PHD OF PHILOSOPHY RESEARCH PROPOSAL

Casual Inference Models and its Empirical Application

Xinglin Lai

12/3/2022

Contents

1 Abstract

There are a large number of causal inference models such as instrumental variables, policy experiments, PSM-DID, etc., but these models are not sufficient to verify micro mechanisms and specific causal relationships. I review the now popular econometric and machine learning models for causal inference and briefly outline the academic progress in analyzing causality with Agent-Based Computational Economics Model (ACE) and Dynamic Stochastic General Equilibrium (DSGE) models. My research claim is that the causal mechanism is essentially a topological structure that can be represented as structural causal models (SCMs) under linear constraints [e.g., Hoover, 2008[29], Pearl, 2009[45], Pearl et al., 2016[47]]. A simple approach is to use instrumental variables of great properties for model identification and inference. I propose the Quantile Instrumental Variables (QIV) method constructed from quantile points and the unbiased and asymptotic proporties of its two-stage least squares estimator for large samples. I will continue to investigate the asymptotic properties of QIV-3SLS estimators for structural causal models, and numerical experiments focusing on complex noise structures and non-linear models will be conducted (expected to be completed this summer). On the other hand, since instrumental variables and systematic estimation cannot answer the causality and causal effects under linearity, I propose to include a decision layer composed of multiple micro agent individuals (ACE) in neural network, giving them the ability to make rational or game decisions, use the input layer as information, and then output the mixed structure of decisions or as the components of the next layer. The combination of the ACE and the layers structure in the network should approximate the real causal structure graph if the distribution of the generated data is nearly identical to that of the actual data and if a particular causal mechanism constructed by the ACE is significantly better than the others. Then We can use an analysis tool of DSGE models (Section 4.3) for comparing static equilibrium analysis and dynamic simulation analysis. Finally, I present my own research experience. I have assisted my mentor, Prof. Shuguang Xiao, in a general project of the National Social Science Foundation and a project of the Humanities and Social Science Foundation of the Ministry of Education. My two working papers use DSGE model and econometric methods of causal inference [Gabaix et al.,2020[18];Peters et al.,2016[47] to study the micro mechanisms of monetary over-issuance driven by local governments in China through fiscal expansion, respectively.

2 Aims and Scopes

1. **Models for Casual Inference** This section focuses on constructing causal empirical methods represented by instrumental variables (IV) as well as systematic equations estimation and causal deductive methods combining generative adversarial networks and Agent-Based computational Economic models (ACE-GAN). I also introduce network interference, a causal empirical approach similar to (ACE-GAN). In fact, the estimating framework of network interference can also be embedded in the model structure of a neural network, representing the causal mechanism with a layer structure. Finally I also discuss some technical details, such as the asymptotic nature of the statistics and numerical optimization.

2. **Generate Unobservable Data** Using the framework of ACE-GAN we can generate instrumental or proxy variables that satisfy the given moment conditions from exogenous variables or co-factors. This facilitates solving empirical challenges such as missing data and unobservable but significant variables.

3. **DSGE models with Micro structures(ACE-DSGE)**To make causal inference more theoretical and to connect micro-mechanisms and macroeconomic analysis, individual irrational decision making, games, information asymmetry, incentive contract design, and disinformation are added to the agent-based model, while incorporating them into general equilibrium analysis.

4. **Empirical Applications of Casual Inference Models** In this section, taking one of my working

papers as an example, I will introduce how causal inference models can be used to study the micro-mechanisms of macroeconomic problems. Of course, in fact causal inference methods will be more effective when used for micro empirical application.

3 Background and Literature Review

The main casual inference econometric models such as DID[62], regression discontinuity design[59][5][14], matched estimation and synthetic control methods(SYC)[1][4] attempt to address estimation bias due to endogeneity, confounding factors, etc. as much as possible based on the available sample, so that the in-sample estimates satisfy unbiasedness and consistency as much as possible. However, the causal inference models that are widely adopted today still suffer from estimation bias that can be easily ignored.

