The Econometrics Journal

Econometrics Journal (2020), volume **00**, pp. 1–44.
doi: 10.1093/ectj/utaa019

Machine learning and structural econometrics: contrasts and synergies

Fedor Iskhakov[*], John Rust[†] and Bertel Schjerning[‡]

[*]*Australian National University, 1018 HW Arndt Bld., Kingsley Place, Acton ACT 2601, Australia.*
Email: fedor.iskhakov@anu.edu.au

[†]*Georgetown University, 583 Intercultural Center, Washington DC 20057-1036, U.S.A.*
Email: jrust@editorialexpress.com

[‡]*University of Copenhagen, Øster Farimagsgade 5, Bld. 25, 1353 Copenhagen K, Denmark.*
Email: bertel.schjerning@econ.ku.dk

First version received: 31 March 2020; final version accepted: 30 June 2020.

Summary: We contrast machine learning (ML) and structural econometrics (SE), focusing on areas where ML can advance the goals of SE. Our views have been informed and inspired by the contributions to this special issue and by papers presented at the second conference on dynamic structural econometrics at the University of Copenhagen in 2018, 'Methodology and Applications of Structural Dynamic Models and Machine Learning'. ML offers a promising class of techniques that can significantly extend the set of questions we can analyse in SE. The scope, relevance and impact of empirical work in SE can be improved by following the lead of ML in questioning and relaxing the assumption of unbounded rationality. For the foreseeable future, however, ML is unlikely to replace the essential role of human creativity and knowledge in model building and inference, particularly with respect to the key goal of SE, counterfactual prediction.

Keywords: *machine learning*, *structural econometrics*, *curse of dimensionality*, *bounded rationality*, *counterfactual predictions*.

The best way for machines to learn is to stand on the shoulders, not the toes, of their human teachers.

Daniel McFadden

1. INTRODUCTION

In this final article, we take the opportunity to offer our own perspectives on areas where structural econometrics (SE) can benefit from recent progress in machine learning (ML). Our views have been informed and inspired by the contributions to this special issue, particularly the lead article, Igami (2020), which draws very interesting analogies between SE and ML in the context of board games, an area where ML has had a number of astounding successes.

Although many parallels exist between the two literatures, it is important to point out the contrasts, including differences in the overall goals of ML and SE. Our views complement and extend Igami's interesting comparison of these two exciting and rapidly evolving scientific literatures. Where possible, we follow his lead by focusing on board games to help make the discussion concrete. We do not attempt to survey the whole of ML, and so our discussion is highly selective and focused on the aspects of ML that are most closely connected to what we do in SE.

We conclude that ML offers a promising class of techniques and algorithms that can significantly extend the set of questions we can analyse in SE. But ML is not a panacea and does not threaten to supersede SE, nor to make the essential human role in model building irrelevant. The practical orientation of the ML literature serves as a helpful role model for SE, and we see major opportunities for empirical work that focuses on improving decision making in the small as well as in the large. In particular, opportunities exist to improve the scope, relevance and impact of empirical work by following the lead of ML by questioning and relaxing the predominant paradigm in SE and most of economics: namely the assumption that individuals and firms have unbounded rationality.

2. HOW DOES ML DIFFER FROM SE AND WILL IT PUT US OUT OF WORK?

It is natural to start with brief definitions of 'machine learning' and 'structural econometrics', even though we hope that most readers will have some familiarity with one or both of these literatures. ML can be defined as the scientific study of the algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions and improve automatically through experience. ML is closely related to the field of artificial intelligence (AI), see Bishop (2006). SE is a branch of econometrics focused on the development of methods for the inference and testing of economic theories and models of individual, firm, and organizational behaviour.

ML and SE share a common interest in prediction and decision making, but the goal of ML is to enable *computers* to do these tasks whereas SE is focused on how *humans* do them. Thus ML is more *practically oriented*, in trying to automate tasks that previously only humans could do well, whereas SE, for reasons we discuss below, has been more *academically oriented*, in trying to understand and model human economic behaviour.[1]

The literature on ML has three broad categories: (a) supervised learning, (b) unsupervised learning, and (c) reinforcement learning, which differ in their assumptions about whether there is a 'teacher' that can help 'train' the computer to predict and recognize patterns in data, and in how to use feedback from the environment to improve performance over time. Under this taxonomy, SE is most closely related to supervised learning, since the goal is to build a succession of models that improve our understanding and ability to predict the behaviour of human decision makers and firms/organizations—our 'teachers'.

ML is associated with a collection of flexible statistical methods for nonlinear regression and the approximation of probability models such as regression trees and neural networks that can be viewed as *sieves* or expanding parametric families that enable nonparametric estimation, similar to the well-known kernel density and local linear modelling methods that have long been used in

[1] Another important goal of ML is *pattern recognition.* Murphy (2012) notes that 'The goal of machine learning is to develop methods that can automatically detect patterns in the data, and then to use the uncovered patterns to predict future data or other outcomes of interest. Machine learning is, thus, closely related to the fields of statistics and data mining, but differs slightly in its emphasis and terminology.'

statistics and econometrics. ML is also associated with model selection methods such as LASSO or those that help select a parsimonious regression model in problems where there can be vastly more potential regressors than observations. These methods, developed in statistics, are now commonly used in ML.

While prediction is certainly one of the goals of statistics and econometrics, the more important goal is *inference*, which 'is concerned with how data supporting general claims about the world deduced from economic models should be calculated' (Lindley, 2017). SE is a subfield of statistics and econometrics that is focused on testing and improving theories and models of economic behaviour. As its name suggests, the goal of SE is to make inferences about the *underlying structure,* which is given by often somewhat abstract objects not directly observable or measurable. Structural objects include the preferences/rewards and welfare of decision makers and their beliefs about the stochastic law of motion of observed and unobserved variables in their environment.

Structural estimation can be viewed as a branch of econometrics with objectives similar to in the literature on *revealed preference*, which attempts to infer the preferences of consumers from their choices. The origins of SE can be traced back to the work of Frisch (1926) and Cobb and Douglas (1928) and to subsequent work at the Cowles Foundation.[2] Since then, SE has evolved in many different directions, including the work on static discrete choice by McFadden (1974) and its extensions to dynamic discrete choice by Wolpin (1984), Rust (1987), Hotz and Miller (1993) and others, as well as applications to dynamic models of economies and games, especially in industrial organization (see, for example Ericson and Pakes, 1995).

SE is fundamentally a process of *human learning*. Even though methods of ML may be quite useful towards furthering that goal, SE is essentially a search for knowledge and understanding, and not just an ability to make better predictions. An important practical goal of SE does involve prediction, but more specifically *counterfactual prediction*—also known as *policy forecasting.* The hope is that by attaining a better understanding of the underlying structure of economic agents, markets, and institutions, we will be better able to predict their endogenous behavioural responses to changes in policies and the environment. With sufficiently good models, economists can design and advise on better policies that affect agents, markets, and institutions and can lead to improved economic outcomes and welfare.[3]

Policy forecasting is an eminently practical application and goal of SE, but it is unclear that our models and methods are up to the task of providing *credible* counterfactual predictions 'in the large'; that is, for some of the most important large-scale policy issues. To take a current example: we are unaware of any micro-founded SE model that is sufficiently realistic and detailed to provide credible guidance on policies to help government policy-makers best deal with the economic and health consequences of the COVID-19 pandemic. Great uncertainty exists concerning the transmission rate and the ultimate fraction of the population that will be infected, and there is debate over policies that might at least 'flatten the curve' of infections so as not to overwhelm the healthcare sector. Yet policies such as mandatory shutdowns and quarantines will have a tremendous economic toll, and contribute to panic that is already reflected in large drops in the

[2] As Rust (2014, p. 821) notes, 'This terminology, and the antecedents of modern structural macro- and microeconometrics (including the simultaneous equations model and some of the earliest work on endogeneity and instrumental variables) can be traced to Cowles, and particularly to Koopmans, and other founding fathers such as Jacob Marschak and Trygve Haavelmo. In fact, the raison d'être of the Cowles Foundation is to promote a tighter integration between theory and measurement in economics.' you understand?

[3] SE has also proved to be essential in comparative institutional design, for example in auctions; see Hubbard et al. (2020b); Cho et al. (2019); Hubbard et al. (2020a). However, we do not have space here to cover the large literature on structural estimation of auctions covered in Hong et al. (2006) or the theoretical and applied literature on mechanism and market design.

stock market as well as in record levels of layoffs. In our highly economically and financially interconnected world there is an inevitable degree of fragility, and these stresses could result in a financial contagion and collapse that could push the world economy into a severe recession.

The first respondents to the current pandemic include calibrated macro-models developed on very short notice, such as Eichenbaum et al. (2020). They graft the SIR (Susceptible, Infected, Recovered) epidemiological model onto a traditional dynamic macroeconomic model. Their analysis leads them to conclude that 'it is optimal to introduce large-scale containment measures that result in a sharp, sustained drop in aggregate output. This optimal containment policy saves roughly half a million lives in the U.S.'

While we do not have any reason to disagree with this conclusion, it is unclear how sensitive their policy advice is to underlying assumptions, particularly on a key parameter R_0, known as the 'basic reproductive number', namely the total number of other people an infected person will transmit the virus to over their remaining lifetime. R_0 is not a structural parameter but an endogenous, policy-dependent quantity: it varies widely over regions and over time and reflects a key behavioural response—people's willingness and ability to engage in 'social distancing'.[4]

Credible models are desperately needed to evaluate the costs and benefits and time paths of the effects of different government policies, including financial bailouts, transfers to the public, and quarantines. We believe that having some model to guide decision making is better than none at all (so this is not a critique of Eichenbaum et al. (2020)'s analysis, which is probably the best possible under short notice). Otherwise, we have to trust critical policy decisions to the intuitive counterfactual predictions made inside the minds of our hopefully wise and well-informed leaders (a process otherwise known as 'human judgement'). Even though we acknowledge the shortcomings of SE and economic models, we are also pretty sure that there is no ML algorithm out there that could provide more credible advice on such momentous questions either.

But what about policy-making 'in the small', that is, for smaller, more well-defined questions? Following the lead article of Igami, we focus on a specific application where ML has achieved incredible practical success in 'policy advice': developing winning strategies for board games, such as chess, shogi, and Go, which by design are complicated for human players. The huge success that ML has achieved in this domain is easy to appreciate: ever since IBM's Deep Blue computer beat Garry Kasparov in 1997, the world's best chess players have been computers.[5]

Are there comparable 'success stories' for SE? The honest answer is that even though we do have a number of them (see Rust, 2019 and further discussion below), none of them have attracted the level of media attention that IBM did in 1997, or that Google and its DeepMind subsidiary achieved two decades later when they announced that, 'Starting from random play, and given no domain knowledge except the game rules, *Alpha Zero* achieved within 24 hours a superhuman level of play in the games of chess and shogi (Japanese chess) as well as Go, and convincingly defeated a world-champion program in each case' (Silver et al., 2017).[6]

[4] For example, Kucharski et al. (2020) found that the travel ban in Wuhan caused R_t (a time-varying version of R_0) to decline from 2.35 before the ban to just 1.05 the week after, while Korolev (2020) estimates R_0 to be 10 or higher using U.S. data.

[5] Note the irony of the term 'computer', which originally referred to women employed to calculate artillery trajectories at Aberdeen Labs in the U.S.A. in World War II. With the advent of the 'von Neumann machine' at the Institute for Advanced Studies in the early 1950s, human computers became obsolete and were replaced by machines; see Dyson (2012).

[6] The reader may wonder how the task of finding winning strategies for board games relates to the main goal of ML noted above: prediction. Note that a *best predictor* can be cast as the mathematical problem of finding a strategy that minimizes an expected loss function. A winning strategy for board games can be cast as the mathematical problem of

Alpha Zero is an example of unsupervised reinforcement learning, since it requires no human teachers and quickly achieved superhuman performance via self-play. This success revitalized AI, which for many decades had promised much more than it actually delivered. But be careful what you wish for: the successes of this new incarnation of AI are revolutionary, but also deeply threatening to many of us 'ordinary humans'. Deep-net speech recognition and translation software and vastly improved industrial robots are starting to put many lower-skilled workers out of jobs. But what about us? Are we still safe from this AI revolution? Will it threaten our livelihoods as structural econometricians by making it easy to train deep neural nets to perform at human or even superhuman levels as the artificially intelligent agents (e.g., the consumers, firms and government institutions) in our economic models, enabling us to make counterfactual predictions by studying how they then respond to different hypothetical environments and policies? If this really is possible, it seems that we can give up learning about economic institutions and tools such as econometrics, dynamic programming and game theory, and simply download *TensorFlow* from Google,[7] formulate an environment we wish to analyse, let the deep nets train themselves via repeated self-play, and *voila!* in just a few hours we will have realistic and detailed simulations of how intelligent agents behave in any given economic environment. When this day arrives, even we will be redundant.[8]

3. DYNAMIC PROGRAMMING: THE LINK BETWEEN ML AND SE

Rust (2019) addressed some of the questions raised above in a survey on the practical applications of dynamic programming (DP), the optimization principle underlying the most dramatic successes in ML and most work on SE. In fact, we argue that DP rather than ML should be given primary credit as the key idea underlying the success of *Alpha Zero.* Yet, achieving superhuman play in board games is one thing, but has DP measurably improved real-world decision making in other domains such as helping individuals and firms to make better decisions? Rust concluded that

> 'DP is an extremely powerful tool for solving a huge class of sequential decision problems under uncertainty. In fact, it is hard to think of problems that cannot be formulated and solved using DP. The flexibility and power of DP has revolutionized the way we do economics, and it has provided a key principle that has produced some of the most impressive achievements in the field of AI. Despite this, I have only been able to come up with a few examples where DP is used and has resulted in demonstrable improvements in real-world decision making. What is the explanation for this paradox?', p. 854

To explain this paradox and the key ideas underlying the *Alpha Zero* program, we provide a bit of notation and introduce the fundamental *Bellman equation* of DP that characterizes the solution to a class of infinite-horizon stationary dynamic programming problems known as *Markovian decision problems* (MDPs, for a more detailed discussion and definition, see Howard, 1960). In a MDP, a decision maker (agent) makes choices in discrete time, and these choices are affected by the agent's *state*, denoted by x. At each time t, the agent can take an action or decision d from a set of feasible actions $D(x)$, and receives a reward (or payoff, or utility, to use the terminology of economics) $u(x, d)$. After taking the decision at time t, the state at time $t + 1$, x_{t+1}, is a realization of

finding a strategy that minimizes a loss function equal to the probability of losing the game, and so it can be viewed as a prediction problem.

[7] TensorFlow is available at `https://www.tensorflow.org`, and an alternative PyTorch Python library is available at `https://pytorch.org`.

[8] Maliar et al. (2019) raised a similar point in their paper, 'Will artificial intelligence replace computational economists any time soon?' They concluded that 'it is unlikely to happen, at least in the near future', and that 'human input is still critical for designing effective machine-learning methods'.

the Markov transition probability $p(x_{t+1}|x_t, d_t)$. The agent chooses a feasible Markovian strategy or decision rule $d = \delta(x)$ (i.e., a rule that specifies which feasible action $\delta(x) \in D(x)$ to take in each state x). It is Markovian in the sense it depends only on the current state x and not on the entire history of previous states. The agent's goal is to choose an optimal strategy δ^* that maximizes their expected discounted utility or reward, where future utilities/rewards are discounted by the agent's *discount factor* $\beta \in (0, 1)$:

$$\delta^* = \underset{\delta}{\operatorname{argmax}}\, E\left\{\sum_{t=0}^{\infty} \beta^t u(x_t, d_t)\right\}. \tag{3.1}$$

The *Bellman equation* is a recursive expression for the optimal discounted reward, known as the *value function* $V(x)$:

$$V(x) = \max_{d \in D(x)}\left[u(x, d) + \beta \int V(x')p(dx'|x, d)\right]. \tag{3.2}$$

We can define the *decision-specific value function* v from V via the following equation :

$$\begin{aligned} v(x, d) &= u(x, d) + \beta \int V(x')p(dx'|x, d) \\ &= u(x, d) + \beta \int \max_{d' \in D(x')}\left[v(x', d')\right] p(dx'|x, d), \end{aligned} \tag{3.3}$$

and from this, the optimal decision rule is simply equal to the decision that maximizes $v(x, d)$:

$$\delta^*(x) = \underset{d \in D(x)}{\operatorname{argmax}}\, v(x, d). \tag{3.4}$$

Note that the Bellman equation can be written compactly as a functional fixed point $V = \Gamma(V)$, where Γ is known as the *Bellman operator*, defined from the right-hand side of the Bellman equation (3.2). Similarly, the decision-specific value function v can be written as the fixed point to a closely related Bellman-like operator $v = \Psi(v)$, where Ψ is defined via the right-hand side of Equation (3.3). As is well known, both Γ and Ψ are contraction mappings, so V and v can be computed via the method of *successive approximations* starting from any arbitrary guesses; that is, iterations of the form $V_{t+1} = \Gamma(V_t)$ and $v_{t+1} = \Psi(v_t)$ will converge to the corresponding fixed points $V = \Gamma(V)$ and $v = \Psi(v)$ as $t \to \infty$ starting from any initial guesses V_0 and v_0, respectively. Note that at each iteration *t all states x and all decisions d are updated synchronously.*

DPs and MDPs are the essential ingredients in the modern literature on SE, particularly for the literature on dynamic discrete choice models. For example, Rust (1987) applied the above framework in a straightforward extension of the work of McFadden (1974) on the static discrete choice model. The latent variables affecting choices are incorporated via *unobserved state variables* ϵ in addition to observed states x. The element ϵ_d of this vector affects the utility of agent's choice d out of a finite set $D(x)$ of feasible alternatives. Under three key assumptions, namely (a) additive separability (i.e., the unobserved state ϵ enters the objective function additively, as $u(x, d) + \epsilon_d$), (b) conditional independence (i.e., the controlled transition probability for the overall state vector (x_t, ϵ_t) factors as $p(x', \epsilon'|x, \epsilon, d) = q(\epsilon'|x')p(x'|x, d)$), and (c) the conditional density $q(\epsilon'|x')$ is an independent zero mean Type 1 extreme value distribution, the fixed-point equation (3.3) can be rewritten as

$$v(x, d) = u(x, d) + \beta \int \sigma \log \left(\sum_{d' \in D(x')} \exp\{v(x', d')/\sigma\} \right) p(dx'|x, d), \tag{3.5}$$

where the expectation with respect to the unobserved ϵ is computed analytically owing to the max-stability property of the extreme value distribution, which has a common scale parameter σ for all components ϵ_d. The integrand in (3.5) is the expectation of the maximum of the value function over the extreme value shocks ϵ:

$$V_\sigma(x') = \int \max_{d' \in D(x')} \left[v(x', d') + \epsilon'_{d'} \right] q(\epsilon'|x') = \sigma \log \left(\sum_{d' \in D(x')} \exp\{v(x', d')/\sigma\} \right), \tag{3.6}$$

which provides a closed-form expression for the agent's expected maximized utility from choosing between alternatives d' from the choice set $D(x')$. McFadden (1981) called this the 'social surplus function'.

