periods.

Neural network will perform model reduction and will find itself the set of statistics that represents best the underlying data. In particular, our analysis makes it possible to see how different is the solution under Krusell and Smith (1998) moments from the one with the actual state space.

Ergodic-set domain. When we solve the consumption-savings problem, we drew the state variables from a prespecified exogenous rectangular domain. However, for Krusell and Smith's (1998) model, we generate state variables by simulating the economy forward. Why don't we use a fixed rectangular domain now? This is because the volume of rectangular domain is huge in high-dimensional problems, and it is prohibitively expensive to attain an accurate approximation everywhere on such a huge domain. In fact, only an infinitesimally small fraction of rectangular domain is generally visited in equilibrium in high-dimensional applications; see Judd et al. (2011) for a discussion. We take advantage of that fact by solving the model on simulated series – we restrict attention to much smaller ergodic-set domain in which the solution "lives". This strategy helps us deal with the curse of dimensionality.

Model reduction. We solve the models with up to $\ell = 1,000$ of agents which corresponds to $2\ell + 1 = 2,001$ state variables. How can the DL method deal with such a huge state space? In addition to the ability to handle multicollinearity, neural networks possess another remarkable property: they automatically perform the model reduction. When we supply a large number of state variables to the input layer, the neural network condenses the information into 64 neurons of two hidden layers, making it more abstract and compact. In a sense, this procedure is similar to the photo compression or principal component transformation when a large set of variables is condensed into a smaller set of principal components without losing essential information; see Goodfellow et al. (2016) for a discussion of neural networks.

Krusell and Smith (1998) find one specific model reduction that works extremely well for their model, namely, they approximate the distribution of state variables with a finite set of moments. They find that in their model, just one moment – the mean of wealth distribution m_t – is a sufficient statistic for capturing all relevant information, reducing their state space to just four state variables (b_t^i, y_t^i, z_t, m_t).

If Krusell and Smith's (1998) construction is the most efficient representation of the state space, the neural network will possibly find this representation as an outcome of training. However, neural network automatically considers many other possible ways of extracting the information contained in the distribution $\{b_1^i, y_1^i\}_{i=1}^{\ell}$ and condensing it into a relatively small set of hidden layers. The output of the machine can look like moments or some other statistics – we will not always be able to understand how the machine handles the information in the hidden layers but this fact does not prevent us from using this miraculous technology in applications.

Calibration and computational details. In the benchmark case, we parameterize the model by $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ with a risk-aversion coefficient of $\gamma = 1$, and assuming $\beta = 0.96$; $\rho = 0.9$; $\sigma = 0.07$; $\rho_z = 0.9$; and $\sigma_z = 0.4\left(1 - \rho_z^2\right)^{1/2}$. We perform training using the ADAM method with the batch size of 10 and the learning rate of 0.001. We fix the number of iterations (which is also a simulation length) to be $b = 500,000$ but we perform the training only each 10th step, so the effective number of iterations is 50,000.

2.2 Numerical results

In Figure 4, we show approximated decision functions and simulated series under 3 layer neural network with 128, 64 and 32 neurons in the three layers respectively.