On the one hand, causal assessment econometric models are often based on the assumption of homogeneous treatment effects, i.e., there is almost no intergroup heterogeneity, which in statistics is attributed to the sample variance of the estimated quantity (the sample variance is infinitely close to the overall variance as the sample size tends to infinity). This homogeneity assumption often leads to estimation bias due to the fact that intergroup heterogeneity is usually subject to a self-selection effect, where people with certain traits actively choose to participate/avoid participating in a policy or enter a policy-relevant treatment/control group. Self-selection effects may lead to biased estimates of policy effects, e.g., a policy that applies tax incentives or relaxes financing constraints to a particular industry may lead to the entry of many operationally efficient firms into that industry, thus weakening the treatment effect of the policy or even the researcher may conclude that the policy has a negative effect. Another problem caused by heterogeneity is that the treatment effect itself is significantly group heterogeneous, and group heterogeneity causes the traditional average treatment effect (ATE) to become insignificant or, more seriously, because we estimate the local average treatment effect (LATE), the within-sample estimates do not reflect the sample aggregate.

On the other hand, the Rubin causal framework[51] cannot tell us a specific casual mechanism. Most of the literatures are based on a theoretical analysis or hypothesis followed by estimation of casual effect, but it is precisely the transmission mechanism that lies at the main problem for many important economic issues. The core variables for general casual problems are dummy variables consisting of zeros and ones, i.e. the implementation or enforcement of a particular policy. In fact, however, the core explanatory variables may be multivariate discrete variables or continuous variables. Therefore estimation methods based on Rubin(1974)[51]'s counterfactual framework are not applicable. Wan(2021)[61] analyses the entrepreneurship-driving effect of urban community infrastructure through micro household survey data and found that community infrastructure has a large positive externality, with a significantly higher proportion of residents starting their own businesses. Wan(2021)[61] is a causal inference study of the average casual effect of community infrastructure on the probability of residents choosing to start a business, but the study ignores the role of transmission mechanisms. When a city has a higher level of economic development, residents have higher levels of income and social benefits, so they have enough capital to support their ambitious goal of entrepreneurship, while higher income and social benefits provide good risk protection for entrepreneurship. Affluent cities also have more revenue to invest in community infrastructure, and the level of infrastructure development in central cities is much higher than in suburban areas. Therefore, it is likely that community infrastructure is not the real reason for the relatively high level of entrepreneurial enthusiasm among this segment of the population, but rather the higher level of economic development of the city in which they live. The higher level of economic development may be related to more factors such as the level of openness to the outside world, policy support, and the level of government management. Therefore, the observation perspective of causal inference from policy evaluation econometric models alone has obvious limitations because they only examine the causal effect between the two and ignore the transmission mechanism, or more generally, the causal mechanism. On the other ha

The problem of heterogeneous treatment effects can be solved by changing the econometric model, and similarly, estimation bias due to potential causality can be re-estimated by adding relevant variables and using an appropriate econometric model, giving rise to two novel empirical directions in economics: non-traditional statistical models and unstructured data.

Non-traditional statistical models in this paper generally refer to statistical learning methods. Machine learning methods have increasingly been used by more economists for empirical research in recent years. For example, estimating casual effects requires the counterfactual control group, but only the treated group can be observed. To solve this problem, the synthetic control method (Abadie et al., 2010)[1] constructs counterfactual individuals as control groups by weighting the average values of different similar individuals. The weights of the SCM are set by a solution that minimises the sum of squared weighted residuals, but control groups with better fitting performance can also be constructed by means of random forest trees (Cicala, 2017)[13]. In addition, statistical learning models such as support vector machines and neural networks have stronger predictive power than traditional parametric and semi-parametric models to construct control groups that are closer to the counterfactual. Varian(2016)[60] constructs future wages for students who were able to score 60 but did not attend university through various predictive model. Specifically, for each individual, Varian(2016)[60] simply modifies their grade to 60 using the trained model to obtain a predicted value of individual's wage, obtaining an unbiased estimate of the treatment effect while preserving individual characteristics.