Rust (1987) showed that, in accordance with (3.4), in the model with unobserved states the optimal decision rule for the agent takes the form

$$\delta^*(x, \epsilon) = \underset{d \in D(x)}{\operatorname{argmax}} \left[v(x, d) + \epsilon_d \right], \tag{3.7}$$

implying in turn that the *conditional choice probability* $P(d|x)$ (i.e., the conditional probability that an agent in observed state x takes decision d) has the form of the *multinomial logit* probability that McFadden (1974) derived for the case of static discrete choice models:

$$P(d|x) = \frac{\exp\{v(x, d)/\sigma\}}{\sum_{d' \in D(x)} \exp\{v(x, d')/\sigma\}}. \tag{3.8}$$

Assuming that the utility function $u(x, d, \theta_1)$ depends on an unknown vector of parameters θ_1 and that the transition probability $p(x'|x, d, \theta_2)$ depends on parameters θ_2, Rust (1987) showed that the parameter vector $\theta = (\beta, \theta_1, \theta_2, \sigma)$ can be estimated by maximum likelihood, given data on the observed states and decisions $\{d_t, x_t\}$ of a given agent. The vector θ is referred to as the *structural parameters*, since once we know these parameters we know the underlying structure of the decision problem, $\{\beta, u, p, q\}$ (where q is the Type 1 extreme value distribution for the unobserved states, ϵ, whose only unknown parameter is σ).

Rust (1987) structurally estimated a model of the bus engine replacement decisions of Harold Zurcher, the maintenance manager whose decision $d \in \{0, 1\}$ concerned whether or not to replace a bus engine when the bus had travelled x miles since the last bus engine replacement. The maximum likelihood problem can be written as a constrained optimization problem:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\{d_t, x_t\}_{t=1}^T | x_0, d_0, \theta) \equiv \underset{\theta}{\operatorname{argmax}} \prod_{t=1}^{T} P(d_t|x_t, \theta) p(x_t|x_{t-1}, d_{t-1}, \theta_2), \tag{3.9}$$

subject to the 'fixed point constraint' $v = \Psi(v)$ given by (3.5) that determines the implicit choice specific value function $v(x, d, \theta)$.

Rust (1987) called the algorithm for computing the maximum likelihood estimator $\hat{\theta}$ in equation (3.9) a *nested fixed point (NFXP) algorithm*, since at each trial value of θ in a standard nonlinear search for an optimizer $\hat{\theta}$, the NFXP algorithm computes $v(x, d, \theta)$ as a fixed-point subproblem (3.5) for all values of (x, d). The NFXP algorithm is computationally intensive owing to the need to repeatedly solve for the fixed point, but using a combination of globally convergent successive approximations with fast Newton–Kantorovich iterations provides a significant advantage over the naive application of successive approximations, which is a standard method to find the fixed point of a contraction mapping (Rust, 2000).[9]

In the ML literature, the functional form (3.8), which we call the multinomial logit model, is imprecisely called the *softmax function*. We agree with Goodfellow et al. (2016), who note that it is more accurate to call it the 'soft argmax function', since $P(d|x)$ is the conditional probability that the agent's optimal choice in state x is d (i.e., the probability distribution of d, the argmax). The term 'softmax' is a more appropriate name for the social surplus function (3.6), also called the 'Emax function' since it equals the expected maximum utility and satisfies

$$\lim_{\sigma \to 0} V_\sigma(x') = \max_{d' \in D(x')} \left[v(x', d') \right], \tag{3.10}$$

where $v(x', d')$ on the right-hand side is the fixed point to equation (3.3), the initial version of the MDP with no unobserved state variables ϵ (equivalently, where the scale parameter of the unobserved shocks is $\sigma = 0$); see Theorem 3 of Iskhakov et al. (2017).

4. CAN ML BREAK THE CURSE OF DIMENSIONALITY?

Rust (2019) discussed the *curse of dimensionality* of solving large-scale DP models as one of the key challenges that have impeded our ability to structurally estimate realistic, high-dimensional DP models and to apply DP to solve various real-world problems. In simple terms, the curse of dimensionality refers to a situation in which the *state variable* x of a DP problem assumes too many possible values to enable effective computation of an optimal strategy $\delta(x)$ that specifies an optimal decision $d = \delta(x)$ in any possible state x. This is certainly the case in most board games, such as chess or Go, where x summarizes the 'board position' at any point in the game. For example, according to *Wikipedia*, in Go there are around 2.08×10^{170} legal board positions. Even if the synchronous successive approximation algorithm $V_{t+1} = \Gamma(V_t)$ were to converge to the fixed point $V = \Gamma(V)$ in relatively few iterations, it is computationally infeasible to evaluate either (3.2) at all possible x states at each iteration, or (3.3) at all possible (x, d) values. Some method of interpolating or approximating values using a much smaller number of states/decisions is necessary.

There are several aspects where the curse of dimensionality appears in DP problems: (a) in problems with vector-valued continuous decisions d, in searching for the vector d that optimizes the value function at each point in the state space, (b) in multivariate integration to compute the conditional expectation of the value function, and (c) in approximating value functions when the state vector x is continuous and multi-dimensional. ML provides a number of methods for ameliorating the latter curse of dimensionality, such as neural networks, which are known to be capable of approximating certain classes of multivariate functions using a vector parameters that does not increase exponentially fast in the dimension of the state vector x: see, for example,

[9] Su and Judd (2012) proposed a constrained optimization approach to structural estimation, MPEC. This method simultaneously computes the value function and the structural parameters, $\hat{\theta}$, by solving problem (3.9) directly as a constrained optimization problem, where the Bellman equation (3.5) is imposed as a set of constraints. Su and Judd concluded that 'the constrained optimization approach can be significantly faster' (p. 2213), but Iskhakov et al. (2016) found that NFXP and MPEC are roughly equivalent for small problems, and that NFXP can be faster for large-scale problems. This is due to the fact that MPEC optimizes not only over the structural parameter θ as NFXP, but over all values of $v(x, d)$, thus solving a substantially higher-dimensional optimization problem.

Barron (1994), and Montanelli and Du (2019). However, while ML is helpful, as we show below, it cannot succeed in breaking the curse of dimensionality. For example, Barron's approximation result assumes that there is an algorithm that can find parameters of the neural network that are global minimizers of a nonlinear least squares problem, but they are notorious for numerous local optima. There is a curse of dimensionality in global optimization that cannot be circumvented by any known optimization method, at least in the worst case; see Nemirovsky and Yudin (1983). Even though ML cannot break the curse of dimensionality, it offers a promising new set of approximate solution methods for large DPs.

For example, research into *reinforcement learning* made great progress by adopting the DP paradigm and introducing asynchronous stochastic algorithms for solving DP problems such as *Q-learning* (Watkins, 1989).[10] Q-learning, originally proposed as a method to solve finite-state MDPs, is an iterative, asynchronous stochastic algorithm for approximating the fixed point $v = \Psi(v)$ of equation (3.3) using the contraction property of Ψ to guarantee that the sequence converges with probability 1 to the true solution $v = \Psi(v)$. The idea is that agents can improve their performance in 'real time' using their current estimate of the decision-specific value function $v_t(x, d)$ by only updating *asynchronously*, that is, only updating the state actually realized (x_t, d_t) at time t according to the updating formula

$$v_{t+1}(x, d) = \tag{4.1}$$
$$\begin{cases} v_t(x_t, d_t) + \alpha_t \left[u(x_t, d_t) + \beta \max_{d' \in D(\tilde{x}_{t+1})} v_t(\tilde{x}_{t+1}, d') - v_t(x_t, d_t)\right] & (x, d) = (x_t, d_t) \\ v_t(x, d) & \text{otherwise,} \end{cases}$$

where $\tilde{x}_{t+1}$ is a realization from the transition probability $p(x'|x_t, d_t)$, namely the next realized state if the agent chose to take action d_t in state x_t, and α_t is a sequence of positive (possibly stochastic) step-size parameters that decrease to 0 with probability 1 as $t \to \infty$ (though not too fast). Tsitsiklis (1995) showed that the Q-learning iteration is a form of *asynchronous stochastic approximation* and provided sufficient conditions for the algorithm to converge to the true fixed point $v = \Psi(v)$ (that is, the true decision-specific value functions) with probability 1 as $t \to \infty$. Thus, Q-learning is in the class of reinforcement learning algorithms since they lead to a sequence of decisions that gradually self-improve and converge to the optimal decision rule δ^* with experience as $t \to \infty$, at least in the case where the MDP has a finite number of possible states and decisions.

Notice also that instead of updating *all* (x, d) in each iteration, Q-learning only updates the *realized state and decision* (x_t, d_t), and thus the iterations are not only stochastic but asynchronous (that is, the values for all states are not being updated at each time step). However, for decision problems with a huge number of possible (x, d) pairs such as chess or Go, most of these pairs will rarely occur, so the convergence of Q-learning is very slow, and the performance of an approximate decision rule is highly dependent on the quality of the initial guess v_0. This problem can be avoided by using deep neural networks to approximate $v_t(x, d)$ with far fewer parameters than the total number of possible states and decisions (x, d). Further progress came from using algorithms such as stochastic gradient descent to adjust the parameters of these deep nets, and the result was a new class of computational strategies for circumventing, or at least dealing with, the curse of dimensionality. In board games, reinforcement learning enabled the training of effective strategies without the need for supervised learning by trying to mimic the recorded play of chess masters. *Alpha Zero* achieved superhuman ability entirely via repeated self-play.

[10] See Bertsekas and Tsitsiklis (1996) and Powell (2010) for other related algorithms called 'neuro DP' or 'approximate DP', respectively.

Economists generally have not used stochastic reinforcement learning algorithms to approximately solve DP problems, although the 'stochastic algorithm' of Pakes and McGuire (2001) is a prominent exception in the context of dynamic games. The most common method used by economists is successive approximations, which is just classical backward induction with 'brute force' solution of sufficiently small problems where all states can be exhaustively enumerated. In infinite-horizon problems, where the value function can be shown to be a fixed point to a contraction mapping, more efficient methods such as policy iteration are often used.[11]

In problems with continuous state variables, the value function, policy function, or some transformation of them, is often approximated with smooth parametric function classes such as Chebyshev polynomials or splines of various types and orders. To ameliorate the curse of dimensionality, such approximations can be used in combination with Smolyak's algorithm for constructing sparse tensor product grids, and can be modified to target a higher quality of approximation in some dimensions than in others, as in Judd et al. (2014). For smooth multivariate functions, sparse grids reduce the number of grid points M needed to obtain a given error tolerance to $O(M\log(M)^{dim(x)-1})$, compared with $O(M^{dim(x)})$ for a full Cartesian grid. However, in discrete-choice applications, the value functions display kinks and the policy functions have discontinuities. In these situations, the *adaptive* sparse-grid algorithm suggested by Brumm and Scheidegger (2017) can significantly improve approximation quality by sequentially refining the sparse grid and recursively adding more points in regions of high curvature, with fewer points in regions of low function variation. However, these methods rely on the construction of regular deterministic grids in $dim(x)$-hypercubes, and therefore they are still subject to the curse of dimensionality.

The 'self-approximating' random Bellman operator method of Rust (1997) completely avoids the need for interpolation and approximation and tensor-based grids. He showed that his random multigrid method can break the curse of dimensionality for a class of models where the decision variable is finite but the state variables are continuous and multidimensional. Although the regularity conditions for the basic version of the algorithm are violated for DP problems where the transition density has large discontinuities or spikes, randomized algorithms may be a promising way to approximate solutions to certain classes of high-dimensional DPs, similar to the way that stochastic reinforcement learning algorithms have proved effective for finding good strategies in chess or Go.[12]

Kristensen et al. (2020) combined smoothing, importance sampling and sieve approximations to solve for the fixed point of the Bellman operator of dynamic discrete choice models. An advantage of this approach is that sieve approximations make it possible to sample directly from $p(x'|x, d)$, and thereby avoid the 'needle in a haystack problem' of simulating peaked conditional

[11] Howard (1960). See also Puterman and Brumelle (1979), who showed that policy iteration is equivalent to solving for V as a zero of a nonlinear operator using Newton's method.

[12] There is no proof we are aware of that stochastic reinforcement learning algorithms can formally break the curse of dimensionality. Bray (2020) points out a key bounding assumption on the transition density $p(x'|x, d)$ that Rust (1997) relied on to prove that his random multigrid algorithm breaks the curse of dimensionality implies that asymptotically as the dimension of the x state vector tends to infinity, 'all but a vanishingly small fraction of state variables to behave arbitrarily similarly to i.i.d. uniform random variables.' Rust et al. (2002) introduced a deterministic weighted tensor product (WTP) approximation algorithm to approximate the fixed points of a class of high-dimensional quasi-linear contraction mappings, which includes the value function of discrete choice DP problems as a special case. They provide regularity conditions under which the WTP algorithm is *strongly tractable*, that is it breaks the curse of dimensionality without the use of randomization and 'attains nearly the same rate of convergence that can be attained for unidimensional problems' (p. 321). However their regularity conditions imply that this result holds only for a class of problems that have 'special structure' that is similar to Bray's result, and can be described intuitively as holding for a class of functions for which 'the dependence of a function on its i-th variable decreases with increasing i.' (p. 322).

densities based on a marginal sampler, but instead automatically generate sample grid points where the problem lives. Using shrinkage estimators and variable selection ML methods to approximate value functions on this random grid, it is then possible to detect sparsity patterns implied by the model and thereby also reduce the required number of simulated grid points.[13]

A benefit for SE when using randomly sampled grid points to approximate the Bellman operator is that it is easier to analyse the consequences for inference that follow from approximating value functions. Kristensen et al. (2020) provided an asymptotic theory for two approximate solution methods: the proposed simulation-based sieve approximation method, and a proposed generalization of the self-approximation method in Rust (1997) that allows for importance sampling. They showed that both methods converge with a $\sqrt{M}$ rate towards Gaussian processes, where M is the number of random grid points. In turn, this result can be used to derive the asymptotic distribution and convergence properties for the resulting approximate estimator used when estimating the structural parameters.

Scheidegger and Bilionis (2019) used a combination of Gaussian process ML and the active subspace method. The former is a grid-free form of supervised ML that uses a set of simulated points to interpolate or approximate the value function without any geometric restriction. The latter is able to discover the patterns that reduce the dimensionality of the space on which the Bellman operator is projected, by detecting the directions of the strongest variability in the value function and subsequently exploiting these directions to construct a response surface. Using this method in conjunction with massive parallel computing, these authors are able to solve very-high-dimensional DP problems with hundreds of continuous state variables.

The SE literature has also developed ways to estimate the underlying structure by *bypassing the need to solve the DP problem*, or at least bypassing the need to solve it accurately. These methods include the Hotz and Miller (1993) estimator for dynamic discrete choice models that inverts nonparametric estimates of conditional choice probabilities $P(d|x)$ to obtain (differenced) estimates of the choice-specific value functions $v(x, d)$, which under certain additional assumptions can be used to obtain estimates of the underlying structure. There is also the Bajari et al. (2007) moment inequality estimator, which uses simulations of states and decisions from nonparametric estimators of conditional choice probabilities $P(d|x)$ and transition probabilities $p(x'|x, d)$ to create a 'forward simulation' estimator of $v(x, d)$ at different (x, d) values. By comparing the optimal simulated decision with other suboptimal decisions d' they exploit inequalities implied by equilibrium and optimality to estimate the structure using only relatively noisy estimates of $v(x, d)$ rather than a precisely calculated solution.