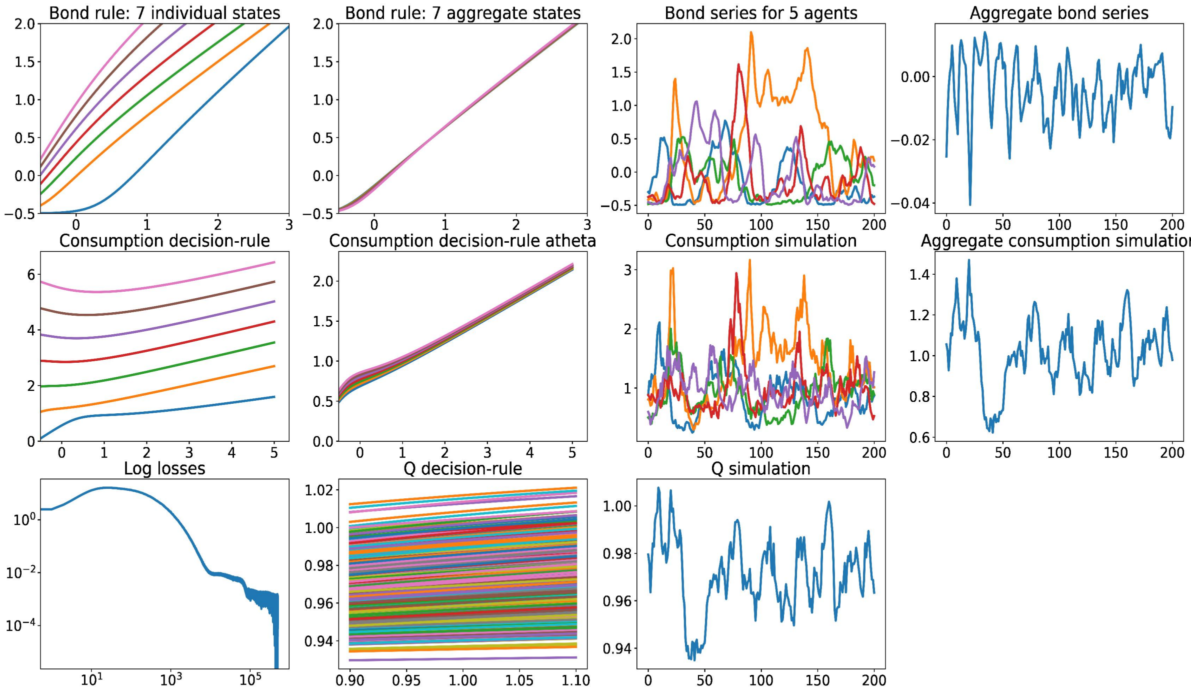

Figure 1. Neoclassical model: the solution constructed using the true state space.

We can see the kink in the decision rules due to the borrowing constraint. In panel 3, we draw bond price function $q(a)$ as a function of aggregate shock for number of simulated distributions $\{b_t^i, y_t^i\}$. We can see that such function is nonlinear. The individual bond holdings reach borrowing constraints which is 0.5 in this experiment. Aggregate bond holdings are close to zero (recall that they should be 0 in equilibrium deviate from zero by $\frac{\overline{b_t}}{\left|b_t^i\right|}$. In figure 2, next plot the same solution for the model in which the state space is replaced with the second moments. (The 1st moment of capital is zero, so we include 1st moment of shocks, 2nd moment of capital and 2nd moment of shocks and a cross term).