Probabilistic graphical models are a kind of probabilistic models that use graphs to express the correlation of variables, which use a node to represent a random variable or a set of random variables, and the edges between the nodes to represent the probabilistic correlation between the variables, i.e. the graph of variable relationships. Depending on the nature of the edges, probability diagrams can be broadly classified into two categories: the first is Bayesian networks (Pearl, 1988)[44] such as Hidden Markov Models(HMM), while the second is Markov networks such as Markov Random Fields (Gemen and Geman, 1984)[22]. Markov random fields are not directed graph models and therefore cannot represent the causal relationships of variables. HMM is a well-known directed graph model that can represent dependencies between variables, but this method is not yet able to represent true causality, so in recent years scholars have proposed many new methods, mainly divided into extracting causal relationships [Rohekar et al.(2021)[49], Addanki et al.(2021)[2]], causal structure learning [Zečeviøc et al.(2021)[63],Kaddour(2021)[31],Akbari(2021)[3]] and causal inference [Lee(2021)[37],Sun(2021)[57]] in three subfields.

In addition to causal structure models based on statistical learning, a number of scholars have developed many causal inference models based on traditional statistical methods in recent years.Peters et al.(2016)[47],[48] proposed the Invariant Prediction Model, which is a causality estimation tool based on the assumption of a linear system of equations and Chow's Test.The Invariant Prediction Model is based on the Additive Noise Model[Peters,2014[46] ; Bühlmann ,2014[11]; Mooij,2014[41]; Hoyer et al.(2009),[30]]. Heinze-Deml et al.,2018[28] proposed an Invariant Prediction Model for non-linear models.

The literature referred to above is concerned with the role of non-traditional statistical models in casual inference. Furthermore, non-traditional statistical models, based on statistical learning methods, are of greater use in data generation. The availability of data for core variables in policy evaluation often troubles scholars, due to the fact that the data may be private or restricted by laws and regulations. Unstructured data such as text and images or even video offer the possibility to address these issues. Hansen(2018)[25] wanted to investigate how transparency policies affect the decision-making process within government, but it was clear that both the two variables were abstract, so the study used text data from the Federal Open Market Committee (FOMC) of a total of 5 million words and 50,000 speeches via Latent Dirichlet Allocation (LDA) method (an unsupervised machine learning method) to extract 40 different topics. By labelling the text data of each cell by different topics, they obtained metrics such as the percentage of different topics and the similarity of speeches between members, expressing the unstructured data as structured data. In addition to textual data, statistics learning models can also predicted and classify images. For example, Engstrom et al.(2017)[16] made use of convolutional neural networks to identify fixed assets such as buildings and roads in satellite images, by counting the number of similar objects in the images as a metric for the level of welfare in a given area.

Although statistics models provide a novel and powerful toolbox for empirical economics. However, data-based empirical approaches still fail to fully answer many questions in policy evaluation. Lucas critique(1976) make economists focus on the impact of self-rational decisions of micro-individuals on the macroeconomic system. Thus relied solely on empirical evidence to describe the dynamic operating mechanisms of the economic system After Lucas critique, economists began to incorporate the optimal choices of different agents into the same model for simultaneous solution. Kydlang & Prescott (1982)[35] proposed the real economic cycle (RBC) model. The RBC model argued that the main source of economic fluctuations was productivity-generating shocks, attributing economic cycles to random shocks, which inspired the dynamic general equilibrium (DSGE) model that was later developed. The New Keynesian model and the DSGE model are based on the RBC but include a variety of economic sectors (manufacturers, financial intermediaries and government), markets and various frictions and distortions to improve data-fitting and forecasting capabilities. DSGE models have a strong micro-foundation and multi-equation linkage estimation, thus they avoid the problems of the Lucas critique. However, the DSGE model still has inevitable flaws.