We are starting to see applications of sequential estimators such as those developed in Hotz and Miller (1993) and Bajari et al. (2007), where ML is used for a first-step estimation of reduced-form policy functions and state transition distributions. However, naively plugging ML estimators into the sample criterion used in the second step may cause substantial bias in estimators of the structural parameters due to regularization bias and overfitting in the first step. For such two-step estimators to be root-n consistent, Chernozhukov et al. (2018) showed that they must be based on orthogonal/locally robust (LR) moment conditions, and Chernozhukov et al. (2016) developed the relevant adjustments to obtain orthogonalized moment conditions for dynamic discrete choice models that adjust for bias owing to first-step ML estimation of conditional choice probabilities to be used with Hotz and Miller (1993).

[13] Renner and Scheidegger (2018) do a similar thing using a type of score function to enhance the training set where it reduces most the uncertainty about the function. This can loosely be considered as a grid-free version of adaptive sparse grids.

Even though ML can potentially address the curse of dimensionality by employing model selection when estimating high-dimensional choice probabilities, data still limit what we can learn about the underlying model structure. But even in the ideal case where ML can recover a precise, sparse representation of $P(d|x)$ that allows us to estimate the structural parameters, we cannot rely on this approach for counterfactual simulations. If choice probabilities $P(d|x)$ fundamentally change shape in the counterfactual and require a different set of variables and basis functions, it is still necessary to solve high-dimensional DP problems whose sparsity structure is unknown. Therefore, ML-based two-step estimators may have to be used in tandem with ML methods that can approximate solutions to high-dimensional problems, such as the work by Norets (2012), Scheidegger and Bilionis (2019), and Maliar et al. (2019).

Now let us return to the question raised at the start of this subsection: Is the 'curse of dimensionality' the key explanation for Rust's paradox? That is, if DP is such a powerful tool, why aren't there more real-world success stories for its application beyond superhuman performance in board games? After all, Google probably did not spend over $500 million in 2014 to acquire DeepMind (the company founded by chess prodigy Demis Hassibis and others that developed the AlphaGo and Alphazero strategies) just for the sake of 'knowledge' or the publicity value of being able to claim it is the world champion in chess. Presumably, Google sees this as a stepping stone to much bigger and better things, including applying reinforcement learning and deep learning to solve much more practical problems that businesses and organizations would pay top dollar for.

Reinforcement learning and deep neural networks cannot break the curse of dimensionality of DP. Chow and Tsitsiklis (1989) proved that there is an unavoidable curse of dimensionality associated with approximately solving continuous-state/continuous-decision DP problems, a curse that cannot be broken by *any algorithm*, including deep neural net approximators and reinforcement learning.[14] However, even if *Alpha Zero* is not strictly optimal (or in equilibrium against itself), it still has demonstrated *superhuman performance.* If this approach for generating extraordinary performance can be replicated in other domains, such as to develop advanced trading/arbitrage algorithms on Wall Street, or more effective real-time pricing algorithms for hotels and airlines, or better inventory management systems for Amazon and Walmart, the value of these contributions for our economy and for Google, Amazon and Walmart could be tremendous—worth potentially many billions or even trillions of dollars. Google would have shown that 'deep Q nets' is indeed a killer application.

The fact that we have not heard that Google has extended the big successes of *Alpha Zero* to other domains may simply be a natural delay while these other applications are being rolled out under the grip of corporate secrecy. However, there is another less optimistic interpretation for why we are not hearing more 'DP success stories' beside *Alpha Zero:* it might be the case that further extensions of deep neural nets and reinforcement learning to more complex real-world domains face fundamental challenges that are very different from the ones that are confronted in developing a winning strategy in chess or Go. For example, we have heard about Google's project for self-driving cars that began in 2009 and led to the spin-off of the Waymo subsidiary. Although it is possible to see 'Google-mobiles' driving in Palo Alto, Phoenix, and other cities where these cars are permitted, they still always have a human co-pilot watching over the automated driver. Owing to a number of well-publicized accidents involving 'driverless vehicles', the eventual roll-out of this version of AI still seems many years away.

[14] Only in cases where DP problems have some additional special structure is it possible to break the curse of dimensionality. For example, the subclass of linear-quadratic DP problems can be solved in polynomial time. The curse can also be broken for the subclass of DP problems with continuous state variables when there are a finite number of possible actions in each state—see the discussion in footnote 12 and also Kristensen et al. (2020).

This is not to diminish the huge success that Google has attained in board games with its combination of reinforcement learning and deep neural networks to approximately solve this DP problem, but there is much less evidence that these successes can be easily transported to other real-world economic problems, such as advising firms how to set hotel or airline prices, or when and how much to invest in inventory, or how much an individual should save for retirement and when is the best time to retire. Solving these problems involves more than just an ability to approximate an optimal strategy: an additional problem is to learn the objective function, constraints and environment in which these agents are operating. We are skeptical that ML by itself (that is, without any human supervision or guidance) constitutes a good solution for these deeper learning problems. To put things somewhat crudely, would you trust a Google deep net to pick your future spouse, to set the length of your prison sentence, to decide on whether your submission to *Econometrica* should be accepted, or to fly the Boeing 737 Max that you are riding on without any human supervision or intervention?

5. THE OTHER MEANINGS OF THE TERM ‘LEARNING’

The daunting problem that is preventing wider successful practical applications of AI is a learning problem that we have been grappling with in SE for decades: the difficulty of *learning about the objective function and environment facing real-world decision makers.* This is a learning problem so difficult that it is far from clear that it is amenable to any quick or easy solution using the available tools from ML.

‘Learning’ is an overloaded piece of terminology that is often used imprecisely in the ML literature. One key use of ‘learning’ involves approximating an optimal strategy. But another aspect of ‘learning’ is *inference*, which we noted above is the main goal of SE: to infer the decision maker’s objective function and beliefs about the environment in which (s)he operates.

In board games such as chess, the objective function is generally known a priori—it can be defined to equal 1 for winning the game and 0 otherwise, and an optimal strategy maximizes the probability of winning the game.[15] The only ‘unknown’ in these games is the optimal strategy, that is, the function that maps any possible state of the game to a feasible move that maximizes the objective (for example the probability of winning). Although it might be reasonable to say that the goal of ML in this instance is to ‘learn’ what this optimal strategy is, it would be clearer to say that the objective is to compute or to approximate the optimal strategy.

Bayesian learning is the predominant example of inferential learning, where observation of data causes the statistician or econometrician to revise their prior beliefs about unknown parameters via Bayes’ rule. Under ideal conditions, with enough data a Bayesian will learn the truth—the posterior distribution converges to a point mass on the true parameter values with probability 1 as the sample size tends to infinity. However, almost from its inception, the literature on SE recognized the *identification problem* that imposes severe limits on our ability to learn the underlying structure in the absence of strong a priori assumptions. When identification fails, the posterior distribution converges to a distribution over *observationally equivalent* parameter values, so we cannot learn the truth even with an infinite amount of data.[16]

[15] An alternative objective is to assume the payoff of a draw is 0 and a loss is –1. A strategy that maximizes expected payoff for this alternative objective is not necessarily the same as a strategy that maximizes the probability of winning the game.

[16] Freedman and Diaconis (1983) and Diaconis and Freedman (1986) discuss the failure of Bayesian posterior to converge to the truth when the parameter space is infinite-dimensional.

For example, the identification problem was recognized in some of the earliest work on SE at the Cowles Commission on the linear simultaneous equation model (an early static structural model). Even if we assume that a given market is in competitive equilibrium, mere observation of a sequence of equilibrium price quantity pairs (p_t, q_t) is insufficient by itself to infer separately the coefficients of linear supply and demand equations. Additional a priori assumptions are required to identify supply and demand, such as *exclusion restrictions* that exclude at least one 'demand shifter' variable from the supply curve equation and at least one 'supply shifter' variable from the demand equation.

There is also an identification problem for static and dynamic discrete choice models. One way to look at it is to consider what general testable implications the hypothesis of utility maximization places on conditional choice probabilities, $P(d|x)$. McFadden (1975) used the term 'mother logit' to describe the way an arbitrary conditional choice probability can be given a utility-maximizing interpretation. To see this, consider the static random utility model with extreme value unobservables with the utility function $u(x, d) = \log[P(d|x)]$: it is easy to see that plugging this into the multinomial choice probability formula (3.8) setting $\sigma = 1$ succeeds in 'rationalizing' the conditional probability $P(d|x)$ regardless of whether $P(d|x)$ is really generated by the random utility maximization model with extreme value taste shocks.

The decision-specific value function $v(d, x)$ plays a key role in the estimation of dynamic discrete choice (DDC) models, and McFadden's mother logit construction directly extends to DDCs, except that the value function $v(x, d)$ takes the place of the single period or static utility function $u(x, d)$. A simplistic way of stating the identification problem is that without further a priori restrictions or assumptions, it is always possible to rationalize any given conditional choice probability $P(d|x)$ as consistent with rational behaviour of a DDC model satisfying the three key assumptions of Section 2 (that is, additive separability, conditional independence, and extreme value unobservables), and, moreover, there are *infinitely many ways* to rationalize this behaviour. It is a rather depressing result that prescribes a fundamental limit on our ability to do inference.

The conditional choice probability can be viewed as a probabilistic choice version of the decision rule $\delta(x)$ and forms the basis for inference about a decision maker's preferences and beliefs. In engineering terminology, structural estimation of DP models is essentially a type of 'inverse DP' that attempts to recover the underlying 'structural objects'—the decision maker's objective function and beliefs about the laws of motion for the state variables x—from observed states and decisions (x, d). Specifically, to carry forward with the MDP example, we define the underlying *structure* of the MDP to be the objects $\Lambda = (\beta, u, p, q)$, where β is the agent's discount factor, u is the reward or utility function, and the pair p, q represents the agent's beliefs about the stochastic 'law of motion' of the environment in which it operates (observed and unobserved states, respectively). With sufficient data on (x, d), we can treat the conditional choice probability $P(d|x)$ and the transition probability $p(x'|x, d)$ as known, since with sufficient data these objects can be estimated nonparametrically. In deference to the older literature on identification, $\{P, p\}$ are referred to as the *reduced form* objects of the dynamic discrete choice model, since these objects can be 'learned'.[17] In the absence of additional a priori restrictions/assumptions, however, the underlying structure cannot be learned: Rust (1994) and Magnac and Thesmar (2002) showed that without further a priori assumptions the DDC model is unidentified; that is, the mapping

[17] There is an important unstated assumption of 'rational expectations' in place here, i.e., that the agent's subjective beliefs about the observed state variable are given by the objectively estimable transition probability $p(x'|x, d)$. Without rational expectations, the agent's beliefs, p, are not generally identified either.

between the structure $\{\beta, u, p, q\}$ to the reduced form $\{P, p\}$ is many to one.[18] But such a priori restrictions/assumptions sometimes do exist. For example, Abbring and Daljord (2020) show that the discount factor in dynamic discrete choice models can be identified using variation that shifts expected discounted future utilities, but not current utilities. This shows the essential role of human intuition in SE, which provides an advantage over 'mindless ML' by showing how to identify otherwise unidentified parameters via economically motivated exclusive restrictions.

Rust (2014) pointed out that the curse of dimensionality and the identification problem are two key 'limits to knowledge' confronting SE. As we discuss in the next section, the literature on ML appears to be unaware of the identification problem. Regardless, these limits to knowledge apply to 'machine learners' no differently than to human ones. It should not be surprising that there are fundamental limits to learning, framed as a process of *inductive reasoning,* just as Gödel's *Incompleteness Theorem* prescribes fundamental limits to the search for knowledge and truth via the process of *deductive reasoning.* Indeed, the identification problem can be regarded as one aspect of the 'problem of induction' that goes at least back to Hume (1739).

For better or worse, a lot of human knowledge relies on *assumptions.* These assumptions can often be tested and revised, but it seems almost obvious that there is no easy and *assumption-free path to knowledge.* Research in neuroscience, such as Griffiths and Tenenbaum (2009), suggests that a key to human intelligence is our innate ability to construct, test, and when necessary revise and update *mental models* of the world. Human knowledge is based on lots of assumptions, many of them wrong. When observations/data suggest that our mental models are sufficiently bad, we have an innate ability to revise our incorrect assumptions and build more accurate mental models. This activity, something humans do more or less automatically and subconsciously, is what structural econometricians have have been trying to do formally and consciously for nearly the last century. Whether this modelling process that occurs in our own brains (whether consciously or subconsciously) can ultimately be replicated by a *general AI* in the form of sufficiently deep neural networks or other forms of ML remains a tantalizing open question. However, for the foreseeable future, it appear that the human brain and human modellers have a decisive advantage in processing huge amounts of diffuse domain-specific knowledge, and the ability to narrow a vast space of possible hypotheses and theories about phenomena in the world to a much smaller, more manageable space via our ability to make good assumptions and models. From everything we can see, ML is still very far away from being able to replicate what appears to be our uniquely human capability.[19]

There is no doubt that AI and ML are making very rapid progress, but there are many applications where these approaches are running up against a wall when it comes to applying them more widely to real-world decision problems. Moreover, the 'wall' is not just the curse of dimensionality, it is also the lack of domain-specific knowledge and understanding that humans possess. This is why we still most often trust human judgement more than the recommendation of a machine. The human being understands what the objective is and has a better understanding than a machine of the often very fuzzy and complicated environment in which they are living. This explains why, outside of very limited and 'standardized' sorts of problems (such as board

[18] See Proposition 2 of Magnac and Thesmar (2002, p. 807), where they show that 'it is always possible to find other structural parameters and, in particular, utility functions, that are compatible with given data'.

[19] This is a well-recognized issue among ML/AI researchers. For example, David Cox, director of the MIT-IBM Watson AI Lab, notes that 'A huge fraction of what we do in our day-to-day jobs is constantly refining our mental models of the world and then using those mental models to solve problems. That encapsulates a lot of what we'd like A.I. to do.' (quoted in Smith, 2020).

games), AI has not taken over and made us all irrelevant (the fears of the late Stephen Hawking notwithstanding).

At the same time, Rust (2019) noted that there are narrower domains for which the application of DP, using SE or ML approaches, could have a revolutionary impact: ‘The most promising domain for application of DP in the short run is to improve firm decision making, particularly in pricing and revenue management, where there is evidence that computer-generated recommended prices have significantly increased revenues and profits.’ As a result, there has been rapid entry and growth in this industry: a report by Markets and Markets (2020) forecasts that it will grow at 9.6% per year from \$14.1 billion in 2019 to \$22.4 billion by 2024.

In computer science, computerized recommendation systems are called *actor/critic algorithms*. The human decision maker or firm is the ‘actor’ who makes the actual decisions, but potentially relying on the advice of a ‘critic’ who recommends certain actions. A prominent example of a successful and widely used actor–critic algorithm is GPS navigation, where the GPS recommends a route to a given destination, but the human driver may or may not follow that advice. In the GPS example, the objective is generally quite clear-cut: how to get from point A to point B with minimum cost (either time or distance). But in more complicated applications, the most painstaking task for an academic or a commercial service trying to play the role of the critic and recommend better decisions to the actor (that is, the individual or firm making the actual decisions) is to understand the *structure of the actor’s decision problem*.

Calculating an optimal solution to the wrong objective, or misspecifying the constraints and opportunities the actor actually confronts may result in unhelpful advice to the actor. As Rust (2019) noted, ‘it is like providing the right answer to the wrong question’, and he concludes that ‘the fuzziness of many real-world decision problems and the difficulty in mathematically modeling them are key obstacles to a wider application of DP in real-world settings’. This is where SE comes into play: to understand the environment and objectives of human and firm decision makers in order to assist them in solving their decision problems. This is a different type of learning than simply ‘learning’ how to approximate the optimal decision rule, the main problem that *AlphaZero* solved for board games.

6. INVERSE REINFORCEMENT LEARNING VERSUS SE

There is a recent growing literature on *inverse reinforcement learning* (IRL),[20] which originated in the paper by Russell (1998) and has evolved largely independently of the literature on SE. Its goal is identical to the inferential goal of SE (Arora and Doshi, 2018, p. 1). In particular, IRL and SE both assume that we observe data (x_t, d_t) on a particular decision maker whose decisions or actions are given by an optimal decision rule $d_t = \delta(x_t)$ from the solution to a Markovian decision problem (MDP).

The methods and approaches to inference used in the IRL literature are, however, quite different from the ones used in SE. IRL is less directly connected to the literature on statistical inference in which SE is grounded. For example, IRL does not pose the inferential problem as one of parameter estimation that can be solved via Bayesian inference or classical methods such as maximum likelihood or the method of simulated moments. IRL uses more informal criteria for measuring ‘goodness of fit’ between the model and data, and different algorithms such

[20] For a recent survey, see Arora and Doshi (2018) and the references therein.

as 'maximum margin optimization' (which seeks a reward function whose implied decision rule $\hat{\delta}_E$ is as close as possible to the true decision rule δ_E) or 'maximum entropy', where the latter 'recovers a distribution over all trajectories, which has the maximum entropy among all such distributions under the constraint that feature expectations of learned policy match that of demonstrated behavior' (Arora and Doshi, 2018, p. 14).

