Machine Learning Methods for Solving and Estimating Heterogeneous Agent Models with Aggregate Shocks

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Pablo Guerrón

Boston College

Models with heterogeneous agents (firms, households, countries) and aggregate shocks are more interesting than ever. Issues such as understanding the origins and evolution of income and wealth inequality, investment decisions, industry dynamics, international trade and migration, and the effects of financial frictions depend crucially on our ability to handle these models formally.

However, solving and estimating heterogeneous agent models with aggregate shocks is challenging. In terms of solution methods, most of the literature has followed variations of the algorithm proposed by Krusell and Smith (1998).[1] While this algorithm provides accurate solutions to many problems of interest, it faces two problems. First, the researcher must select the moments of the distribution of endogenous state variables that best summarize the information that agents need to compute equilibrium variables. Second, the researcher needs to specify a law of motion from these moments and estimate it from simulated data. Both tasks are often too unstructured. As soon as we leave the familiar territory of variations of the neoclassical growth model with one or two endogenous aggregate state variables, economists have limited guidance about which moments summarize the distribution of interest accurately and about what is the right shape of the law of motion. Moreover, the basic Krusell-Smith algorithm does not scale well regarding the number of state variables and suffers from an acute case of the "curse of dimensionality."

Regarding the estimation of the model, a researcher could employ techniques such as the particle filter (Fernández-Villaverde and Rubio-Ramírez, 2007) to evaluate the likelihood of the model. However, the particle filter is computationally costly, especially if nested with a Krusell and Smith (1998)-type algorithm.

To address these challenges, we propose the use of machine learning techniques. More concretely, we will discuss in detail how to approximate the law of motion for the moments of the distribution using a neural network and how to evaluate the likelihood of the model. Because of space constraints, we will skip the discussion of how to employ random forest trees to select moments of the distribution to feed into the law of motion. These techniques open the door to the solution and estimation of a large class of models with heterogeneous agents and aggregate shocks that were, up to this moment, intractable. Furthermore, the techniques are particularly suitable for massive parallelization employing either graphics processing units (GPUs) or the newly available tensor processing units (TPUs).

Concerning the law of motion of the moments of the distribution, we employ a nonlinear approximation technique based on neural networks. The network is trained in simulated data until convergence. We let the shape of the network be determined endogenously and not imposed ex-ante. Our approach has four crucial advantages. First, the *universal approximation theorem* (Hornik, Stinchcombe, and White, 1989; Cybenko, 1989) states that a neural network with at least one hidden layer can approximate any Borel measurable function mapping finite-dimensional spaces arbitrarily well. In particular, the theorem does not require that the approximated function be differentiable and can handle cases with kinks and occasionally binding constraints. Second, the neural network coefficients can be efficiently estimated using gradient descent methods and back-propagation. These algorithms allow for a more straightforward coding and shorter implementation time than other approaches. Third, neural networks are more economical, for middle and high dimensions, than other approximators. More concretely, Barron (1993) shows that a one-layer neural network achieves, for functions on the first moment of the magnitude distribution of the Fourier transform, integrated square errors of order $O(1/n)$, where n is the number of nodes. In

[1] See Den Haan (1996); Den Haan (1997), Algan, Allais, and Den Haan (2008), Reiter (2009, 2010), Den Haan and Rendahl (2010), Maliar, Maliar, and Valli (2010), Sager (2014), and Pröhl (2015) (a recent survey of the field can be found in Algan, Allais, Den Haan, and Rendahl, 2014).

comparison, for series approximations (polynomials, spline, and trigonometric expansions), the integrated square error is of order $O(1/(n^{2/d})$, where d is the dimensions of the function to be approximated. In other words: the "curse of dimensionality" does not apply to neural networks that approximate functions of the broad class considered by Barron (1993). This advantage is fundamental because even going to $d = 3$ or $d = 4$ saturates alternatives for nonlinear approximators such as Chebyshev polynomials. Fourth, neural networks extrapolate outstandingly. This is, in practice, key for models with complex ergodic distribution of endogenous equilibrium variables. Neural networks have well-behaved shapes outside their training areas. In contrast, Chebyshev polynomials (or other series) more often than not display explosive behaviors outside the fitted area.[2]