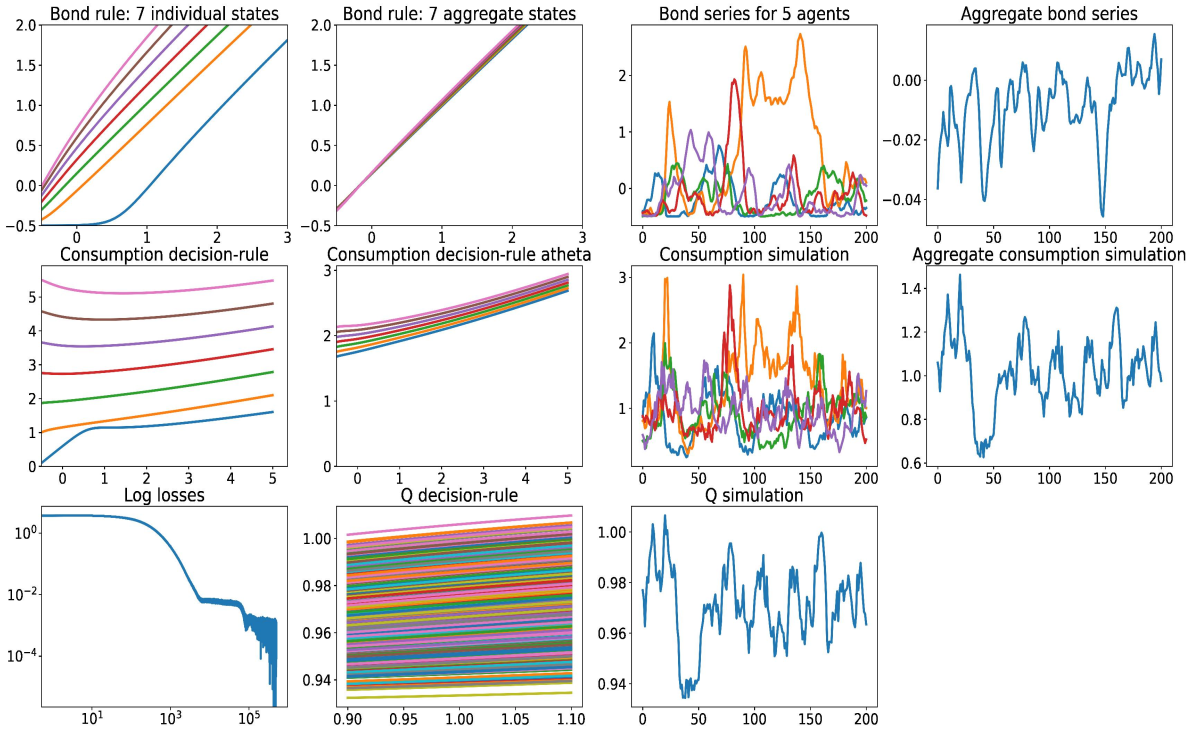

Figure 2. Neoclassical model: the solution constructed using the first and second moments.

The solutions in Figures 1 and 2 look relatively close. To get better assessment, we provide additional figure in which we compare the solutions under the true state space, under second moments and under third moments.

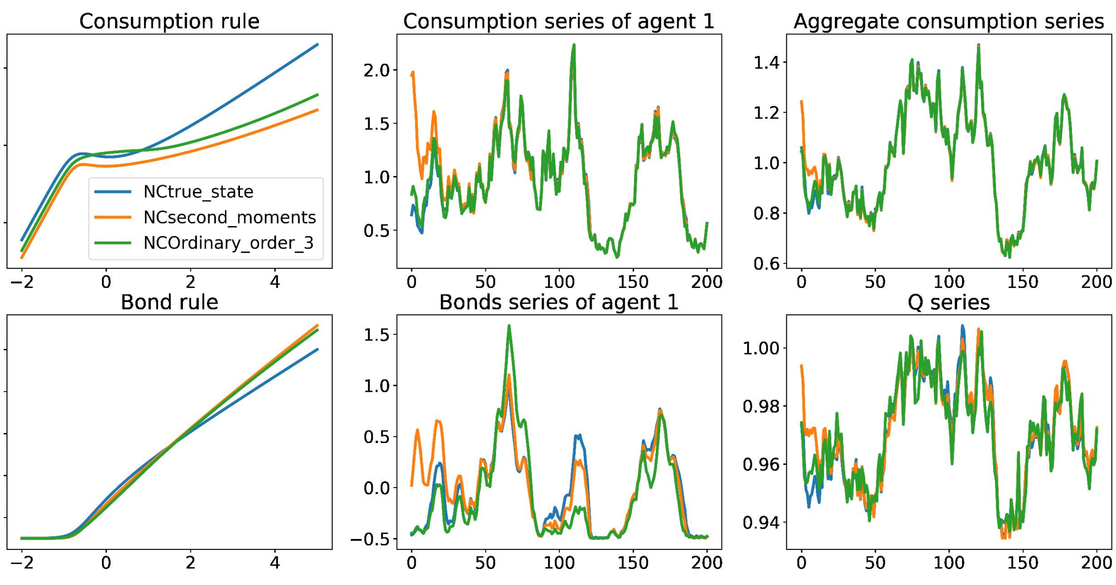

Figure 3. Neoclassical model: a comparison of 3 solutions constructed using different sets of state variables.

We now see that the solutions obtained look relatively close, in particular, the simulated series are relatively close in all 3 cases.

Can we use Krusell and Smith (1998) analysis? In the model with savings through capital, Krusell and Smith (1998) suggested how to produce accurate solutions to the HANC model.

Specifically, they parameterized the individual and aggregate decision functions by

$$\begin{aligned} c_t^i &= C\left(k_t^i, y_t^i, z_t, K_t\right), \\ \ln K_{t+1} &= b_0 + b_1 \ln z_t + b_2 \ln K_t, \end{aligned}$$

They find that R^2 of the second regression is of order 0,999999 meaning that we can describe accurately the future aggregates in terms of today's aggregates. We can implement the same approach in the economy with bonds by using

$$\begin{aligned} c_t^i &= C\left(b_t^i, y_t^i, z_t, m_t\right), \\ \ln m_{t+1} &= b_0 + b_1 \ln z_t + b_2 \ln m_t, \end{aligned}$$

where m_t is a selected set of moments (statistics). There are two reason why Krusell and Smith (1998) analysis can potentially produce inaccurate solutions in the model with bonds. First, it could be that the selected set of moments is not sufficient to capture the information about the distributions. However, our companion figure 3 shows that the second moment solutions obtained are not too far from the solution obtained under the actual decision rules.