The IRL literature seems unaware of the substantial work on the identification problem in the SE literature, although the survey by Arora and Doshi (2018) indicates that they independently rediscovered it: 'A critical challenge, first noted by Ng and Russell (2000) is that many reward functions (including highly degenerate ones such as a function with all rewards values zero) explain the observations', so 'many reward functions in the set of all reward functions can generate policies that realize the observed demonstration. Thus, IRL suffers from an ambiguity in solution.' (p. 8).

Both IRL and SE are related to, but distinct from, the literature on *supervised learning*. In this context, standard supervised learning algorithms are focused on recovering a decision rule δ from the data $\{x_t, d_t\}$ on the states and decisions of a given expert, using various loss functions and ways of approximating the function δ. Where SE and IRL depart from the supervised learning and nonparametric estimation literatures is that they are not just focused on recovering the optimal decision rule δ_E, but rather on the underlying *structure* that leads to it, i.e., the reward function of the agent.

Igami's article discussed the evolution of computerized board games and noted that many of them are forms of supervised learning, since the algorithms were 'trained' using the observed play of chess masters. The latest generation of computer algorithms, such as *Alpha Zero*, are trained by self-play with copies of itself rather than by supervised learning, and thus does not need to observe human experts in order to achieve superhuman playing ability.

As we noted, it appears to us that this success of unsupervised learning may be limited to board games, where the terminal payoffs are known a priori and it is possible to use computers to rapidly simulate play in millions of training games. In real-world applications, such as the successful application of approximate DP methods to locomotive allocation and fleet sizing at the Norfolk Southern railroad reported in Powell et al. (2014), the objective function is far fuzzier, and substantial work needs to be invested to produce a computer simulation model that can accurately simulate traffic demands, locomotive movement, maintenance and breakdowns. The computer model that performed these simulations, called PLASMA, required a separate multi-year development effort in which the team 'carefully calibrated the model against several years of historical performance'. The authors note that, 'This required us to painstakingly examine detailed assignments and compare high-level performance metrics. The process required that we iteratively identify and correct data errors, enhance the model, and make occasional improvements to the basic algorithm' (Powell et al., 2014, p. 10).

It is unclear whether ML can, without any human intervention or guidance, learn enough about the objective function and the environment to produce an accurate simulation model to train the DP algorithm to make better locomotive allocation/scheduling decisions in the way that *Alpha Zero* was trained in board games. However, Powell et al. (2014) demonstrate that a *combination* of human and ML can result in a demonstrably superior performance to human judgement only, i.e., to scheduling locomotives by human experts at Norfolk Southern without the benefit of the computerized recommendation system. These hybrid human/AI actor/critic algorithms hold great promise for future applications to challenging real-world problems.

7. RELAXING THE ASSUMPTION OF UNBOUNDED RATIONALITY

IRL and SE can be viewed as types of supervised learning that are 'trained' from observation of human experts *under the assumption that the human experts behave optimally.* Simon (1992) questioned this assumption, which he referred to as *unbounded rationality*, and noted that 'by the middle 1950s, a theory of bounded rationality had been proposed as an alternative to classical omniscient rationality [and] a significant number of empirical studies had been carried out that showed actual business decision making to conform reasonably well with the assumptions of bounded rationality but not with the assumptions of perfect rationality' (p. 357).

Unfortunately, Simon's work has not had the influence it deserves in economics, and despite a resurgence of interest in *behavioural economics,* unbounded rationality remains the predominant paradigm.[21] To convince you of Simon's view that most real agents *satisfice* rather than optimize, let us return to the example of board games. In dynamic games, the principle of unbounded rationality implies that humans play *Nash equilibrium strategies.* The traditional application of game theory has no role for differential ability, since all agents are assumed to have unbounded memory and computational capacity. But if this is true, it should have been impossible for Deep Blue to beat Garry Kasparov in 1997, and for the subsequent generations of computerized algorithms for other board games such as shogi and Go to achieve superhuman performance. If humans are unboundedly rational, these computerized strategies could at best only equal human performance.

In fact, even the current best AI strategies (such as *Alpha Zero*) are also boundedly rational, as evidenced by the continuing evolution of strategies and steadily improving chess ratings, such as measured by the ELO score.[22] At his peak, Kasparov's ELO rating was 2,851, and the current (human) world champion, Magnus Carlsen, has an ELO rating of 2,840; see Alliot (2017). The computer program *Stockfish* has an ELO rating of 3,334, and *Alpha Zero* has an ELO in excess of 3,500. Rutman (2018) noted that the fractions of drawn games is an increasing function of the players' ELO scores. Extrapolating the linear trend, he makes a prediction that the draw rate will equal 1 when the ELO rating exceeds 5,200. Thus, perfectly played chess may well be boring: every game ends in a draw. We already know this is true of checkers, since the combinatorics of the state space is small enough (with 'only' 5×10^{20} possible board positions) to enable the game to be solved; see Schaeffer et al. (2007).[23]

Given the strong evidence that Simon and others have offered that shows that unbounded rationality is often a bad assumption, why do IRL and SE continue to rely on it so heavily? After all, what is the practical value of inferring the objective function of an unboundedly rational agent if that agent already knows their objective function and their environment and is already perfectly capable of solving their own decision problem? Indeed, in light of the

[21] See DellaVigna (2018) for a growing literature on *structural behavioural economics.*

[22] The ELO score is a measure of relative playing ability in zero-sum games such as chess, named after its creator, Arpad Elo. Alliot (2017) is an empirical study of 26,000 chess games played by masters using the Stockfish computer program to evaluate mistakes made by human chess masters. The study demonstrates that 'conformance' of human moves with those chosen by Stockfish results in a measure of 'playing ability' he calls *conformance.* 'I have demonstrated along this article that conformance was highly correlated with the game outcome, and we know that computer programs are currently much stronger than human beings.' (p. 40).

[23] We already knew from Zermelo's 1914 theorem that in perfectly played two-person alternating move, zero-sum games of complete information with no chance element, the only possibilities are that either player can force a draw or otherwise one of the two players will always win.

discussion of the identification problem and the ability of 'mother logit' to rationalize any type of observed behaviour, isn't there a danger that IRL and SE could become almost tautologically meaningless exercises? The evident unwillingness of so many researchers in SE to confront these questions and devote more attention to practically oriented applications similar to those studied in ML is the reason why we characterized SE as being more 'academically oriented' in Section 2.

In defence of SE, it is important to recall that the mathematics of DP and game theory were developed as idealized models of human strategic behaviour, for example as in von Neumann and Morgenstern (1953), for which unbounded rationality is a reasonable starting point, similar to the assumption of a 'perfect gas' underlying Boyle's law in physics or the idealized but highly abstract Arrow–Debreu model of 'perfect competition' in economics. As subsequent work on game theory and dynamic models in economics has shown, even if the individual agents in an economy display unbounded rationality, it does not necessarily imply that the market equilibrium is desirable, because of externalities, information frictions and exercise of market power. Thus, SE can have practical value for policy-making even if unbounded rationality holds at the individual agent level.

One of the major intellectual contributions of SE is to calculate measures of 'market efficiency' or social welfare and use these as metrics to guide policy analysis, a key component of a scientifically based theory of regulation and market design. For example, Igami and Uetake (2020) provided a structural dynamic analysis of endogenous technological innovation, investment, mergers and consolidation in the hard disk drive industry: 'Our counterfactual simulations suggest the current rule-of-thumb policy, which stops mergers when three or fewer firms exist, strikes approximately the right balance between pro-competitive effects and value-destruction side effects in this dynamic welfare trade-off'.[24] There is also steadily accumulating evidence that SE models, even under the unbounded rationality assumption, can provide accurate counterfactual predictions of the behavioural and welfare effects of a range of policy changes. For example, there is an increasing use of validation studies using controlled human experiments where structural models accurately predict the behavioural changes for the treatment group using only data on the control group to estimate the model and make the counterfactual predictions; see for example the monograph by Wolpin (2013) and the references therein.

But if unbounded rationality does not always hold, it seems to open up a much wider scope for practical application to policy-making, such as to advise firms on more profitable business strategies, as Simon (1992) suggested. To use his terminology, instead of imposing the assumption of unbounded rationality as the basis for *positive theories* of economic behaviour, relaxing it paves the way for *normative theories* of how boundedly rational agents 'should behave'. The success of *Alpha Zero* in board games and the success stories of how DP has improved decision making in business applications recounted in Rust (2019) suggest that not only is this more practical normative orientation possible, but there is tremendous scope for its application to a range of real-world settings. This is the practical orientation that Simon (1992) suggested, but which for whatever reason was relegated to *operations research* and not considered a part of mainstream economics: 'So remote were the operations researchers from the social science community that economists wishing to enter the territory had to establish their own colony, which they called "management science".' (p. 350).

[24] SE analyses *do* affect thinking and policy-making, as evidenced by the fact that this paper was specifically cited in the chapter 'Evaluating the risk of declining competition' of the Economic Report of the President, CEA (2020).

8. ML AND SE UNDER BOUNDED RATIONALITY

Of course, the researchers doing ML and SE are boundedly rational themselves, and it is important to avoid hyping the potential benefits of these tools to assist real-world decision makers. There is a big gap between developing algorithms that can play better chess and developing algorithms that can, for example, calculate more profitable pricing strategies for hotels and airlines. There are two key limits to our own rationality: the curse of dimensionality, and the inferential problem of learning the decision maker's objectives and their environment.

Our primary way of dealing with the curse of dimensionality is to side-step it via the *principle of decomposition* originally proposed by Dantzig and Wolfe (1961) for linear programming: under certain circumstances, it is possible to approximately decompose a large overall decision problem into a set of independent, smaller and more tractable subproblems. Thus, rather than use DP to try to solve the *overall* problem of an individual or firm, we focus on the less ambitious goal of trying to assist them in solving some of their smaller, easier subproblems. For individuals, a good example is the computer GPS algorithm that helps them get to their destinations faster. For firms, examples include providing recommended prices, or advising a steel wholesaler on how much inventory to carry and when to restock individual products, or advising a rental firm on when to replace its equipment. Decomposition works when the objective function is additively separable, and where there is sufficiently weak interaction between the subproblem in question and the other problems the firm must solve. Additional examples of approximate decomposition are noted below.

The other obstacle to the normative application of DP is the inferential problem of learning about the firm's objective and the environment in which it operates. As we noted, this is easier for board games such as chess since the payoff function and the 'environment' (that is, the rules of the game) are known. But for individuals and firms, both the objective and the environment are typically unknown or incompletely known. For example, predicting the payoff to different airline pricing strategies depends critically on knowledge of *consumer demand*, including *differential willingness to pay* by different types of consumers. It also depends on the pricing strategies of competing airlines and alternative modes of transportation. These objects cannot automatically be 'learned' by ML algorithms because they require models capable of making accurate *counterfactual predictions*: what would demand have been if the airline had chosen higher or lower prices than it actually charged?

There are two key ways we learn about our environment to improve our decision making: (a) we make assumptions and build models (hopefully good ones), and (b) we gather data and undertake experiments. These two modes of learning interact, and data and experiments often reveal flaws in our assumptions, causing us to revise our models and beliefs so that they are more in line with reality. Presumably this process leads to better decisions and higher payoffs.[25] In the section, we will discuss how SE deals with inferential problems that naive application of ML largely ignores. Before doing this, however, it is useful to briefly review how an unboundedly rational decision maker combines decision making and inferential learning. This is a very difficult topic, and not much is known about how it works beyond specific cases, but Kiefer and Nyarko (1989) and Nyarko (1991) are two important contributions that give us insights into how it is done.

Their studies can be framed as a problem of a firm that is a Bayesian decision maker that chooses a pricing strategy for its product with incomplete knowledge of its (linear) demand curve. One

[25] Mao and Zheng (2020) show how even a misspecified structural model can lead to more accurate predictions when additional data become available.

of the key insights of their analysis is the tradeoff between *exploitation and experimentation.* Given the firm's prior belief about its demand curve, there is an optimal price it should charge to maximize its profits given its prior beliefs. However, the firm can vary its price to gain further information about the intercept and slope of its demand curve, but this experimentation is costly as it involves deviating from the initial prior belief about the optimal price. Kiefer and Nyarko (1989) consider the problem of discounted expected profit maximization, where the posterior belief about the demand curve is the firm's state variable in the DP problem. This fully endogenizes learning and decision making. They prove that the posterior beliefs converge with probability 1, but not necessarily to a unit mass on the true demand curve. Firms that are sufficiently impatient choose to not engage in sufficient price experimentation and never fully learn their true demand curves. Note that from Simon's perspective this could be interpreted as a form of *satisficing,* yet the overall strategy is fully optimal in a dynamic sense, and thus corresponds to learning and decision making with unbounded rationality. The important general conclusion is that *it is not always worthwhile for an unboundedly rational firm to fully learn its environment, and this incomplete understanding is costly since it leads to suboptimal decisions relative to a 'first best' world in which the firm did completely know its environment.*[26]

By and large, most empirical work in SE assumes that the agents (firms or individuals) have full knowledge of their preferences and environment,[27] and focuses on the inferential problems *we* face as econometricians, assuming away the learning problems that most real-world agents face. This is understandable to some extent, since except for special cases where the Bayesian updating formula can be expressed in terms of parametric families called *conjugate priors*, the posterior distribution will be an arbitrary probability distribution, and thus solving the DP involves solving a problem with an infinite-dimensional state variable. Even if we approximate this with a high-dimensional state variable, the curse of dimensionality typically makes it prohibitively difficult to solve most Bayesian DP problems numerically.

In the remainder of this section, we return to the bounded rationality viewpoint, where the agents we study are not necessarily behaving optimally. Our objective, taking the viewpoint of 'policy-making in the small', is to learn enough about their decision problem to be able to solve it ourselves using DP, using it as an 'actor–critic' algorithm to provide recommendations to the agent on better courses of action. However, we are unable to solve most problems as Bayesian DPs. The reason is not only the computational problem noted above, but also because few real-world agents trust us to undertake controlled experiments in the environments in which they operate. Therefore we adopt a second-best strategy that is the one taken by the literature on *adaptive control*—namely a two-stage approach of (a) first learn about the agent's objective and environment using historical data on the agent, and then (b) solve the DP using the estimated structure for an optimal decision rule that constitutes our recommended 'policy advice' to the agent. There is a crucial third step: using experimentation to improve and update our model in cases where the agent in question is willing to allow it. In the next section we discuss how much can be learned about the agent and their environment using only *non-experimental data.*

[26] When there is incomplete learning of the demand curve, Kiefer and Nyarko (1989) show that the firm stops experimenting and chooses a limiting price for its product that maximizes its expected profits with respect to the limiting belief over a *partially identified* set of demand functions that are observationally equivalent to the true demand function at the fixed limiting price the firm sets. Nyarko (1991) shows that if there are independent shocks to the demand curve that the firm observes prior to setting its price each period, then if the cost function is nonlinear, the firm will generally fully learn its demand curve and thus asymptotically set a fully informed or 'first best' profit maximizing price.

[27] The papers that explicitly models firms as Bayesian learners include Abbring and Campbell (2005), Crawford and Shum (2005), Hitsch (2006), and Doraszelski et al. (2018). A recent review of this literature is provided by Aguirregabiria and Jeon (2020).

9. CHALLENGES TO LEARNING FROM NON-EXPERIMENTAL DATA

As human modellers, we have the ability to make relevant simplifying assumptions that can reduce some of the inferential challenges. For example, if we are studying firms it is often the case that the reward or utility function $u(x, d)$ is not an 'unknown' that needs to be inferred. Similar to board games (where the objective function is known a priori, namely maximize the probability of winning), many firms would probably agree that their objective is, at least approximately, to maximize the expected discounted profits.[28]

To solve the firm's DP, however, we still need to have knowledge of its *environment*, which is embodied by the transition probability $p(x'|x, d)$ relating its current state and decision (x, d) to future states. Note that we previously referred to p as the decision maker's *beliefs*, but here we will impose the assumption of *rational expectations* and treat p is an *objective probability measure* under the assumption that the decision maker will do better if they have a correct understanding of reality as opposed to a mistaken belief about it.[29] Even under the assumption of rational expectations, it is still a challenging econometric problem to estimate $p(x'|x, d)$, which along with β is required to solve the firm's DP problem even in cases where $u(x, d)$ is known.

A huge challenge to inference or 'learning' about $p(x'|x, d)$ using non-experimental data stems from the fact that the data observe is typically *endogenous*(or what statisticians called *confounded*), another inferential problem that most of ML ignores. Suppose the agent's choice d is governed by some possibly suboptimal decision rule $d = \delta(x)$. Even if we perfectly observe (x_t, d_t), there may not be sufficient variation in d for each value of x to be able to predict $p(x'|x, d)$ for values of $d \neq \delta(x)$. For example, if d is a home-seller's decision on how to set the listing price, where x includes observed characteristics of the house and outcomes such as whether it is actually sold and how much it sells for, we may not observe homeowners setting unusually high or low prices. Thus, we may not have enough variation in the data to predict whether there are counterfactual listing price strategies that could enable the home-seller to sell faster, or for more money.[30]

For dynamic discrete choice models, the additive separability, conditional independence and extreme value assumptions that we discussed in Section 2, combined with the assumption of

[28] Of course, the discount factor β is still an important unknown, although there are promising new methods for identifying it; see Abbring and Daljord (2020). We have also abstracted from questions of risk, whether the firm is a public or private firm, and whether the stock market valuation of the firm equals the 'fundamental value' (that is, the expected discounted value of future dividends). Empirical evidence provided by Campbell and Shiller (1988) suggests that the latter does not hold, and work dating back to Lintner (1956) provides substantial evidence of 'dividend smoothing' behaviour that is inconsistent with the objective of maximizing expected discounted dividends. For private firms, the owner's degree of risk aversion comes into play, so the optimal strategy for such firms does require a knowledge of their utility function $u(x, d)$, see Gupta and Rust (2018). When it comes to individual decision making, another literature, which we do not include in this review, builds on alternative models of preferences, such as the recursive preferences of Epstein and Zin (1989), or the present biased preferences of Laibson (1997) and O'Donoghue and Rabin (1999); see Chan (2017) and Kemptner and Tolan (2018) for recent applications.