Specifically, deficiencies in DSGE models are divided into two main parts: identification problems and modelling error. Identification problem refers to the difference between the number of parameters to be estimated and the number of observable variables in the DSGE model, making some of the parameters unidentifiable [Romer (2016)[50]]. The modelling error consists of two aspects: First, the stochastic shock assumption is too arbitrary. Romer(2016)[50] argues that stochastic shocks in DSGE models cannot explain negative TFP shocks because technological progress is usually a non-reducible process and economic facts show that macro policies are indeed effective, so the assumption has been questioned by some scholars.Second, Stiligiz(2018)[56] argues that the micro-foundations of DSGE based on classical economic theory including the representative rational household and the assumption of complete information have been doubted and even replaced by new developments in branches of economics such as behavioural economics, game theory and information economics. DSGE models may also ignore the economic effects and heterogeneity from micro individuals, which may result from the interaction of a large number of individual behaviours. As a consequence, the predictive and explanatory power of DSGE models is not yet able to address many important economic issues such as financial crises. In addition to the above points, scholars such as [Korinek (2017)[34],Galí (2017)[19], Blanchard (2017)[9]] have also criticised and suggested improvements to the DSGE model in terms of different aspects of modelling errors.

The development of mathematical tools such as evolutionary theory, game theory and optimisation theory since the mid-20th century has led to great challenges for models that consider perfectly rational and homogeneous subjects, as represented by the DSGE model. Some scholars found that economic systems are essentially complex adaptive systems composed of a large number of subjects. An effective way to study these systems is to use computer simulations rather than traditional mathematical analysis and econometric tests. For this reason, the formation and development of Agent-based Computational Economics (ACE) started. Marimon (1989)[39] was the first scholar to introduce the adaptive capacity of agents into his research, and he investigated the adaptive capacity of agents in the context of the rational-expectation monetary model developed by Kiyotaki et al(1997)[32]. In the context of the expected money model established by Kiyotaki et al(1997)[32], he investigated how agents learn and coordinate their behaviour with each other. In addition, the ACE model's researches on the learning ability of subjects have begun to show strengths in economic game theory. Since the publication of a special issue of the economics journal "Games and Economic Behavior on artificial intelligence" in 2001, more than 40 articles in the journal have used features such as subject learning ability and heterogeneity to study game theory problems in economic behaviour in a decade, which has become a new micro perspective and popular research topics in economic game research. In addition to game competition, early agent models such as Schelling (1971)[53] and Epstein & Axtell (1996)[17] considered individual learning, adaptation and innovation as interactive behaviours and used them to study trade transactions and market mechanisms. However, the early studies were based on conventional network structures such as uniform lattices, which ignored the complex characteristics of network structures and the behaviours arising from subject interactions. More recently, the ACE approach relies on network typologies to investigate the complexity of interactions, linkages and game behaviours between heterogeneous agents and their emergent features, such as the emergence of currencies and the diffusion of new technologies. For example, Mauro&Richiardi(2011)[20]

and Gabriele et al. (2012)[58] develop an order-driven market model to analyse the role of Herding Effects in the formation of market network structures and wealth distribution. By simulations they find that when imitation is abundant, the wealth distribution of individuals induced by transactions is heavy-tail, which is consistent with the actual performance of the market. The ACE model is therefore an effective way of modelling and analysing the interactions between heterogeneous agents and the resulting macroeconomic dynamics, providing a robust micro-foundation for data-based causal inference.

4 Program and Design of the Research Investigation

4.1 Models for Causal Inference

The causal inference is divided into two parts: causal mechanism analysis and causal effect estimation. The focus of my study will be on causal mechanism analysis. A causal mechanism is essentially a topological structure that can be represented as structural causal models (SCMs) under linear model constraints [e.g., Hoover, 2008[29], Pearl, 2009[45], Pearl et al., 2016[47]]. Thus, if unbiased consistent estimates can be obtained for structural causal models, the underlying causal mechanisms can be estimated according to the positive and negative direction of the coefficients as well as the degree of significance. Therefore, the simplest approach is to identify and infer the model using instrumental variables of good nature.