So, we look at the second criteria, namely, the aggregate law of motion which is unnecessary under our DL method but indispensable under Krusell and Smith (1998) analysis. We assess the regression with 2 moments

$$\begin{aligned} c_t^i &= C\left(k_t^i, y_t^i, z_t, \overline{m}_1, s_1, s_2.s_{12}\right), \\ \ln \overline{m}_{1,t+1} &= b_0 + b_1 \ln z_t + b_2 \ln\left(\overline{m}_{1,t}, s_1, s_2.s_{12}\right), \; R^2 = 0.832 \\ \ln \overline{s}_{1,t+1} &= b_0 + b_1 \ln z_t + b_2 \ln\left(\overline{m}_1, s_1, s_2.s_{12}\right), \; R^2 = 0.979 \\ \ln s_{2t+1} &= b_0 + b_1 \ln z_t + b_2 \ln\left(\overline{m}_1, s_1, s_2.s_{12}\right), \; R^2 = 0.663 \\ \ln s_{12t+1} &= b_0 + b_1 \ln z_t + b_2 \ln\left(\overline{m}_1, s_1, s_2.s_{12}\right), \; R^2 = 0.526 \end{aligned}$$

Thus, we cannot run Krusell and Smith (1998) type of regressions because they explanatory power is low. (Again, our implementation of Krusell and Smith (1998) method avoids these regressions and this is why we obtain sufficiently accurate solutions even if we use moments).

In sum, like participants of 2008 JEDC we were unable to find statistics (moments) that can be used for the model with bonds in order to facilitate the method parallel to Krusell and Smith (1998). The problem appears to be that we cannot predict the future variables (in form of bonds, moments or even DL features) by using the current variables. We tried some other possibilities but none of them worked. We cannot rule out entirely that there is some ingenious reduced representation of the state variables that will allow us to forecast the future aggregates using just aggregate laws of motions but we were not able to find it even when we knew accurate solutions. Guessing it in general can be hard.

3 Heterogenous agent new Keynesian model (HANK) with savings through bonds

We consider a stylized HANK model with savings through bonds, see, e.g., Debortoli and Gali (2018). To economize on space, we limit ourselves to explaining how the HANK equations differ from those in Krusell and Smith (1998) model.

Agent i's maximization problem is as follows:

$$\max_{\left\{c_t^i, n_t^i, b_{t+1}^i\right\}_{t=0}^{\infty}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u\left(c_t^i, \ell_t^i\right) \right] \tag{6}$$

$$\text{s.t. } c_t^i + b_t^i = \frac{R_{t-1}}{\pi_t} b_{t-1}^i + w_t y_t^i n_t^i + \Pi_t, \tag{7}$$

$$b_{t+1}^i \geq \overline{b}, \tag{8}$$

where c_t^i is the amount of consumption of agent i in period t, $\ell_t^i + n_t^i = 1$, $w_t = z_t$, where R_{t-1} is nominal interest rate; π_t is inflation; and Π_t is the profit distributed equally across the agents. The problem (6)-(8) differs from the HANC model in how the wage w_t and rate of return on bond $\frac{R_{t-1}}{\pi_t}$ are determined.

Output depends on aggregate efficiency labor H_t, on exogenous aggregate productivity shock z_t and on a distortion from inflation $\triangle_t^p$:

$$Y_t = z_t H_t \triangle_t^p$$

Here, the real wage is not equal to the marginal product but is determined by marginal cost m_t according to $w_t = z_t m_t$. In turn, marginal cost m_t is determined by the Phillips curve

$$m_t = \frac{1}{\mu} + \frac{1}{\mu\kappa} \left\{ \pi_t \left(\pi_t - \pi^*\right) - E_t \left[\Lambda_{t+1} \pi_{t+1} \left(\pi_{t+1} - \pi^*\right) \frac{Y_{t+1} \triangle_t^p}{Y_t \triangle_{t+1}^p} \right] \right\}$$

where where $c_t^i = \left[\frac{\psi\left(1-\ell_t^i\right)^{-\eta}}{w_t} \right]^{-\frac{1}{\gamma}}$ and $\Lambda_{t+l} \equiv \beta^l \frac{E_t\left[\int_{j\in U_{t+l}^{illiq}} \left(c_{t+l}^i\right)^{-\gamma} di\right]}{\int_{j\in U_t^{illiq}} \left(c_t^i\right)^{-\gamma} dj}$ with U_t^{illiq} being the set of agents at time t who are not against their illiquid asset borrowing constraint an μ and κ are the parameters.