[29] In certain situations, the assumption of rational expectations has been shown to fail. For example, in a structural estimation of a life-cycle model of female labour supply, Schneider (2020) is able to identify the magnitude of the overestimation of job arrival probability after a spell of non-employment using German data from 1986 to 2006. The perceived probability of having a job offer is about 66% higher than actual probability, leading to an increase of between 11% and 17% in the length of employment breaks related to child-bearing, and a decrease in labour earnings of between 12% and 16%. See Kubler and Scheidegger (2019), p. 1 who 'introduce the concept of "self-justified equilibria" as a tractable alternative to rational expectations equilibria in stochastic general equilibrium models with heterogeneous agents.'

[30] For example, in some markets, *underpricing* (for example, setting a list price below the sale price expected eventually) may be an effective strategy if it attracts sufficiently many buyers and a bidding frenzy that drives the ultimate sale price far above the low initial list price.

optimality, are sufficient to enable us to identify $p(x'|x, d)$. Under these assumptions, the optimal decision rule takes the form $d = \delta(x, \epsilon)$, where ϵ is a vector extreme value distributed 'taste shocks' that are *conditionally independent*; that is, ϵ_t is independent of ϵ_{t-1} conditional on x_t. These taste shocks have unbounded support so that there is a positive probability of observing the agent taking any feasible action $d \in D(x)$. For the parametric specifications $p(x'|x, d, \theta_2)$ discussed in Section 2, the parameters θ_2 can be consistently estimated by partial likelihood, and it is also possible to consistently estimate p nonparametrically as well. Note that the assumption of optimality is not essential to our ability to estimate p: as long as the agent behaves according to any decision rule $d = \delta(x, \epsilon)$ that provides a positive probability of observing any feasible choice $d \in D(x)$, we can identify p: the key assumption is the conditional independence of the unobserved variables ϵ.[31]

Complications arise when we relax the conditional independence assumption and allow for the possibility of serially correlated unobserved state variables in the model. For example, suppose the true environment that the firm is facing has the form $p(x', \epsilon'|x, \epsilon, d)$, where ϵ is a vector of serially correlated states that the agent observes but we do not observe. If we ignore this serial correlation and try to estimate a misspecified model of the form $p(x'|x, d)$, it creates something akin to an 'omitted variable bias' that can cause inconsistent estimates of the effect of decisions d on observed next-period states x' because the econometric model will try to use d to capture the effect of the serial correlation transmitted via the unobserved ϵ component of the state variable.[32]

The problem of endogeneity results in 'spurious causality', which will be discussed more in the next section, but the problem has long been understood in the SE literature, dating to the work at the Cowles Foundation on the structural estimation of linear simultaneous equation models. This early literature recognized that naive attempts to estimate demand curves can easily produce spuriously upward-sloping rather than the theoretically predicted downward-sloping relationships between price and quantity demanded. Solving the problem of endogeneity to make correct inferences about the environment is the essential first step before we can provide useful advice to help firms or other decision makers do better.

Among the arsenal of tools for dealing with endogeneity is the method of *instrumental variables*, which is effective for linear models, but there are many other forms of endogeneity, including biases resulting from noisy measurements and incompletely observed data, including problems of 'sample selection bias' and problems created by other forms of 'endogenous sampling'; see Heckman (1979) and Manski and McFadden (1981) for examples in static structural models, and Hall and Rust (2020) for an example of how the problem arises in the case of dynamic structural models. The econometrics literature has come up with ingenious solutions to many of the inferential problems associated with endogeneity and censoring/truncation of data, and shown how badly inferences can be affected when these problems are ignored.

However it is not always possible to use instrumental variables to solve endogeneity problems: these variables may not exist, or we may be interested in estimating nonlinear models, where the method is often inapplicable. An important literature on *control functions* introduced by Heckman and Robb (1985) deals with endogeneity problems in simultaneous equations models that arise in 'economic models of agent's optimization problems or of interactions among agents' by providing 'a convenient procedure to estimate one of the functions in the system using reduced

[31] The conditional independence assumption is analogous to a conditional independence assumption in the treatment effects literature known as the *unconfoundedness assumption*; see Rubin (1990).

[32] Imai et al. (2009), Norets (2009), and Reich (2018) have all provided Bayesian and classical likelihood estimators for dynamic discrete choice models with serially correlated unobservables, but all three studies rely on the maintained assumption of optimal behaviour by the agent.

form residuals from the other functions as additional regressors' (Blundell and Matzkin, 2014). The latter paper introduces a new property of functions called *control function separability* and 'provides a complete characterization of the structural systems of simultaneous equations in which the control function procedure is valid' (p. 271).

Another approach is to exploit the assumption of optimality, which combined with assumptions on the distribution of various unobserved *latent variables* is another way to deal with endogeneity and a variety of other econometric problems. For example, Merlo et al. (2015) investigated the home-seller's problem empirically: how to set the list price and what 'reservation value' the seller should accept to maximize the ultimate expected proceeds from selling their home. Despite a wealth of observed 'hedonic characteristics' of homes in their sample (e.g. square footage, quality of school district, etc.), Merlo et al. (2015) concluded that the unobserved characteristics ϵ of homes are the source of a strong endogeneity problem: homes with superior unobserved characteristics had higher list prices, received more offers, and sold for higher prices. Failing to account for this omitted variable endogeneity bias results in an obvious form of spurious causality: namely, the prediction that setting a higher list price will result in more offers and a higher final transaction price. Of course, this is not the case, as setting too high a list price will lead to the opposite result: few or no offers and an inability to sell the home.

By assuming that sellers use optimal selling strategies and by modelling the unobservable characteristics ϵ of homes parametrically (as log-normally distributed), Merlo et al. (2015) were able to solve the endogeneity problem without the use of instrumental variables via the method of simulated moments (MSM) introduced by McFadden (1989).[33] Using simulations of optimal selling strategies, the authors estimated the parameters of $p(x'|x, d)$ by finding values that minimize the distance (a quadratic form) between a vector of simulated moments and observed moments constructed from the data set. These moments include the trajectory of list prices for the home over time, the number of offers received, and the final transaction price. The optimality assumption forces increases in list prices to reduce the rate of arrival of offers, since a positive relationship would imply much higher list prices than we actually observe.[34]

Once we rely on the assumption of optimality, we lose the ability to provide normative advice to the home-seller. In effect, we assume that they do not need our advice, and our maintained assumption prevents us from testing whether their actual home-selling strategies are optimal or not. Can SE deal with endogeneity and other econometric problems that confound inference for the 'environmental objects' that agents face while also relaxing the assumption of unbounded rationality?

Examples such as Cho and Rust (2010), Misra and Nair (2011), McClelland and Rust (2018), Hall and Rust (2020), Barkley et al. (2020), and Cho et al. (2020) have shown that this is indeed possible. Due to the identification problem, when we weaken one assumption (in this case the assumption of optimality), we need to supply other identifying assumptions to infer the agent's environment. As Manski (2003) and others have shown, there are tough tradeoffs imposed by the identification problem, and weakening certain assumptions in an econometric model can cause other key structural items of interest to become either unidentified or only partially identified.

[33] For an alternative 'instrument-free' approach to demand estimation that uses the maintained assumptions of optimality and equilibrium, see Miller and MacKay (2019).

[34] Merlo et al. (2015) did relax the unbounded rationality assumption for buyers, modelling them as *automata* who arrive probabilistically and make offers to the seller from a probability distribution they estimated as part of the underlying 'environment' for their model. Modelling the buyers as fully rational would lead to problems of multiple equilibria, since the home-selling problem can be regarded as a bargaining problem with two-sided incomplete information between buyer and seller.

In the best case, there may be no serious endogeneity or data issues that confound our ability to infer the underlying environment. Cho and Rust (2010) and McClelland and Rust (2018) are examples: they study the optimal replacement of rental cars and rental machinery, respectively, using econometric duration models for the time that rental cars and machines spend in rental contracts and in the lot waiting to be rented, as well as their resale values and maintenance costs. Using these econometric estimates, the authors used DP to calculate optimal replacement strategies for the cars and machines and compared them with the replacement policies the firms actually used (which were estimated nonparametrically). They demonstrated that the firms were using suboptimal replacement rules and that by adopting the DP replacement strategy the firms' profits could be significantly increased.[35]

Econometric problems can also arise when we do not fully observe agents' decisions. We can, however, address these problems by treating unobserved decisions as latent variables, together with using other observed variables that are correlated with unobserved decisions as a basis for identifying $p(x'|x, d)$. For example, Misra and Nair (2011) studied the compensation schemes that a large contact lens producer offered to its sales employees. They relaxed the assumption that the company offers an optimal compensation plan, but maintained the assumption that its sales people behave optimally given that compensation plan. Sales effort is unobserved by the econometrician. It is natural to assume that sales (which are observed) are related to sales effort. By making reasonable parametric assumptions about this relationship (that is, a linear relationship between effort and sales), they were able to estimate both the disutility of sales effort, and the productivity of sales efforts on an agent-by-agent basis. Treating these preference and productivity parameters as *structural* (i.e., invariant to changes in the compensation scheme), they were able to design an improved compensation scheme, a modified form of which the company ultimately adopted. The new plan was a win-win for the agents and the company: most sales agents earned more, and company profits increased by \$12 million.

These examples illustrate that it is possible to relax the assumption of optimality and still be able to infer/identify the agent's environment, as represented by the transition probability $p(x'|x, d)$. This requires different types of identifying assumptions to replace the powerful assumption of optimality. In many cases, this requires the use of semi-parametric or highly flexible parametric estimators that involve estimating infinite-dimensional 'nuisance parameters' such as the agent's decision rule δ, in order to consistently estimate the 'parameters of interest' characterizing the agent's environment $p(x'|x, d)$. Although econometricians have been working on semi-parametric estimation for decades, and hundreds of different estimators for various types of problems have been proposed and studied, the ML literature has reinvigorated this area by offering even more choices of flexible estimation methods that can help us develop new estimators in SE that relax strong assumptions such as optimality, and other parametric functional form assumptions.

For example, Druedahl and Munk-Nielsen (2020) (in this issue) show how ML-inspired methods can provide much greater flexibility in capturing richer patterns of income dynamics compared with traditional parametric approaches to specifying the law of motion $p(x'|x, d)$ that are often used in the SE literature. They introduce a flexible model of income dynamics that they call

[35] There are issues of endogeneity relating to how companies set the rental rates for cars and machines. The authors of both papers noted what they call the *flat rental puzzle:* namely, rental rates do not decline with the age of the car or machine. To evaluate counterfactual replacement policies that keep cars and machines longer than the firm currently keeps them, the authors relied on assumptions provided by the firms about how much rental rates would have to be reduced to induce customers to rent the older equipment. It seems that profits could be increased further if the firms estimated structural econometric models of customers' willingness to pay for rental equipment of various ages. By raising rental rates of new equipment and discounting rates charged for older equipment, it seems likely the firms could significantly increase their revenues relative to their flat rental rate schedules.

linked regression trees (LRTs), which are inspired by regression trees in the statistical and ML literature. They show that LRTs can capture higher-order patterns of dependence, skewness and kurtosis without the need for parametric assumptions.

Another example is Hall and Rust (2020), who use flexible semi-parametric estimators in their study of the purchasing, sales and inventory holding decisions of a firm in the steel service-centre industry. What these firms do is simple to describe: they buys large quantities of steel in the world market at wholesale prices and sell smaller quantities to retail customers at a markup. The 'parameters of interest' in this case are the parameters of a Markovian model of the stochastic process for wholesale steel prices. Knowing this process is crucial for advising the firm on how to optimally 'buy low and sell high' by strategically timing its purchases. This problem would be simple if we could observe daily wholesale prices traded on some centralized market, say the New York Mercantile Exchange. But, steel is almost entirely traded in a search market involving private, bilateral negotiations between buyers and sellers. Thus there is no market where we observe wholesale prices of particular steel products on a daily basis. Instead, the data are *endogenously sampled*—we only observe wholesale prices on the random dates that the firm actually purchases steel, not on the days it does not buy steel (which constitute more than 80% of the days in our sample).

This irregular, endogenous sampling of wholesale prices makes it very challenging to infer the true underlying stochastic process. Hall and Rust (2020) showed how it can be done using MSM, with simulations of the firm's strategy δ that determines the timing and quantities of steel it purchases. The key idea is to sample the simulated data endogenously in exactly the same way the actual data are sampled, namely by recording purchase prices and quantities only on days when simulated purchases occur. It is possible to do this while relaxing the assumption of optimality, but this introduces infinite-dimensional nuisance parameters consisting of a conditional probability governing the decision of whether or not to purchase more steel (given the current price and quantity of steel held), and a separate conditional probability density determining the quantity of steel purchased. These objects can be estimated nonparametrically by a variety of flexible parametric estimators called *sieves* (see Kristensen et al., 2020 for examples), which can include methods that are popular in ML such as deep neural networks. After using these methods with MSM to estimate the parameters of interest, Hall and Rust (2020) then used DP to solve for the optimal purchasing strategy, which they show has the form of an (S, s) inventory rule. They compared counterfactual predictions of expected profits under an (S, s) rule with the profits the firm actually earned during the sample period. Depending on the product, discounted profits were predicted to increase by 5% to 27% under the optimal (S, s) purchasing rule.[36]

A final example is Cho et al. (2020), who analysed the pricing decisions of a hotel operating in a particular market of a major American city. Hotel prices vary widely over different seasons and holidays, given the finite capacity constraints of the local hotel market. Variability in demand relative to this fixed capacity leads to a spurious upward-sloping inferred demand for hotels owing to the fact that hotels raise their prices to ration scarce capacity on days when demand is high, and lower their prices significantly on days when demand is low. A further problem is that the number of customers 'arriving' to this market is unobserved, as is the number of customers who book

[36] This study is also a good illustration of the principle of decomposition. The firm that Hall and Rust (2020) investigated carried over 9,000 individual steel products. Each product requires at least two continuous state variables: the current price p_t and the quantity held in inventory q_t. Thus the overall problem of the firm is a DP with over 18,000 state variables. However, if there are no aggregate storage space constraints or financial constraints, Hall and Rust (2020) show that the firm's problem decomposes into the solution of 9,000 two-dimensional DPs, which is far easier to solve than a single 18,000-dimensional DP.

at competing hotels. Furthermore, the hotels do not observe the number of arriving consumers choosing the 'outside good' , namely to not stay at any of the hotels in this local market. These censoring/truncation issues preclude the use of standard methods for demand estimation such as the widely used BLP estimation method of Berry et al. (1995), because there is not enough information to construct the market shares that this method relies on. A second problem is that there are no relevant demand or supply shifters that could serve as the exogenous instrumental variables needed by the BLP method.

Instead, Cho et al. (2020) make a conditional independence assumption similar in spirit to the unconfoundedness assumption of Rubin (1990), which is widely used in the literature on 'treatment effects'. This assumption states that, conditional on observed covariates x, there are other variables z, exogenous 'price shifters', which affect how the hotel sets its prices but do not affect demand. For this reason they refer to it as a *conditional exogeneity assumption*; that is, prices can be treated as econometrically exogenous conditional on a set of observed covariates x that serve as 'sufficient statistics' along with price for determining demand. Although conditional exogeneity is a strong assumption, Cho et al. (2020) show that it can be complemented with latent variable modelling approaches in cases where we do not directly observe the variables x underlying the conditional exogeneity assumption but may only have noisy proxies for them. In their empirical illustration, they demonstrate that this assumption combined with flexible mixture modelling results in plausible downward-sloping estimates of demand for hotels in this market. Using the identified demand, the authors are able to test various hypotheses, such as the hypothesis that the hotel in question is setting profit-maximizing prices, and the hypothesis that the hotels in this market are engaging in *algorithmic collusion* coordinated by the computerized revenue management systems they subscribe to. They strongly reject both hypotheses.

The estimator that Cho et al. (2020) constructed, using both regression and maximum likelihood approaches, takes the form of a mixture model owing to the fact that a key explanatory variable driving hotel prices and occupancy rates—the number of consumers arriving to the market A—is unobserved. They showed that it is possible to nonparametrically identify the conditional distribution of expected occupancy A given an observed demand shifter x, and uncover a separate underlying structure for unobserved consumer heterogeneity that is key to determining differential willingness to pay for hotels. The resulting estimator is a microeconometrically motivated model that resembles the deep neural network models used in the ML literature, but the latter are often proposed without any theoretical justification or attempt to identify the underlying structure of arrivals and preferences; instead, the deep nets are promoted as flexible functional forms that 'can produce superior predictive accuracy as compared to a standard linear regression or logit model' (Bajari et al., 2015, p. 484).