In recent years many scholars have started to investigate methods for constructing instrumental variables. [Coppola et al. ,2019[15]; Gabaix et al. (2020)[18] proposed Granular Instrumental Variables(GIV) to extract heterogeneity shocks from panel data and use them as instrumental variables to estimate causal effects. Jason et al.(2017)[27] proposed using multiple (weak) instrumental variables and using neural networks to obtain optimal estimates of treatment variables on endogenous variables, and then using a two-stage neural network approach to reduce the total prediction error. Bennett et al. (2019)[8] proposed a neural network-based GMM approach to solve the problem of instrumental variable estimation in the case of high dimensionality and complex functions.

The instrumental variable construction methods mentioned above all have corresponding drawbacks. Models such as instrumental variable estimation based on neural network models (DeepIV [Hartford et al.[27]],KernelIV[Singh et al.(2019)[55]] and DeepGMM[Bennett et al.(2019),[8]]) require large sample data sets and high dimensional variable data, which is not in line with the common situation of small samples and lack of data faced in general research. Granular Instrumental Variables (GIV) also require corresponding panel data and the assumption that GIV requires individual heterogeneity shocks to be independent of other variables is more stringent. This also limits the scope of GIV.

I present the unbiased and asymptotic properties of the Quantile Instrumental Variables (Quantile IV) method constructed from quantile points and its two-stage least squares estimators for large samples. Preliminary Monte Carlo simulation experiments show that Quanttile IV-GMM estimation for linear, nonlinear models is more accurate and robust than the results of OLS. Next I will continue to study the asymptotic nature of Quantile for structured causal models, and numerical experiments that focus more on complex noise structures and nonlinear models will also be conducted. This working paper is expected to be uploaded to the Arxiv website in the summer of 2022.

The combination of instrumental variables and systematic estimation is a causal inference in the form of data empirical evidence, but does not answer causality and causal effects of micro mechanisms. agent-based computational economics (ACE) can solve the following problem: when we determine that X is the causal variable of Y, but it is not clear the mechanism of action from X to Y or we need to determine the most important of several known action paths $Z_1, Z_2, ..., Z_M$ in the most important conduction channel. ACE can combine the potential paths of action $Z_1, Z_2, ..., Z_M$ by simulating and evolving them with analytical solutions given by game basis or microeconomic models or given behavioral models, and then comparing the estimation performance of different transmission mechanisms. The optimal estimation performance is of course almost identical to the real data, and this is exactly what generative adversarial networks (GAN) [Goodfellow et al.,2014[23],Arjovsky et al.,2017[6],Salimans[52],Gulrajani et al.,2017[24]] the basic idea. Therefore, I suggest a causal estimation framework combining ACE and GAN combined (ACE-GAN).

An important function of the ACE-GAN framework is to construct microscopically grounded counterfactual objects for different scenarios. Using a (Bayesian) moment matching approach it can be considered to construct data generating mechanisms for counterfactual objects with given parameters and make statistical inferences. Real data are used as real individuals. We additionally construct a discriminator network to train the generator network to be able to deceive the discriminator network, so that it can gradually generate increasingly realistic data as it is trained. the generator network of the GAN model is divided into two parts. The first part is the initial connection using a neural network model such as a multi-layer perceptron or a random forest tree, and the second part takes the predictions from this part as input and outputs the final data through a convolutional layer. The neural network composition of the generator is often an unknown black box. Typically researchers find the optimal hyperparameters and optimal model structure through methods such as grid search or Bayesian optimization, which do not have a solid statistical basis or interpretability. My suggestion is to have a decision layer of multiple micro-agents in the neural model network, giving them the ability to make rational or game-based decisions, using the input layer as information, and then outputting the sum of the decisions or the components of the variables in the next layer. If the distribution of the generated data is nearly identical to the actual data, and a particular causal mechanism constructed by ACE is significantly better than others, then it can be argued that the mechanism implicit in ACE-GAN can represent the core part of the fact, and some tricky economic mechanism problems can be solved by an analytical tool similar to the DSGE model (Section 4.3).