The profit is given by the difference between output and labor income $\Pi_t \equiv Y_t - w_t H_t$.

The distortion is determined by how much the inflation π_t differs from its target level π^*:

$$\triangle_t^p \equiv 1 - \frac{1}{2\kappa\left(\mu - 1\right)} \left(\pi_t - \pi^*\right)^2,$$

where κ and μ are the parameters.

Finally, the nominal interest rate is determined by the Taylor rule

$$R_t \equiv R_* \left(\frac{\pi_t}{\pi_*}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_*}\right)^{\phi_y}, \tag{9}$$

where R_t and R_* are the gross nominal interest rate at t and its long-run value, respectively; π_* is the target inflation; where ϕ_π and ϕ_y are the parameters in the Taylor rule. We assume that the interest rate satisfies the zero or effective lower bound ZLB or ELB respectively.

3.1 The solution procedure for HANK with bonds

We solve the HANK model by using a similar DL solution method (to save on space, we do not describe it in details but explain how it differs from the DL algorithm used to solve the HANC model). We parameterize with neural networks the share of consumption and labor decision functions both of which are transformed using a sigmoid function into an interval $[0, 1]$. Furthermore, we parameterize the Lagrange multiplier which is transformed to nonnegative range using an exponential function. Finally, we parameterize two aggregate variables m_t, π_t. As in HANC, the model has $2\ell+1$ state variables $\{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t$.

We parameterize with neural network

$$\text{aggregate variables } m_t,\, \pi_t \text{ in terms of } \left(b_t^i, y_t^i, \{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t\right)$$
$$\text{individual variable } c_t^i,\, \ell_t^i \text{ in terms of } \left(\{b_t^i, y_t^i\}_{i=1}^{\ell}, z_t\right)$$

We perform the following sequence of computations:

1. Given m_t, find
$$w_t = m_t z_t$$

2. Find profits
$$\Pi_t = Y_t - w_t H_t$$
Alternatively, $\Pi_t = \frac{Y_t}{\triangle_t^p}(1-m_t) - \frac{1}{2\kappa(\mu-1)}(\pi_t - \pi^*)^2 \frac{Y_t}{\triangle_t^p}$. Also, we have $Y_t \frac{m_t}{\triangle_t^p} = w_t H_t$.

3. Given π_t, find
$$\triangle_t^p \equiv 1 - \frac{1}{2\kappa(\mu-1)}(\pi_t - \pi^*)^2$$

4. Find
$$Y_t = z_t H_t \triangle_t^p$$

5. Find the nominal interest rate from the Taylor rule
$$R_t \equiv R_* \left(\frac{\pi_t}{\pi_*}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_*}\right)^{\phi_y}.$$

We next formulate the loss functions for training:

1. The Phillips curve provides the condition that identifies the marginal cost m_t
$$m_t\mu - \frac{1}{\kappa}\left\{\pi_t(\pi_t - \pi^*) - E_t\left[\widetilde{\Lambda}_{t,t+l}\pi_{t+1}(\pi_{t+1} - \pi^*)\frac{Y_{t+1}\triangle_t^p}{Y_t\triangle_{t+1}^p}\right]\right\} = 1$$

2. The equation that identifies π_t is the market clearing condition
$$\int_0^1 b_t(j)\,di = 0$$

3. The intratemporal first-order condition identifies labor
$$n_t^i = 1 - \left[\frac{c_i^{-\gamma} w_t y_t^i}{B}\right]^{-1/\eta}$$

4. The intertemporal first-order condition identifies the consumption and savins choices

$$\Psi^{FB}\left\{\frac{w^i}{c^i}-1,\eta^i-1\right\}=0$$

$$\beta R_t E_t\left[\frac{\left(c^i_{t+1}\right)^{-\gamma}}{\left(c^i_t\right)^{-\gamma}\pi_{t+1}}\right]-h^i_t=0$$

$$c^i_t+b^i_t=\frac{R_{t-1}}{\pi_t}b^i_{t-1}+w_t y^i_t n^i_t+\Pi_t\equiv w^i_t.$$

3.2 Calibration

The following values and targets are from Debortoli and Gali (2018)