Can ML, with its focus on predictive accuracy, removing the role of human oversight, and neglecting data and endogeneity problems, nevertheless still do better? The study by Keane and Neal (2020) in this issue suggests probably not. They analyse the effect of climate-induced weather changes on agricultural yield using deep neural networks (DNN) to estimate highly nonlinear agricultural crop production functions with a high-dimensional set of daily weather inputs. They find that, although ML methods such as DNNs do produce near perfect in-sample fits and show no sign of overfitting, a key limitation of DNNs is their inability to make accurate counterfactual predictions in different environments. Using a hold-out sample to consider the model's ability to forecast counterfactual weather changes that have no histrionical precedent, they find that the out-of-sample performance is surprisingly poor. Yet, they demonstrate that DNNs can be a powerful tool when used in tandem with SE. Specifically, they show how a DNN discovered patterns in the data that helped them to develop new and improved structural panel data models with superior out-of-sample performance. The resulting SE model outperforms both DNNs and traditional state-of-the-art panel data models.

So far, we have largely taken a 'single-agent' focus and considered how to relax the assumption of unbounded rationality while still being able to infer enough of the underlying structure, namely the environment as captured by $p(x'|x, d)$. A related rationality assumption typically imposed in strategic interactions is the hypothesis of *Nash equilibrium*, i.e., that agents' behaviour is governed by strategies that are mutual best responses. However, Nash behaviour presumes the ability that agents can optimize, so it may be an even more questionable assumption in many circumstances. Nash equilibrium is especially problematic in settings where there can be multiple equilibria: how do the agents coordinate on one of them? In a model of Bertrand price competition where firms can make cost-reducing investments, Iskhakov et al. (2018) show that even a simple, small-scale, dynamic extension to the classic Bertrand pricing game can result in hundreds of millions of Markov perfect equilibria. Apart from the daunting task of solving dynamic games with such a huge multiplicity of equilibria, they raise serious questions about identification and inference even if we maintain the Nash equilibrium assumption. But what happens if we relax it?

Aguirregabiria and Magesan (2020) showed that with an additional exclusion restriction it is possible to test whether players' beliefs represent the probability distribution of the actual behaviour of other players conditional on the information available, and thus to test the Nash equilibrium assumption. In an empirical application based on a dynamic game of store location by retail chains, they find significant evidence that players' beliefs are biased and that the assumption of unbiased beliefs can result in substantial attenuation bias in the estimate of competition effects.

In the previous section, we provided an example of one game, chess, where we can convincingly dismiss the unbounded rationality hypothesis that human chess champions play Nash equilibrium strategies. In chess, there is a premium on the ability to 'look ahead' to consider the future consequences of future moves, and computer strategies such as *Alpha Zero* demonstrate superiority over humans in this dimension. However, what about simpler games such as tennis, where the ability to look ahead many stages (versus focusing on winning each point) is not important to success in the game? Although Walker and Wooders (2001) claim that play by tennis professionals is consistent with the Nash/minimax equilibrium prediction, a reanalysis of a much larger data set by Anderson et al. (2020) leads to the opposite conclusion: they construct alternative serve strategies that significantly improve a server's chances of winning the game, contrary to the Nash equilibrium prediction that a server should be indifferent to serving to any particular direction (e.g., serving wide, down the T or to the receiver's body). So Nash equilibrium may fail to predict behaviour even by experts who play games that are 'mentally simple'. See also Camerer (2003) for numerous other examples where human behaviour is inconsistent with the predictions of Nash equilibrium and the logic of backward induction.

The bottom line of this section is that inference with non-experimental data provides a host of challenges, especially when we try to relax the assumption of unbounded rationality, including the twin hypotheses that human behaviour is consistent with backward induction and Nash equilibrium. We have illustrated that there are solutions to a variety of econometric endogeneity and data problems under bounded rationality, but they are highly case-specific, and require a lot of human insight/understanding and cannot just be solved by the mindless application of 'smart machines'. At the same time, the ML methods provide a set of promising new tools for SE that can reduce the need for arbitrary parametric functional form assumptions. The most promising path forward is via human/machine collaboration: using ML to automate parts of the inferential process but overall still being guided by human insight and intuition. That is, ML will not make SE irrelevant any time soon.

10. CAUSALITY, COUNTERFACTUALS AND CONTROLLED EXPERIMENTS

This section was added at the suggestion of the editor, Jaap Abbring, who commented:

> 'You note that ML by itself is not good in dealing with endogeneity and causality, which may require structural modeling. Judea Pearl, one of the pioneers of AI, has essentially been making the same point and proposed a way forward using structural directed acyclical graphs (recursive nonparametric versions of the Structural Equation Models of early econometrics and social science) as structural models. He has been quite adamant that economists have all but forgotten the early lessons of Haavelmo and failed to pick up on this new "causal revolution" at their own peril.'

Abbring correctly represents Pearl's views, which are stated even more forcefully in Pearl (2018):

> 'Current machine learning systems operate, almost exclusively, in a statistical, or model-free mode, which entails severe theoretical limits on their power and performance. Such systems cannot reason about interventions and retrospection and, therefore, cannot serve as the basis for strong AI. To achieve human level intelligence, learning machines need the guidance of a model of reality, similar to the ones used in causal inference tasks.'

While we agree with Pearl's view, it is qualified by the term *strong AI*, which refers to a type of general intelligence, similar to what we possess. However, with respect to more narrowly defined tasks such as board games—the domain of *weak AI*—ML has achieved superhuman performance without having an underlying complete 'model of reality'. In this article we have focused on these narrowly defined problems, and raised the question of how much further beyond board games the 'weak AI' of ML can be successfully extended without the active collaboration and guidance of the 'strong AI' provided by human expertise and judgement. Pearl (2019) also raised the question of robustness: if the environment that an actor–critic algorithm is 'expecting' changes sufficiently, will its performance break down and result in counterproductive forecasts and recommendations? Strong AI is presumably robust, but weak AI may be 'brittle' and thus untrustworthy unless it is constantly reviewed and supervised by human judgement.[37]

Unfortunately, we do not have the space to do full justice to Pearl's work and the vast literature on causal inference, another discipline with its own terminology and intellectual barriers to entry that deter many of us from benefiting from its insights and different perspectives. Similar to 'learning', the term 'causality' is an overloaded term, in part due to the well-worn phrase 'correlation does not imply causality' and its imprecise use in econometrics, such as by Granger (1969), who defined causality as a type of correlation in time series. This is why we attempted to avoid the term in the previous draft of this paper: in the context of our remarks here, we only need a very modest definition of 'causality'—for us it is equivalent to knowledge of the true transition probability $p(x'|x, d)$ that captures how different decisions affect the future distribution of states, x', given the current state x. Knowledge of preferences (β, u) and the environment p is sufficient to determine an optimal decision rule δ, which is the primary goal of the more practically oriented perspective on improving decision making that we have taken here.

Our more modest definition of causality as knowledge of $p(x'|x, d)$ does typically capture an intuitive notion of cause and effect. For example, if d represents a decision of whether to retire

[37] Holland (1986) used the term 'brittleness' to express the lack of robustness of special-purpose AI methods that do poorly and are unable to adapt whenever the environment deviates from what the algorithm was designed to expect. In the DP literature, there is a growing literature on 'robust DP under ambiguity' whose performance is less sensitive to misspecifications or changes in the environment. See Blesch and Eisenhauer (2020) for an application of this to the bus engine replacement problem in Rust (1987).

and x includes the person's earnings, then the decision to retire causes earnings to fall to zero, while a decision to collect Social Security benefits causes pension benefits to become positive. In more complex cases, say if d represents the decision to smoke and x includes an indicator for a heart attack, the notion of causality is more complicated, since other things besides smoking can cause a heart attack. For most decisions it is sufficient to have a probabilistic belief of how an action such as smoking increases the risk of negative outcomes such as a heart attack, even though this reflects an incomplete understanding and inability to fully control the underlying causal mechanism. Most of science, even quantum mechanics, is based on this type of incomplete probabilistic understanding of reality.

Thus, while we see a great deal of kinship between the notions of causality in Pearl (2009) and SE, particularly his distinction between *predictions, interventions*, and *counterfactuals*, our interest is more in improving decision making rather than in providing a satisfactory general definition of what causality is. Pearl's terminology is similar to that of SE, including the *structural causal model* (SCM) defined in terms of Bayesian networks and directed acyclic graphs (DAGs) that restrict the pattern of dependence among collections of random variables. Pearl's definition was inspired in part by early work in SE where the DAG structure is equivalent to triangularity or 'recursive' structure for linear simultaneous equation systems; see Simon (1953) and Strotz and Wold (1960).[38]

It is less clear whether DAGs are useful ways of *discovering* causal structure versus *illustrating* assumptions used to infer that structure. For example, figure 1 of Rust (1988) used a DAG to illustrate the pattern of dependence in the controlled process $\{x_t\}$ implied by the conditional independence assumption on the unobserved state variables $\{\epsilon_t\}$; however, as we noted in Section 8, our ability to infer $p(x'|x, d)$ is consequence of the assumption, not the DAG that illustrates it. Even though we have found DAGs to be quite useful in analysing and computing equilibria of a class of games we call *directional dynamic games* (DDGs, see Iskhakov et al., 2016), it not easy to see how to apply Pearl's abstract SCM approach in specific empirical applications. The benefits of Pearl's general way of thinking about causality may become clearer from concrete empirical applications such as Imai and Kim (2019), where it is easier to see how DAGs contribute new insights into the identification of causal structure.[39]

An alternative approach to the identification of causal structure is the 'Rubin Causal Model' (RCM) (also known as the Neyman–Rubin Causal Model, see Rubin, 1974, 1990) based on the analysis of *potential outcomes* inspired by classical randomized controlled experiments (RCEs).[40] One of the key identifying assumptions in the RCM is a conditional independence (or *unconfoundedness* or *strong ignorability*) assumption that states, in effect, that conditional on observed covariates x, the assignment of the treatment is random; that is, potential outcomes are independent of the treatment assignment, so the data can be considered to have come from a 'virtual RCE' conducted for each value x. Of course, decisions made by a firm or individual are not 'treatments' and are endogenously determined rather than randomly assigned. There are also multiple outcomes from any given decision as captured by the conditional probability $p(x'|x, d)$. However, as we noted in Section 8, a conditional independence assumption on the unobserved

[38] Triangularity has been extended to nonlinear simultaneous equations systems; see Newey et al. (1999).

[39] Imai and Kim (2019) consider whether researchers should use fixed effects regression models to infer causal inference using panel data, and use DAGs to highlight 'two key causal identification assumptions that are required under fixed effects models and yet are often overlooked by applied researchers: (1) past treatments do not directly influence current outcome, and (2) past outcomes do not affect current treatment.' (p. 468).

[40] RCEs are also called randomized controlled trials (RCTs), and in the machine learning literature the approach is known as 'A/B testing'.

factors ϵ affecting the agent's decisions can be interpreted as the virtual random assignment of decisions according to the conditional choice probability $P(d|x)$. It can be viewed as an analogue of the unconfoundedness assumption, and when it holds it is sufficient to identify the causal structure captured by $p(x'|x, d)$, as shown in Rust (1987; 1988).

Despite the attempt by Pearl (2009) to unify different ways of thinking about causality using DAGs, he is critical of the two main approaches that are used to identify it: (a) the SE or 'path analysis or structural equation modeling (SEM)' dating back to Wright (1921) and Haavelmo (1943); and (b) the RCM or 'Neyman–Rubin potential-outcome model' of (Rubin, 1974; Splawa-Neyman et al., 1990). He notes that these two languages are mathematically equivalent yet concludes that neither approach has become a standard in causal modelling: 'the structural equation framework because it has been greatly misused and inadequately formalized (Freedman, 1987) and the potential-outcome framework because it has been only partially formalized and (more significantly) because it rests on an esoteric and seemingly metaphysical vocabulary of randomized experiments and counterfactual variables that bears no apparent relation to ordinary understanding of cause-effect processes in non-experimental settings.' (pp. 134–135).

From our perspective, a pragmatic way to resolve these disagreements is via the adage 'the proof of the pudding is in the eating'. We believe that successful methods and assumptions for identifying causal structure will reveal themselves in empirical work, and ultimately in concrete demonstrations of how this knowledge helps individual decision makers make better decisions (i.e., improving decision making in the small). We believe that more successes on smaller-scale problems can pave the way for SE and SCM to be more useful and credible in larger-scale policy-making as well (i.e., improving decision making in the large). The study by Hitsch and Misra (2018) is one such example, which uses advances in causal inference (RCM approach) and ML to evaluate the profitability of target marketing strategies more cost-effectively using a single randomized sample of customers. They conclude that 'while machine learning methods may generally improve the profits from a targeting policy, a pronounced increase in profits only occurs when we use methods that draw on recent ideas from both the causal inference and machine learning literatures'.

Although Pearl (2018) argues that 'model-free' approaches such as ML are incapable of making reliable counterfactual predictions, we believe that ML can do this, at least in a certain restricted sense. Consider the case of board games: at any state x of the game, *Alpha Zero* has to decide which of many different alternative feasible moves d is the best. To do this, it uses an estimate of the Q-value (or, in the language of SE, its current estimate of the decision-specific value function) $v_t(x, d)$ to make this choice, taking the move $d = \delta_t(x) = \text{argmax}_{d' \in D(x)}\, v(x, d')$ that has the highest Q-value. So in this sense, *Alpha Zero* has to be able to make counterfactual predictions of the value (i.e., the probability of winning) of different possible decisions.

Q-learning and related methods from reinforcement learning (RL) such as *real-time DP* (Barto et al., 1995) can be characterized as *model-free learning algorithms*, since they are updating methods defined in terms of the Q-values, $v(x, d)$, that bypass the need to identify and estimate the environment $p(x'|x, d)$. Note from Equation (3.3) that $v(x, d)$ is an implicit function of $p(x'|x, d)$ and it also constitutes a sufficient statistic for determining the optimal decision rule in Equation (3.4). Thus, as long as it is possible to train an algorithm using only *simulations from* $p(x'|x, d)$, the method will indeed be model-free in the sense that it is not necessary to actually estimate and identify p itself. This is relevant for training done in real time, since 'nature' provides the necessary simulated realizations from $p(x'|x, d)$ that Q-learning needs to update $v(x, d)$. In the process, the method also produces (in real time) counterfactual predictions of the consequences of alternative decisions, at least in the restricted sense discussed above.

There is a larger sense of counterfactual prediction in SE that is relevant to policy evaluation, namely, changes in policies that affect the solution to the decision maker's DP problem that imply a different decision rule δ_Π that is optimal in a counterfactual world where there is a new policy/environment denoted by Π that affects either the decision maker's utility function u or the environment p itself. A real-world example might be a change in tax policy that affects an individual's willingness to work, or a firms' willingness to hire them. In the board game example, we can imagine trying to predict how the play of a game might be affected if we were to change the rules of the game, such as allowing pawns to move backwards instead of only forwards. Can ML handle these sorts of counterfactual prediction exercises? A simple thought experiment suggests that, at least for the board game example, the answer is yes. We can imagine re-doing the Q-learning exercise, but instead of under the existing rules of chess, we impose the counterfactual new rules Π. Then we train the algorithm the same way as it was trained under the current rules of chess, resulting ultimately in its learning (i.e., converging to) a new strategy δ_Π that is optimal under the new rules Π.

The important caveat to this conclusion is that the model-free methods we are aware of require a simulator to train the algorithm. The simulator could be 'nature' or it could be (in the case of *Alpha Zero*) a computer simulation. Using 'nature' as a simulator is not very desirable in many real-world applications, since the performance of a randomly initialized Q-learning algorithm is likely to be pretty poor, and thus the cost of training in terms of the number of mistakes the algorithm will make early on while it still has much to learn will probably be unacceptably high. Typically, hundreds of thousands or millions of iterations or training instances are required before the algorithm starts to reach an acceptable level of performance. It is hard to think of many real-time applications where the individual or firm could afford to be so patient and wait for so long for the algorithm to be trained.

Of course Google trained *Alpha Zero* via computer simulations. Although the training occurred in a matter of hours in terms of wall-clock time, millions of games were played on its massively parallel network of computers, something that would take eons if done at the speed that human chess players make their moves. So, in effect, the success of model-free RL algorithms such as Q-learning depends *on the existence of very accurate and rapid simulators of the environment*, that is, the ability to draw realizations from $p(x'|x, d)$. While it is possible to do this for board games, as we noted in Section 4 where we discussed the work of Powell et al. (2014) regarding the use of approximate DP methods for locomotive scheduling at Norfolk-Southern, perhaps the most time-consuming and difficult part of the exercise was the econometric problem of developing an accurate simulator of locomotive demand, travel, maintenance and breakdowns, embodied in their PLASMA model. So we conclude that when it comes to applying ML to most real-world problems, methods such as Q-learning are not truly 'model free'—they require considerable human input and ingenuity and the methods of SE and the causal modelling literature to identify and estimate an accurate model of the environment $p(x'|x, d)$ that can serve as the simulation platform for training the model.

This brings us to the last topic of this section: the importance of *controlled experimentation.* This can be of huge value in three different ways: (a) *to learn/train better decision rules* as embodied by the decision rule δ, (b) *to learn about the environment* as embodied by the transition probability $p(x'|x, d)$, and (c) *to evaluate/demonstrate the accuracy of counterfactual predictions and the value-added of advice* from ML- and SE-inspired actor–critic algorithms.