An approach similar to this is a causal model based on network interference [Bandiera et al.,2009[7],Bond et al.,2012[10],Bursztyn et al, 2014[12],Leung et al.,2022[38]]. individuals in network interference are real N subjects, and by studying their network structure, e.g. e.g. social connections , to study the causal effects of a policy or treatment measure. This approach is suitable for estimating accurate assessments of causal effects, but not for testing causal mechanisms, but this approach provides unbiased and consistent estimates of the causal effect of $D \in \{0, 1\}$ on $Y \in \mathcal{R}$. D can be output as a randomly assigned binary variable via a common connectivity function such as Softmax, so network interference can be present in the predictor as part of the GAN. Next if we consider D as a binary variable subject to a logistic distribution with conditional expectation $\mathbb{E}(D|Z)$ influenced by $Z \in \mathcal{R}^K$, then we have constructed a causal mechanism from Z to Y. Similarly, by repeating the above process and replacing variables on both the cause and effect sides, we construct a causal mechanism from Y to Z. Comparing the fit of the two causal mechanisms to the actual data (e.g. by discriminant correctness) or selecting the causal structure of the minimum objective function by OLS and (Bayesian Posterior) Max Likelihood Estimation(MLE), while maintaining unbiased consistent estimates of network interference, constitutes a causal prediction method similar to Score-Based casual inference methods.Without doubt, in combination with the Quantile IV method I proposed above, GIV for panel data or DeepIV for high-frequency, high-dimensional, large-sample data, it is possible to get closer to unbiased consistent effective property estimates than OLS or MLE. For the objective function of the combination of predictors and discriminators, I suggest combining generalized method of moments(GMM) estimation. For example, let the correctness of the discriminator's judgment error be $\widehat{DIS} \in (0, 1)$ and the conditional expectation (or median) of the predictor's predicted target variable be $\widehat{\mathbb{E}(Y)}$. The moment conditions corresponding to the two estimators are $\mathbb{E}[\widehat{DIS}] = 1, \mathbb{E}[\widehat{(Y)}] = \mathbb{Y}$, respectively. To further approximate the distribution of the real data by moment matching, second-order central moments, i.e., variance, can be added. Another more direct way is to use Kullback-Leibler (KL) divergence, Jensen-Shannon divergence or Wasserstein metric to measure the distance between two data distributions, and these same statistics can be used to derive the moment conditions.

The estimation problem in the ACE-GAN framework still has many difficulties: for example, the estimation of the inverse variance matrix and the selection of the optimizer. The variance matrix can be considered to estimate the variance by computer resampling, i.e., Bootstrap, as predetermined coefficients to construct the objective GMM function, but the first-order derivatives of the objective function for many parameters may not exist, which may lead to inconsistent estimates, and the specific formulation of the convergence rate may be very complicated in the presence of consistent estimates. For the optimization solution, the traditional (Quasi) Newton-Raphson method and the commonly used AdaGrad and Adam methods may need to be redesigned due to the problems that the objective function of the GMM may not be convex and the gradient vanishing that may result from the optimization in the high-dimensional parameter case.

4.2 ACE-GAN for Variables Generation

Another benefit of using ACE-GAN models is that by a generative neural network with a micro-foundation, researchers are able to generate instrumental variables from exogenous variables such as random shocks sampled from known distributions or micro-data generated by ACE models. Indeed, some scholars have already used neural network to predict endogenous explanatory variables for weak instrumental variables, suggesting that the good predictive power of neural networks provides more robust estimators for the one-stage estimation of 2SLS/GMM [Varian(2016)[60], Mullainathan and Spiess (2017)[42], Hartford et al.[26]]. It is important to note that the ACE-GAN methods can be used to generate not only instrumental variables but also proxy variables to solve the problem of lack of data by obtaining unknown data from known data.