Parameter	Description	Target/Source
$\gamma=1$	Risk aversion	standard
$\eta=1$	Labor supply elasticity	standard
$\beta=.975$	Discount factor	
$d=.1$	Depreciation rate	
$\alpha=\frac{1}{3}$	Capital share	standard
$\mu=\frac{10}{9}-$	Elasticity of substitution among goods	profits share of 10%
$\kappa=\frac{9}{105}$	Price adjustment cost	avg. price duration of 1 year
$\bar{b}=-0.4$	Liquid asset borrowing constraint	75% of people have liquid assets
$\rho_z=0.9777$	Persistence of idiosyncratic shock	persistence of annual wage =.92
$\sigma_z=0.1928$	Standard deviation of idiosyncratic shock	standard deviation of annual wage=0.7
$\rho_y=0.9$	Persistnece of TFP shock	standard
$\sigma_y=0.016$	Standard deviation of TFP shock	standard

We set at $\pi^*=1.02$. To have steady state in the bond Euler equation, we need

$$R=\frac{\pi^*}{\beta}=1.05699,$$

$$\triangle\equiv 1-\frac{1}{2\kappa\left(\mu-1\right)}\left(\pi-\pi^*\right)^2=1$$

$$m=\frac{1}{\mu}=0.899999$$

$$\begin{aligned} w &= mc \\ Y &= \triangle \\ C &= Y \\ H &= 1 \end{aligned}$$

Hence, the utility parameter B is

$$B=\frac{C^{-\gamma}w}{\left(3-H\right)^{-\eta}}$$

Finally, the profit is

$$\Pi=Y-wH=1-mc$$

3.3 Results

We first show the decision rules for the key model variables as well as simulated series in Figure 4.

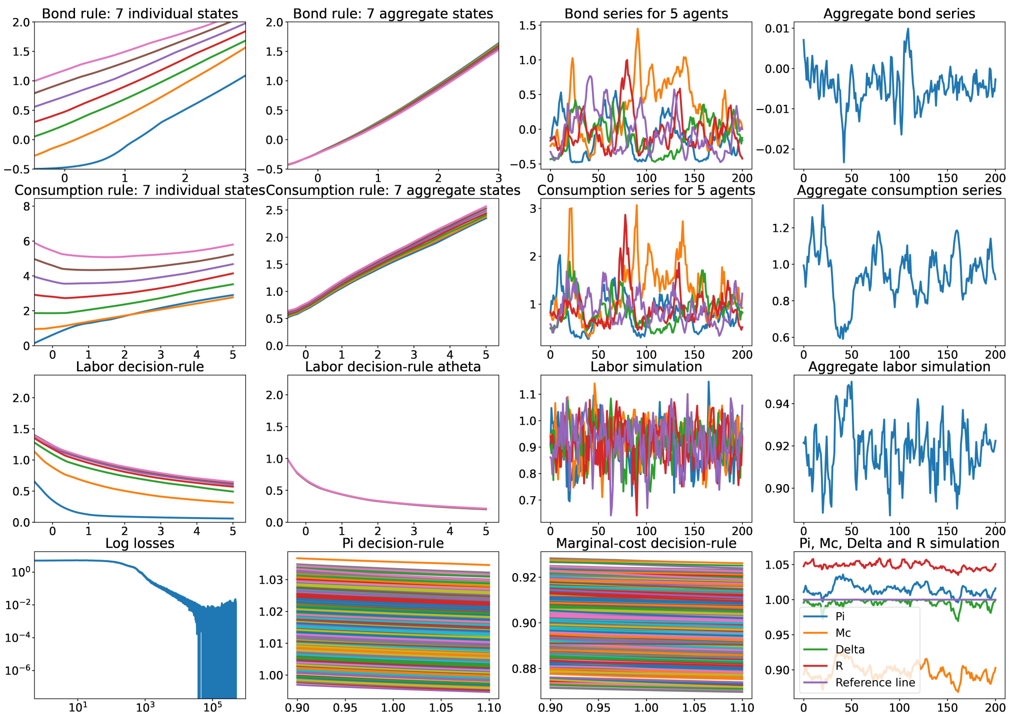

Figure 4. New Keynesian model: the solution constructed using the true state space.

We see that the individual decision rules for bonds have the kink for low productivity state, we observe the corresponding kink in the consumption function. The labor decision rules are decreasing in wealth as expected. For the aggregate state, we see some differences in the decision rules but such differences are much smaller. The time series for individual bonds, consumption and labor fluctuate much more than the corresponding aggregates. The mean bond is close to zero as implied by equilibrium. Unit free loss function (errors in the model equations) are of order $10^{-3} - 10^{-4}$. Finally, we plot the decision rules for inflation and marginal cost depending on the aggregate shock under different realizations of idiosyncratic uncertainty as well as time series for these variables.