We have already discussed item (a), but item (b) is worth further discussion because it has been used with great success for learning about causality/the environment in situations where non-experimental data are subject to endogeneity and different types of statistical confounding. The most often used type of controlled experimentation is RCEs, where members of a population are randomly assigned to treatment and control groups. The focus of RCEs is to assess the *treatment effect* such as the scalar 'average treatment effect', although heterogeneous treatment effects (i.e., ones that vary with the observed characteristics x of individual subjects) can be estimated with enough data as well. The study by Hitsch and Misra (2018) is an example where the treatment consists of mailing a catalogue to a subset of customers in the treatment group, and the treatment effect of interest is how much the catalogue increased the spending of customers who received the catalogue compared with those in the control group who did not.

Another important study, Dubé and Misra (2019), used data from an RCE to learn about demand (including heterogeneous willingness to pay by different customers) and improve the pricing decisions of the online job-posting platform ZipRecruiter.com.[41] The company conducted an RCE in September 2015 on 7,867 prospective customers (firms considering posting job ads on ZipRecruiter.com) who were randomly assigned one of ten different monthly prices for the service, including the flat \$99 per month charge that the company had set as its existing or status quo pricing policy. These firms can be considered to be randomly assigned to the control group, and the others were randomly assigned one of nine possible price 'treatments' ranging from \$19 per month to \$399 per month. The treatment effect of interest was the 'conversion rate', namely the binary outcome of whether the prospective customer signed up to use ZipRecruiter.com at the quoted price.

Dubé and Misra (2019) used the data to estimate a binary logit model of a prospective customer's decision to sign up for ZipRecruiter.com, where the price and intercept coefficients of the logit model are allowed to be arbitrary functions of a vector x of 12 covariates that were reported by the recruiter on the account signup form. An important aspect of their estimation is the combination of Bayesian estimation and ML methods, including LASSO as tool for model selection (to determine how the logit intercept and slope depend on x). This enabled them to rapidly simulate draws from the posterior distribution for the coefficients of the logit model given the customer's observed characteristics x via an innovative statistical approach they call the 'weighted likelihood bootstap' (WLB) algorithm. Using their econometric model of customer behaviour they can calculate either what they call 'personalized prices' (i.e., prices that depend on the customer's characteristics x), or an optimal uniform price that is the same for all customers.

They calculated the optimal uniform price to be \$327 per month, resulting in a 55% increase in expected profit over the company's existing \$99 per month price, even though the higher price reduces the conversion rate from 25% to 12%, so 'Ziprecruiter.com was pricing in the inelastic region of demand prior to the experiment' (p. 17). They also computed optimal personalized prices and found that these increase expected profits by an additional 19%, or an increase of 86% over the company's existing \$99 per month uniform price. To demonstrate that these counterfactual predictions could actually materialize, they conducted a validation study, using their estimated model to predict conversion rates from a second RCE conducted between October and November 2015. A total of 5,315 prospective customers were randomly assigned to one of (a) the control group, i.e., \$99 per month, (b) a price of \$249 per month, or (c) the optimal personalized price calculated from their econometric model, except that ZipRecruiter.com specified that no personalized price be higher than \$499 per month. Their model provided remarkably accurate predictions of conversion rates and expected revenue per customer in all three groups. For example, for the 2,485 customers who received the personalized price 'treatments', the actual

[41] A previous version of this paper was presented at the conference leading to this special issue at the University of Copenhagen in May 2018 under the title 'Scalable price targeting'.

conversion rate was 15% whereas their model predicted 14%. The expected gain in profits from switching from the $99 per month uniform price to personalized pricing was accurately predicted by their econometric model.

Thus, not only did Dubé and Misra (2019) use experimentation to identify consumer preferences (i.e., learn about the firm's environment), but they also used it in the third way we mentioned, i.e., to demonstrate the accuracy of their model's counterfactual predictions and confirm that their recommended prices would significantly increase the profitability of ZipRecruiter.com.[42] The power of the combination of experimentation and ML, both as a way to 'learn' the environment and as a way to validate and test the accuracy of counterfactual predictions and recommended decisions from models, leads us back to one of the original questions of our article: can this process of experimentation and decision making be 'automated', and does it obviate the need for human judgement in model selection and development that is the traditional hallmark of SE?

Although RCEs and ML often give many students a giddy feeling that these tools are so powerful that they don't need to invest in other tools or think too deeply about the problem at hand, we still see an overwhelming role for human judgement in many different aspects of the Dubé and Misra (2019) study. Human judgement was required in the design of the price experiment, such as how many treatment arms to use and at what prices. Human judgement played a key role in selecting the econometric model (a binary logit model), even though ML played an important role in helping to identify an agnostic yet parsimonious specification of customer-specific heterogeneity.[43] Furthermore, human judgement played a role in terms of the general form of pricing mechanism that the authors evaluated. We can imagine many other ways of pricing the services provided by ZipRecruiter.com: instead of a monthly flat fee, customers could be charged per ad posted, or per application received, and there could be pricing for add-on services such as collection of reference letters, or tools to help employers screen applications to identify promising candidates, etc. There is also a larger question of dynamic history-dependent pricing schemes, and questions of network externalities in job-posting sites that may make 'underpricing' a longer-run optimal strategy.[44]

Thus, even though we deem the Dubé and Misra (2019) study to be a major success by providing very convincing evidence that ZipRecruiter.com can change its pricing strategy and significantly increase its profits in the short run, it is less clear whether this is consistent with a longer-term optimization. This is another place where human judgement plays a key role: to the extent that higher short-term pricing is judged to also be in the long-term interests of the firm, it is consistent with the decomposition principle that we discussed in Section 8. But it is probably not feasible at present to design an optimal dynamic pricing strategy that takes all of the considerations noted above into account.

[42] Of course, theirs is not the first study to demonstrate the accuracy and credibility of SE models via out-of-sample predictive tests using RCE data. For example, Todd and Wolpin (2006) estimated a SE model for individuals in the control group of a RCE conducted in poor rural areas of Mexico, called Progresa. They re-solved the model to reflect a subsidy families in the treatment group received if they kept their children in school and compared the predicted behaviour to the actual behaviour of the treatment group, finding 'that the model's predicted program impacts track the experimental results' (p. 1384) and 'this evidence lent sufficient support to the model to use it to simulate the effects of a number of counterfactual policy experiments, which illustrate a menu of options that might be available to policymakers.' (p. 1408).

[43] See also Bonhomme et al. (2019), who propose a *grouped fixed effects* estimator that is a flexible two-step procedure for dealing with unobserved heterogeneity in panel data models based on a first-stage k-means clustering, a method often used in ML.

[44] Job sites such as ZipRecruiter.com can be considered 'platforms', and the literature on 'two-sided markets' has shown that complicated dynamic pricing strategies can emerge when there are competing platforms and the possibility of 'tipping effects'—see Dubé et al. (2010) for an analysis in the context of video gaming and Bandyopadhyay et al. (2013) for job platforms.

There are also well-known limits on RCEs. They are often prohibitively costly, and in many cases infeasible for legal or ethical reasons. For example, a real estate agency may find it very difficult to undertake an RCE that randomly varies the list prices of its customers. Even when they are feasible, there is often a steep tradeoff between the confidence in the conclusions and the range of 'treatments' that can be tested in an RCE. Finally, there are many other types of experimentation besides RCEs, and although they are considered the 'gold standard' for demonstrating efficacy in many scientific disciplines, and government agencies such as the U.S. Food and Drug Administration require them, Deaton and Cartwright (2018, p. 2) note that RCEs 'even when feasible are dominated by other methods'. For example, our discussion of Bayesian learning and control in Section 8 revealed that an optimal dynamic experimentation/learning strategy will not typically involve RCEs but rather a more sophisticated experimentation process that exploits prior knowledge of the problem and considers the tradeoff between exploitation and experimentation.

This is why we do not believe that experimentation and ML completely dominate and will replace SE studies using non-experimental data. It makes sense to use all available data whenever possible, and these days many firms have huge amounts of it. Given that the cost of modelling and simulation is relatively low and continually decreasing, it seems obvious that the starting point should be SE analyses of existing (generally non-experimental) data. Unfortunately, our review of SE approaches to identification and inference for non-experimental data in Section 8 leads us to conclude that the assumptions and methods that are appropriate for different problems are highly case-specific and depend on the nature of endogeneity and data problems that vary from problem to problem. It does not seem possible at present to prescribe a single generic approach that enables us to solve these problems. In other words, we cannot *automate* the process of inference, but need to rely on human understanding and analysis to understand the econometric problem at hand, provide the appropriate identifying assumptions, and employ the appropriate estimation method to deal with it. Overall, we believe that different ways of looking at identification and inference issues can be very productive, although there is an ever-present danger of becoming siloed owing to differences in terminology and how we look at things.

In many cases, a firm or organization will be hesitant to go to the next step and undertake additional costly data gathering or experimentation until it sees sufficiently convincing evidence of what the payoff might be. With enough initial data, and provided our models and assumptions are sufficiently good, it will generally be more cost-effective to build a computer simulation model and crash test alternative policies inside the virtual world of the computer than to undertake trial and error experimentation in reality. Experimentation and additional data gathering are called for when the counterfactual predictions from the model seem non-robust or dependent on assumptions we are not very confident about. Ultimately, fuller out-of-sample predictive tests and model validations are probably required to convince a firm to change its behaviour in response to predictions of ML and SE models. The studies by Todd and Wolpin (2006), Cho and Rust (2010) and Dubé and Misra (2019) are examples where counterfactual predictions from SE models are tested and validated via controlled experiments, randomized or otherwise. We are aware of only a few examples, for example Misra and Nair (2011), where a firm has been sufficiently convinced by the 'first stage' modelling results using non-experimental data to jump directly to the implementation of a counterfactual superior decision rule predicted by the model.[45]

[45] Credibility also depends on the level of 'transparency' of the model and the methods used to build it. Human trust of AI systems often depends on having an *intuitive understanding* of the recommendations from the systems, which is

We also believe that there are huge opportunities for ML and SE to rapidly and cost-effectively improve the experimental design of controlled experiments to evaluate whether model-generated policies can actually increase firm profitability, as well as how it affects customer welfare. Essentially, we can design various types of *virtual controlled experiments* via computer simulations to evaluate how long a field experiment must be run, how much data would need to be collected, and how much a field experiment is likely to cost in order to obtain conclusive evidence as a prelude to undertaking an actual controlled field experiment. Although trial and error experimentation may be a human analogue of real-time DP and Q-learning, it brings us back to the fundamental question we have asked in this article, namely which do we trust more: unaided human judgement or the recommendations of a machine? Ultimately, it seems that a combination of the two approaches is probably the best way to proceed, via the human/AI computerized recommendation systems and actor–critic algorithms we discussed in Section 5.

In summary, SE and ML can help to improve decision making in two key ways: (a) by helping to design and discover better policies for the firm (e.g., pricing or investment policies, etc.) and (b) by helping to design more cost-effective field experiments to evaluate whether the predicted profitability improvements from such policies can actually be realized. In the case of board games, the 'field experiments' that were required in order to demonstrate the superiority of the *DeepBlue* and *Alpha Zero* strategies were much simpler: they only required showing that they could consistently defeat the world's best human chess players. The superiority of the ML strategies was convincingly established back in 1997 when *DeepBlue* beat Garry Kasparov. Since then, the evolution of computerized strategies has continued within virtual controlled experiments known as 'computerized tournaments'.

However, considerable work remains to be done to convince firms and government policymakers that policy recommendations from computer models developed using ML and SE are credible. As Manski (2013, p. 2) notes, 'policy predictions often are fragile. Conclusions may rest on critical unsupported assumptions or on leaps of logic. Then the certitude of policy analysis is not credible.' We are optimistic that a combination of improvement in methodology for estimating more accurate SE models under weaker but more realistic assumptions combined with an increased use of controlled experiments to validate the counterfactual predictions of these models will pave the way towards their more widespread use, and will help to improve the credibility of the profession as a whole. However, we remain humble and advise starting out less ambitiously—following the lead in ML—by focusing on credible demonstrations that these tools can improve decision making 'in the small'.

11. CONCLUSION

Machine learning (ML) and structural econometrics (SE) have different goals, although both can be seen as different approaches to the holy grail of 'artificial intelligence'. Both literatures have

inconsistent with treating them as 'black boxes' without need of further explanation. Recent studies of human/AI decision teams by Bansal et al. (2019) shows that 'In settings where the human is tasked with deciding when and how to make use of the AI system's recommendation, extracting benefits from the collaboration requires the human to build insights (i.e., a mental model) about multiple aspects of the capabilities of AI systems.' Byrne (2019) stresses the importance of the use of counterfactuals to enable humans to better understand and trust AI recommendation systems, in a new field called *explainable AI* (XAI): 'Counterfactuals have become essential in many aspects of Artificial Intelligence. . . . Perhaps the most striking use of counterfactuals in AI at present is in explainable AI. . . . To increase trust in such systems by human users, and accuracy in their training by designers, there is a need to enable AI systems to provide dynamic or ad hoc explanations of their decisions in ways that are intelligible to humans.' p. 6276 provided powerful tools for learning, and particularly learning about the strengths, weaknesses and limitations of human decision making, in spite of economists' continuing fixation on the assumption of unbounded rationality. We described the angst that ML, a new, very successful and powerful form of AI, has created for many lower-skilled workers around the world, and raised the prospect that its success may even make us, structural econometricians, redundant. This may be an inevitable by-product of the scientific progress that has created the mathematical, statistical and computational tools that may ultimately take us to the doorstep of 'unbounded rationality' itself.

In the course of trying to better understand and replicate ourselves, we may be collectively paving the way to our own ultimate demise, as Benzell et al. (2015) suggested: 'Will smart machines do to humans what the internal combustion engine did to horses – make them obsolete?' This is a fear shared by the late Stephen Hawking, who was one of numerous signatories to an open letter that warns 'we could one day lose control of AI systems via the rise of superintelligences that do not act in accordance with human wishes – and that such powerful systems would threaten humanity. Are such dystopic outcomes possible? If so, how might these situations arise? . . . What kind of investments in research should be made to better understand and to address the possibility of the rise of a dangerous super-intelligence or the occurrence of an "intelligence explosion"?'

We think this is still a long way off and prefer to conclude on a less apocalyptic note. At least in the short run, we see great opportunities and no immediate threats to our existence as structural econometricians from progress on ML. Instead, we feel that McFadden (2021) provides the right perspective on the opportunities and challenges:

> 'What are the lessons here for econometricians? You should not simply dismiss learning machines and their computer operators as deplorables. You should instead think of them as your worst possible students – ignorant, arrogant, and disinterested. If you can figure out how to break through and teach them to respect and use the scientific content of economics, they may prove to be your best research associates. To keep structural econometrics vigorous and relevant, you need to continue to move aggressively to embrace the innovations in data collection and analysis created by computational advances.'

ACKNOWLEDGEMENTS

The authors are grateful to the Editors for making this special issue possible. We acknowledge Jaap Abbring, Mitsuru Igami, Tobias Klein, Simon Scheidegger, Harry Paarsch, Sanjog Misra, Günter Hitch and Frank Vella for insightful comments and feedback. We also acknowledge the feedback from Øystein Daljord, prior to his untimely death, which is a terrible loss for the profession.

REFERENCES

Abbring, J. H. and J. R. Campbell (2005). A firm's first year. Discussion Paper 2005-046/3, Tinbergen Institute, Amsterdam.

Abbring, J. H. and Ø. Daljord (2020). Identifying the discount factor in dynamic discrete choice models. *Quantitative Economics 11*, 471–501.

Aguirregabiria, V. and J. Jeon (2020). Firms beliefs and learning: models, identification, and empirical evidence. *Review of Industrial Organization 56*, 203–35.

Aguirregabiria, V. and A. Magesan (2020). Identification and estimation of dynamic games when players' beliefs are not in equilibrium. *Review of Economic Studies 87*, 582–625.

Alliot, J.-M. (2017). Who is the master? *IGGA Journal 39*, 3–43.

Anderson, A., J. Rosen, J. Rust and K. Ping Wong (2020). Disequilibrium play in tennis. Working paper, Department of Economics, Georgetown University, Washington D.C..

Arora, S. and P. Doshi (2018). A survey of inverse reinforcement learning: challenges, methods and progress.

Bajari, P., C. L. Benkard and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*, 1331–70.

Bajari, P., D. Nekipelov, S. P. Ryan and M. Yang (2015). Machine learning methods for demand estimation. *American Economic Review Papers and Proceedings 105*(5), 481–85.

Bandyopadhyay, S., F. Iskhakov, T. Johnson, S. Lee, D. McArthur, J. Rust, J. Watson and J. Watson (2013). Can the economics job market be improved?. In Z. Neeman, M. Niederle, A. Roth and N. Vulkan (Eds.), *Oxford Handbook of Market Design*. Oxford University Press: Oxford.

Bansal, G., B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld and E. Horvitz (2019). Beyond accuracy: the role of mental models in human–AI team performance. In AAAI 2019.

Barkley, A., J. R. Groeger and R. A. Miller (2020). Bidding frictions in ascending auctions. *Journal of Econometrics*. forthcoming.

Barron, A. (1994). Approximation and estimation bounds for neural networks. *Machine Learning 14*, 115–33.