This approximation can be applied both to discrete- and continuous-time models. In our first set of applications, we focus on continuous time because it enjoys several numerical advantages with respect to discrete time. First, continuous time naturally generates sparsity in the matrices characterizing the transition probabilities of the discretized stochastic processes. Intuitively, continuously moving state variables like wealth only drift an infinitesimal amount in an infinitesimal unit of time. Thus, in an approximation that discretizes the state space, households reach only states that directly neighbor the current state. Second, the optimality characterizing consumption has a simpler structure than in discrete time, since it equates the marginal utility of consumption (or profits) to the partial derivative of the value function with respect to wealth (or assets). Third, it is easier to capture occasionally binding constraints in continuous time than in discrete time as the optimality conditions hold with equality everywhere in the interior of the state space. Fourth, the dynamics of the cross-sectional wealth distribution are characterized by a simple Kolmogorov forward equation. The discretization of this equation yields an efficient way to simulate a time series of the cross-sectional distribution (although this can also be performed in discrete time, as in Ríos-Rull 1997, Reiter 2009, and Young 2010, at some additional cost).

With respect to estimation, we exploit the fact that the adjoint of the infinitesimal generator generated by the solution of the Hamilton-Jacobi-Bellman equation that encodes the dynamic optimization problem of the agents turns out to be the likelihood function of the model. This is remarkable, since it means that, in practice, evaluating the likelihood function –given some data– amounts to transposing and inverting a sparse matrix that has already been computed. Once we have the likelihood function, we can either maximize it or implement the Bayesian approach by nesting the procedure in a McMc.

We have already solved and estimated a version of the model by Brunnermeier and Sannikov (2014) augmented with heterogeneous agents. This model is particularly interesting, both in its implications and in terms of methodology. Regarding its implications, this model teaches us much about the role that the leverage level plays in aggregate fluctuations. Regarding methodology, the model's solution is highly nonlinear and, thus, a naive application of the Krusell-Smith algorithms fails at delivering an accurate solution. We have already computed the equilibrium dynamics of the model, evaluated the likelihood of the model, and presented our results in several talks.

But much remains to be done regarding the fine tuning of the methodology and plenty of other applications are possible. Among those, we already have preliminary work on models with menu-cost pricing, models with heterogeneous firms and bumpy investment, and models with technological innovation.

[2]Similarly, approaches, such as Smolyak interpolation, that alleviate the "curse of dimensionality" in standard problems are harder to apply here precisely because we know little ex-ante about the shapes of the ergodic distribution. Neural networks are more resilient to sparse initial information regarding such shapes.

References

Algan, Y., O. Allais, and W. J. Den Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32(3), 875–908.

Algan, Y., O. Allais, W. J. Den Haan, and P. Rendahl (2014): "Solving and Simulating Models with Heterogeneous Agents and Aggregate Uncertainty," in *Handbook of Computational Economics*, ed. by K. Schmedders, and K. L. Judd, vol. 3, pp. 277 – 324. Elsevier.

Barron, A. R. (1993): "Universal Approximation Bounds for Superpositions of a Sigmoidal Function," *IEEE Transactions on Information Theory*, 39(3), 930–945.

Brunnermeier, M. K., and Y. Sannikov (2014): "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104(2), 379–421.

Cybenko, G. (1989): "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, 2(4), 303–314.

Den Haan, W. J. (1996): "Heterogeneity, Aggregate Uncertainty, and the Short-Term Interest Rate," *Journal of Business & Economic Statistics*, 14(4), 399–411.

Den Haan, W. J. (1997): "Solving Dynamic Models with Aggregate Shocks and Heterogeneous Agents," *Macroeconomic Dynamics*, 1(02), 355–386.

Den Haan, W. J., and P. Rendahl (2010): "Solving the Incomplete Markets Model with Aggregate Uncertainty Using Explicit Aggregation," *Journal of Economic Dynamics and Control*, 34(1), 69–78.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087.

Hornik, K., M. Stinchcombe, and H. White (1989): "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks*, 2(5), 359 – 366.

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

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.

Pröhl, E. (2015): "Approximating Equilibria with Ex-Post Heterogeneity and Aggregate Risk," Research Paper 17-63, Swiss Finance Institute.

Reiter, M. (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, 33(3), 649–665.

——— (2010): "Solving the Incomplete Markets Model with Aggregate Uncertainty by Backward Induction," *Journal of Economic Dynamics and Control*, 34(1), 28–35.

Ríos-Rull, J.-V. (1997): "Computation of Equilibria in Heterogeneous Agent Models," Staff Report 231, Federal Reserve Bank of Minneapolis.

Sager, E. (2014): "Solving the Incomplete Markets Model With Aggregate Uncertainty: The Method of Mixtures," Bureau of Labor Statistics, Price Research Division.

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.