In Figure 5, we show the solution that we constructed using the second moments of the bond and income distributions as the state space. We can see that the decision rules look visibly different from those in Figure 3, so we do not make formal comparison analysis as we did for the growth model.

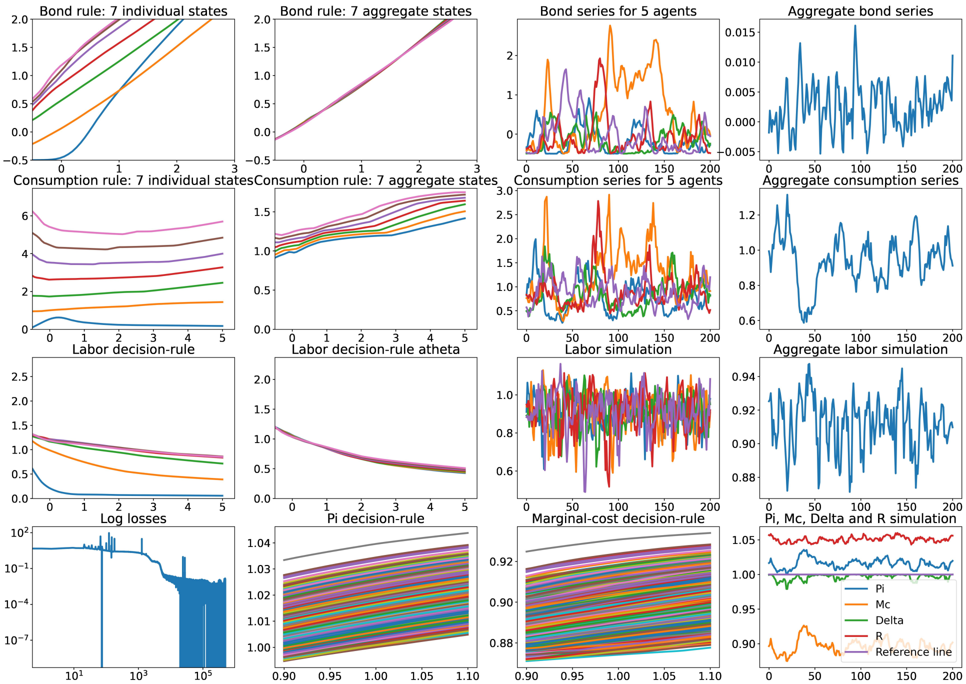

Figure 5. New Keynesian model: the solution constructed using two moments.

We tried to use higher moments and other statistics but we could not make the resulting solutions sufficiently close to those produces under the true state space. Thus, it is more difficult for Krusell and Smith (1998) method to approximate the solutions because the aggregate laws do not always available.

4 Concluding comments

Krusell and Smith (1998) method provided a simple and elegant method of solving heterogeneous agent economies that made a number of models tractable that are intractable otherwise. This method was fruitful in helping to resolve consumer and producer models.

However, some models in the literature cannot be solved with Krusell and Smith (1998) method or it is unknown how they can be solved. with such method. We have shown one such model but there are others. Also, there are debates in the literature whether or not bounded rationality (reduced state space) can always give the right equilibrium.

We offered an alternative DL framework that works with the true state space even if it is such state space has hundred of variables such as the HANC and HANK models. We were able to solve the models that do not allow for Krusell and Smith (1998) aggregation and we show that it is hard to find a set of aggregate statistics (moments) that can be used to characterize the aggregate economy. The complications are related to the fact that the mean of the bond distribution is zero.

Recent literature moves toward introducing distributions in macroeconomic models and it is not clear whether or not KS method will work in all such models and how accurate and reliable it can be. The alternative method proposed here is simple, reliably and useful.

References

[1] Ahn, SeHyoun, G. Kaplan, B. Moll, T. Winberry and C. Wolf, 2018. "When Inequality Matters for Macro and Macro Matters for Inequality," NBER Macroeconomics Annual, University of Chicago Press, vol. 32(1), pages 1-75.

[2] Azinovic, M., G. Luca and S. Scheidegger (2020). Deep equilibrium nets. SSRN: https://ssrn.com/abstract=3393482.

[3] Auclert, Adrien. 2019. "Monetary Policy and the Redistribution Channel." American Economic Review, 109 (6): 2333-67.