Barto, A., S. Bradtke and S. Singh (1995). Learning to act using real time dynamic programming. *Artificial Intelligence 72*, 81–138.

Benzell, S. G., L. J. Kotlikoff, G. LaGarda and J. D. Sachs (2015). Robots are us: some economics of human replacement. Working paper 20941, National Bureau of Economic Research, Boston, MA.

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*, 841–90.

Bertsekas, D. and J. Tsitsiklis (1996). *Neuro-Dynamic Programming*. Athena Scientific: Belmont, MA.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag.

Blesch, M. and P. Eisenhauer (2020). Robust investments under risk and ambiguity.

Blundell, R. and R. L. Matzkin (2014). Control functions in nonseparable simultaneous equations models. *Quantitative Economics 5*, 1137–73.

Bonhomme, S., T. Lamadon and E. Manresa (2019). Discretizing unobserved heterogeneity. Working paper, University of of Chicago.

Bray, R. L. (2020). On the limits of using randomness to break a dynamic program's curse of dimensionality. Manuscript, Northwestern University.

Brumm, J. and S. Scheidegger (2017). Using adaptive sparse grids to solve high-dimensional dynamic models. *Econometrica 85*, 1575–612.

Byrne, R. M. (2019). Counterfactuals in explainable artificial intelligence (XAI): evidence from human reasoning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*.6276–82.

Camerer, C. (2003). *Behavioral Game Theory*. Princeton University Press: Princeton, NJ.

Campbell, J. Y. and R. J. Shiller (1988). Stock prices, earnings, and expected dividends. *Journal of Finance 43*, 661–76.

CEA (2020). *The Economic Report of the President*. The White House and Executive Office of the President.

Chan, M. K. (2017). Welfare dependence and self-control: an empirical analysis. *Review of Economic Studies 84*, 1379–423.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey and J. M. Robins (2016). Locally robust semiparametric estimation.

Cho, S., G. Lee, J. Rust and M. Yu (2020). Semi-parametric instrument-free demand estimation: relaxing optimality and equilibrium assumptions. Working paper, Department of Economics, Georgetown University, Washington D.C..

Cho, S. and J. Rust (2010). The flat rental puzzle. *Review of Economic Studies 77*, 560–94.

Cho, S.-J., H. J. Paarsch and J. Rust (2019). An empirical analysis of informationally-restricted dynamic auctions of used cars. Department of Economics, University of Central Florida.

Chow, C. Tsitsiklis (1989). The complexity of dynamic programming. *Journal of Complexity 5*, 466–88.

Cobb, C. W. and P. H. Douglas (1928). A theory of production. *American Economic Review 18*, 139–65.

Crawford, G. S. and M. Shum (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica 73*, 1137–73.

Dantzig, G. and P. Wolfe (1961). The decomposition algorithm for linear programs. *Econometrica 29*, 767–78.

Deaton, A. and N. Cartwright (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine 210*, 2–21.

DellaVigna, S. (2018). Structural behavioral economics. In B.D. Bernheim, S. DellaVigna and D. Laibson (Eds.), *Handbook of Behavioral Economics – Foundations and Applications 1*. 613–723. Amsterdam: Elsevier.

Diaconis, P. and D. Freedman (1986). On the consistency of bayes estimates. *Annals of Statistics 14*, 1–26.

Doraszelski, U., G. Lewis and A. Pakes (2018). Just starting out: learning and equilibrium in a new market. *American Economic Review 108*(3), 565–615.

Druedahl, J. and A. Munk-Nielsen (2020). Higher-order income dynamics with linked regression trees. *Econometrics Journal 23*.

Dubé, J.-P. and S. Misra (2019). Personalized pricing and customer welfare. Working paper, University of Chicago Booth School of Business, Chicago. https://ssrn.com/abstract=2992257.

Dubé, J.-P. H., G. J. Hitsch and P. K. Chintagupta (2010). Tipping and concentration in markets with indirect network effects. *Marketing Science 29*, 216–49.

Dyson, G. (2012). *Turing's Cathedral*. Pantheon Books: New York.

Eichenbaum, M., S. Rebelo and M. Trabandt (2020). The macroeconomics of epidemics. Working paper, UCLA.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica 57*, 937–69.

Ericson, R. and A. Pakes (1995). Markov-perfect industry dynamics: a framework for empirical work. *Review of Economic Studies 62*, 53–82.

Freedman, D. (1987). As others see us: a case study in path analysis. *Journal of Educational Statistics 12*, 101–223.

Freedman, D. and P. Diaconis (1983). On inconsistent bayes estimates in the discrete case. *Annals of Statistics 11*, 1109–18.

Frisch, R. (1926). Sur un problème d'économie pure. *Norsk Matematisk Forenings Skrifter 16*, 1–40.

Goodfellow, I., Y. Bengio and A. Courville (2016). *Deep Learning*. MIT Press: Cambridge, MA.

Granger, C. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica 37*, 424–38.

Griffiths, T. and J. Tenenbaum (2009). Theory based causal induction. *Psychological Review 116*, 661–716.

Gupta, S. and J. Rust (2018). A simple theory of why and when firms go public. Working paper, Georgetown University Department of Economics, Washington D.C.

Haavelmo, T. (1943). The statistical implication of a system of simultaneous equations. *Econometrica 11*, 1–12.

Hall, G. J. and J. Rust (2020). Estimation of endogenously sampled time series: the case of commodity price speculation in the steel market. *Journal of Econometrics*. *(forthcoming)*.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*, 153–61.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions. In J. J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data: Econometric Society Monograph, Volume 10*. Cambridge: Cambridge University Press.

Hitsch, G. (2006). An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science 25*, 25–50.

Hitsch, G. J. and S. Misra (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. Working paper, University of Chicago Booth School of Business, Chicago, IL.

Holland, J. H. (1986). Escaping brittleness: the possibilities of general-purpose learning algorithms applied to parallel rule-based systems. In R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), *Machine Learning, an Artificial Intelligence Approach*. 593–623.Morgan Kauffman: Burlington, MA.

Hong, H., H. J. Paarsch and R. Haley (2006). *An Introduction to the Structural Econometrics of Auction Data*. MIT Press: Cambridge, MA.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*, 497–529.

Howard, R. (1960). *Dynamic Programming and Markov Processes*. MIT Press: Cambridge, MA.

Hubbard, T. P., S. Klonner and H. J. Paarsch (2020a). The value of financial intermediation: evidence from chit-fund auctions in Tamil Nadu. Department of Economics, University of Central Florida.

Hubbard, T. P., H. J. Paarsch and W. M. Wright (2020b). Hụi: a case study of a sequential double auction of capital. Department of Economics, University of Melbourne.

Hume, D. (1739). *A Treatise on Human Nature*. Oxford: Clarendon Press.

Igami, M. (2020). Artificial intelligence as structural estimation: economic interpretations of Deep Blue, Bonanza, and AlphaGo. *Econometrics Journal 23*.

Igami, M. and K. Uetake (2020). Mergers, innovation, and entry-exit dynamics: consolidation of the hard disk drive industry, 1996–2016. *Review of Economic Studies*

Imai, K. and I. S. Kim (2019). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science 63*, 467–90.

Imai, S., N. Jain and A. Ching (2009). Bayesian estimation of dynamic discrete choice models. *Econometrica 77*, 1865–99.

Iskhakov, F., T. Jorgensen, J. Rust and B. Schjerning (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitive Economics 8*, 317–65.

Iskhakov, F., J. Lee, J. Rust, B. Schjerning and K. Seo (2016). Constrained optimization approaches to the estimation of structural models: comment. *Econometrica 84*, 365–70.

Iskhakov, F., J. Rust and B. Schjerning (2016). Recursive lexicographical search: finding all Markov perfect equilibria of finite state directional dynamic games. *Review of Economic Studies 83*, 658–703.

Iskhakov, F., J. Rust and B. Schjerning (2018). The dynamics of Bertrand price competition with cost-reducing investments. *International Economic Review 59*, 1681–731.

Judd, K. L., L. Maliar, S. Maliar and R. Valero (2014). Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control 44*, 92–123.

Keane, M. and T. Neal (2020). Comparing deep neural network and econometric approaches to predicting the impact of climate change on agricultural yield. *Econometrics Jounal 23*.

Kemptner, D. and S. Tolan (2018). The role of time preferences in educational decision making. *Economics of Education Review 67*, 25–39.

Kiefer, N. M. and Y. Nyarko (1989). Optimal control of an unknown linear process with learning. *International Economic Review 30*, 571–86.

Korolev, I. (2020). Estimating the parameters of the SEIRD model for COVID-19 using the deaths data. Working paper, Binghampton University, Binghmapton NY.

Kristensen, D., P. K. Mogensen, J. M. Moon and B. Schjerning (2020). Solving dynamic discrete choice models using smoothing and sieve methods. *Journal of Econometrics, forthcoming*.

Kubler, F. and S. Scheidegger (2019). Self-justified equilibria: existence and computation. SSRN Working paper, University of Zurich, Zurich, Switzerland.

Kucharski, A., T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk and R. M. Eggo (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet, Infectious Diseases*, *20*, 553–8.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*, 443–77.

Lindley, D. (2017). *Statistical Inference*. 1–7, London: Palgrave Macmillan.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review 46*, 97–116.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Eonometrica 70*, 801–16.

Maliar, L., S. Maliar and P. Winant (2019). Will artificial intelligence replace computational economists any time soon? Technical report, CEPR Discussion paper 14024, Centre for Economic Policy Research, Lonodon UK.

Manski, C. and D. McFadden (1981). *Structural Analysis of Discrete Data*. MIT Press: Cambridge, MA.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer Verlag: Berlin, Germany.

Manski, C. F. (2013). *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge University Press: Cambridge.

Mao, J. and Z. Zheng (2020). Structural regularization. Manuscript, Xiamen University.

Markets, Markets (2020). Revenue management market by solutions (risk management, pricing and revenue forecast management, revenue analytics, revenue leakage detection, channel revenue management) by services (professional, managed) by deployment mode — global forecast to 2024. Technical report TC4241, Markets and Markets, New York.

McClelland, J. and J. Rust (2018). Strategic timing of investment over the business cycle: machine replacement in the US rental industry. *Jahrbücher für Nationalökonomie und Statistik 238*, 313–69.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of Econometrics*, 105–42. New York, NY: Academic.

McFadden, D. (1975). On independence, structure, and simultaneity in transportation demand analysis. Technical report, Urban Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data*, 198–272. Cambridge, MA: MIT Press.

McFadden, D. (1989). A method of simulated moments for estimation of discrete models without numerical integration. *Econometrica 57*, 995–1026.

McFadden, D. (2021). Epilogue. *Journal of Econometrics*. *forthcoming*.

Merlo, A., F. Ortalo-Magne and J. Rust (2015). The home selling problem: theory and evidence. *International Economic Review 56*, 457–84.

Miller, N. and A. MacKay (2019). Estimating models of supply and demand: instruments and covariance restrictions. Working paper, Georgetown McDonough School of Business , Washington, D.C.

Misra, S. and H. Nair (2011). A structural model of sales-force compensation dynamics: estimation and field implementation. *Quantitative Marketing and Economics 9*, 211–57.

Montanelli, H. and Q. Du (2019). New error bounds for deep ReLU networks using sparse grids. *SIAM Journal on Mathematics of Data Science 1*, 78–92.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

Nemirovsky, A. and D. Yudin (1983). *Problem Complexity and Method Efficiency in Optimization*. New York, NY: Wiley.

Newey, W., J. Powell and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*, 565–603.

Ng, A. Y. and S. Russell (2000). Algorithms for inverse reinforcement learning. In *Proc. 17th International Conf. on Machine Learning*. 663–70. Morgan Kaufmann, Burlington, MA.

Norets, A. (2009). Inference in dynamic discrete choice models with serially correlated unobserved state variables. *Econometrica 77*, 1665–82.

Norets, A. (2012). Estimation of dynamic discrete choice models using artificial neural network approximations. *Econometric Reviews 31*, 84–106.

Nyarko, Y. (1991). On the convergence of Bayesian posterior processes in linear economic models: counting equations and unknowns. *Journal of Economic Dynamics and Control 15*, 687–713.

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *American Economic Review 89*(1), 103–24.

Pakes, A. and P. McGuire (2001). Stochastic algorithms, symmetric Markov perfect equilibria, and the 'curse' of dimensionality. *Econometrica 69*, 1261–81.

Pearl, J. (2009). *Causality Models, Reasoning and Inference, 2nd edition*. Cambridge University Press: Cambridge.

Pearl, J. (2018). Theoretical impediments to machine learning with seven sparks from the causal revolution.

Pearl, J. (2019). The seven tools of causal inference with reflections on machine learning. *Communications of the ACM 62*, 54–60.

Powell, W. (2010). *Approximate Dynamic Programming Solving the Curses of Dimensionality*. New York, NY: Wiley.

Powell, W., B. Bouzaiene-Ayari, C. Lawrence, C. Cheng, S. Das and R. Fiorillo (2014). Locomotive planning at Norfolk Southern: an optimizing simulator using approximate dynamic programming. *Interfaces 44*, 567–78.

Puterman, M. and S. Brumelle (1979). On the convergence of policy iteration in stationary dynamic programming. *Mathematics of Operations Research 4*, 60–69.

Reich, G. (2018). Divide and conquer: recursive likelihood function integration for hidden Markov models with continuous latent variables. *Operations Research 66*, 1457–70.

Renner, P. and S. Scheidegger (2018). Machine learning for dynamic incentive problems. SSRN Working paper, University of Zurich, Zurich, Switzerland.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

Rubin, D. (1990). Formal modes of statistical inference for causal effects. *Journal of Statistical Planning and Inference 25*, 279–92.

Russell, S. (1998). Learning agents for uncertain environments (extended abstract). In *Proceedings of the 11th Annual Conference on Computational Learning Theory*. 101–03. ACM Press.

Rust, J. (1987). Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica 55*, 999–1033.

Rust, J. (1988). Maximum likelihood estimation of discrete control processes. *SIAM Journal of Control and Optimization 26*, 1006–24.

Rust, J. (1994). Structural estimation of Markov decision processes. In *Handbook of Econometrics*, Volume *4*, 3081–143. Elsevier: Amsterdam, the Netherlands.

Rust, J. (1997). Using randomization to break the curse of dimensionality. *Econometrica 65*, 487–516.

Rust, J. (2000). Nested fixed point algorithm documentation manual, version 6.

Rust, J. (2014). The limits of inference with theory: a review of Wolpin (2013). *Journal of Economic Literature 52*, 820–50.

Rust, J. (2019). Has dynamic programming improved decision making? *Annual Review of Economics 11*, 833–58.

Rust, J., J. Traub and H. Wozniakowski (2002). Is there a curse of dimensionality for contraction fixed points in the worst case? *Econometrica 70*, 285–329.

Rutman, J. (2018). God's chess rating. https://unclejerry9466728.wordpress.com/2018/12/20/172, accessed March 30, 2020.

Schaeffer, J., N. Burch, Y. Björnsson, A. Kishimoto, M. Müller, R. Lake, P. Lu and S. Sutphen (2007). Checkers is solved. *Science 317*, 1518–22.

Scheidegger, S. and I. Bilionis (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science 33*, 68–82.

Schneider, U. (2020). Identifying and estimating beliefs from choice data – an application to female labor supply. Working paper, University of Groningen Netherlands.

Silver, D. et al, (2017). Mastering the game of Go without human knowledge. *Nature 550*, 354–58.

Simon, H. (1992). Rational decision-making in business organizations. In A. Lindbeck (Ed.), *Nobel Lectures, Economics 1969–1980*, 343–71. World Scientific Publishing, Company: Singapore.

Simon, H. A. (1953). Causal ordering and identifiability. In W. C. Hood and T. Koopmans (Eds.), *Studies in Econometric Method*. Wiley and Sons.

Smith, C. S. (2020). Computers that teach themselves: scientists are exploring ways to help machines develop 'common sense'. *New York Times*. *April 9, 2020*, F6.

Splawa-Neyman, J., D. M. Dabrowska and T. P. Speed (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science 5*, 465–72.

Strotz, R. H. and H. O. A. Wold (1960). Recursive vs. nonrecursive systems: an attempt at synthesis (part I of a triptych on causal chain systems). *Econometrica 28*, 417–27.

Su, C. and K. L. Judd (2012). Constrained optimization approaches to the estimation of structural models. *Econometrica 80*, 2213–30.

Todd, P. and K. I. Wolpin (2006). Assessing the impact of a school subsidy program in Mexico: using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review 96*, 1384–417.

Tsitsiklis, J. (1995). Asynchronous stochastic approximation and Q-learning. *Machine Learning 16*, 185–202.

von Neumann, J. and O. Morgenstern (1953). *Theory of Games and Economic Behavior*. Princeton University Press: Princeton, NJ.

Walker, M. and J. Wooders (2001). Minimiax play at Wimbledon. *American Economic Review 91*(5), 1521–38.

Watkins, C. (1989). *Learning from delayed rewards*. Ph.D. thesis, Cambridge, UK: Cambridge University.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy 92*, 852–74.

Wolpin, K. I. (2013). *The Limits of Inference Without Theory*. MIT Press: Boston, MA.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 557–85.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website.

ONLINE SUPPLEMENT

Co-editor Jaap Abbring handled this manuscript.