[4] Bayer, C. and R. Luetticke, (2018). Solving heterogeneous agent models in discrete time with many idiosyncratic states by perturbation methods. Manuscript.

[5] Bayer, Christian, R. Luetticke, L. Pham-Dao and V. Tjaden, 2019. "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," Econometrica, Econometric Society, vol. 87(1), pages 255-290, January.

[6] Boppart, T., P. Krusell and K. Mitman, (2018). The performance of policy rules in heterogeneous-agent models with aggregate shocks. Journal of Economic Dynamics and Control 89(C), 68-92.

[7] Debortoli, Davide and Jordi Gali, 2018. "Monetary policy with heterogeneous agents: insights from TANK models." Manuscript.

[8] Den Haan, W., (1997). Solving dynamic models with aggregate shocks and heterogeneous agents. Macroeconomic Dynamics 1(02), 355-386.

[9] Den Haan, W., (2010), Comparison of solutions to the incomplete markets model with aggregate uncertainty. Journal of Economic Dynamics and Control 34, 4–27.

[10] Duarte, V. (2018). Machine learning for continuous-time economics. Manuscript.

[11] Duffy, J. and P. McNelis, (2001). Approximating and simulating the real business cycle model: parameterized expectations, neural networks, and the genetic algorithm, Journal of Economic Dynamics and Control 25(9), 1273-1303.

[12] Fernández-Villaverde, J., S. Hurtado, and G. Nuño, (2018). Financial frictions and the wealth distribution. Manuscript.

[13] Gaspar, J. and K. L. Judd, (1997). Solving large-scale rational-expectations models. Macroeconomic Dynamics 1, 45-75.

[14] Goodfellow, I., Y. Bengio, and A. Courville, (2016). Deep learning. Massachusetts Institute Technology Press.

[15] Iskhakov, F., Jørgensen, T., Rust, J. and B. Schjerning (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. Quantitative Economics, 8(2):317–365 (lead article).

[16] Iskhakov, F. and M. Keane (2020). Effects of Taxes and Safety Net Pensions on Life-Cycle Labor Supply, Savings and Human Capital: the Case of Australia, Journal of Econometrics, 2020.

[17] Jirniy, A. and V. Lepetyuk, (2012). A reinforcement learning approach to solving incomplete market models with aggregate uncertainty. Manuscript.

[18] Judd, K., (1998). Numerical Methods in Economics. MIT Press, Cambridge, MA.

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

[20] Lepetyuk, V., L. Maliar and S. Maliar, (2020). When the U.S. catches a cold, Canada sneezes: a lower-bound tale told by deep learning. Journal of Economic Dynamics and Control, forthcoming.

[21] Kaplan, Greg, G. L. Violante and J. Weidner, 2014. "The Wealthy Hand-to-Mouth," Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 45(1 (Spring), pages 77-153.

[22] Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." American Economic Review, 108 (3): 697-743.

[23] Krusell, Per and Anthony Smith, 1998. "Income and Wealth Heterogeneity in the Macroeconomy" Journal of Political Economy 106, no. 5, 868-896.

[24] Krusell, Per and Anthony Smith, 1998. "Income and Wealth Heterogeneity, Portfolio Choice and Equilibrium Asset Returns" Macroeconomic Dynamics,1, 1997, 387–422.

[25] Maliar, L. and S. Maliar, (2014). Numerical methods for large scale dynamic economic models in: Schmedders, K. and K. Judd (Eds.), Handbook of Computational Economics, Volume 3, Chapter 7, Amsterdam: Elsevier Science.

[26] Maliar, L., Maliar, S., and P. Winant, (2018). Deep learning for solving dynamic economic models. CEF-2018 presentation, https://www.youtube.com/watch?v=u7_CdytTEe8&feature=youtu.be.

[27] Maliar, Lilia, Serguei Maliar, and Pablo Winant. 2019. "Will Artificial Intelligence Replace Computational Economists Any Time Soon?" CEPR working DP 14024.

[28] McNelis, P. (2005). Neural networks in finance: gaining predictive edge in the market. Elsevier Academic Press.

[29] Powell, W. (2010). Approximate dynamic programming. Wiley, A. John Wiley & Sons.

[30] Sutton R. and A. Barto (2018). Reinforcement learning: an introduction. The MIT Press, Cambridge, Massachusetts, London, England.

[31] Villa, A. and V. Valaitis (2019). Machine learning projection methods for macro-finance models, Manuscript.