GAN methods are often used to generate unstructured data such as video, images and text. All these data can be transformed into tensor, i.e., vectorized matrices. In fact, data including scenarios such as financial markets and online shopping can be transformed into tensor data. Therefore, GAN can extract data from high-frequency and high-dimensional data, while ACE provides the microscopic basis for data extraction. For example, engstrom2017 et al. (2017)[16] uses the amount of infrastructure identified from images as a measure of welfare level. By analogy, suppose now that we need to construct an indicator of the financing constraints[21] to firms , but obviously this indicator is not observable and it will not get robust conclusion to use data from listed companies with specificity. It can be considered to use the ACE model to construct a mapping f of observable variables such as return on assets and equity ratio Z to a set of variables A that indicate the degree of financing constraints, such as interest rate premium, i.e., $Z \to^f A$. The mapping f can be a complex neural network structure. The different indicators P between listed and unlisted companies of the same product category on the online shopping platform is then used as the output variable to form the ACE-GAN framework, i.e., $Z \to^f A \to^G \hat{P} \approx P$, where G denotes the predictor of the GAN. $P.P$ can be numerical variables such as number of positive reviews, sales volume, pricing differences, textual data such as customer reviews and news reports, or even satellite imagery. Since the Chinese government imposed environmental regulation policies and proposed carbon peaking, traditional industries such as steel and chemicals have faced stricter financing constraints. Due to the difficulty of financing, these companies cut their production capacity and thus reduce their polluting gas emissions, so the difference between the satellite cloud image and the rest of the region where these companies are located will change. Define the variables X and Y non-independently as $X \asymp Y$. As long as $Z \asymp A$ and $Z \asymp P$ are satisfied, $\hat{P}$ and A can be generated by ACE-GAN framework. In fact, comparing the before-after differences of $\hat{P}$ and P is a special Difference-in-Difference (in-Difference) estimator, and $Z \to^f A \to^G \hat{P} \approx P$ is also a special synthetic control method [1][4]. Thus $\hat{P} \approx P$ implies that it is reasonable to use $A = f(Z)$ as a proxy variable for the missing variable. Extending this conclusion further, the above conclusion is also reasonable when A is also a tensor, as it can be proved by replacing P with A.

4.3 ACE-DSGE model

Software

The software used for most ACE and DSGE studies are now not the same. Most DSGE models use *matlab* and *dynare* to build models, estimate parameters and perform simulation, but *dynare* does not yet support the solution and estimation of non-linear models. Non-linear conditions include, but are not limited to, inequality constraints, stochastic decisions of individuals, transitions in economic states, etc., which are essentially impossible to be considered in a linear system of equations. In addition, if the researcher defines more complex non-linear dynamical systems, this is also difficult to solve by the perturbation method employed by *dynare*. If the original equations are simplified using log-linearisation, important conclusions may be lost. Most of the ACE models are based on the *netlogo* and its built-in programming language. *netlogo* provides source codes and parameter setting functions from many ACE models, which allows researchers to reproduce and improve an ACE model. *netlogo* also provides visual graphing tools, which make it easy for researchers to export model results and use them directly in writing papers. However, to combine ACE and DSGE models, researchers need to write their models in the same computer language whenever possible.

The advantage of the *Python* implementation of the DSGE model is that I can easily use the third-party libraries such as $Numpy$,$Scipy$,$scikit-learn$ and $Tensorflow\&Keras$ to solve vector operations (in the presence of heterogeneous individual and sector summation problems), loss function optimisation, equation solving (in the presence of non-linear equations and implicit function) and also combining ACE and DSGE models.

Simulation analysis for complex nonlinear models

The simulation analysis of nonlinear Dynamic stochastic general equilibrium(DSGE) models is relatively different from ordinary DSGE models, the most obvious example being that the commonly used impulse response and variance decompositions are not directly obtained by log-linearizing the DSGE model. Therefore, the analysis of this non-linear DSGE model is based on stochastic simulations or steady-state analysis in different policy scenarios. Koop (1996)[33] proposed a generalised impulse response function for non-linear models. Lanne(2016)[36] proposed a variance decomposition method under non-linear models.

There is a fundamental difference between the ACE model and the DSGE model in terms of the object of analysis. To combine them, the variables in the ACE model need to be mapped or reduced to fewer dimensions of macroeconomic system variables such as the formation of product prices, typically, such as the summation of output in the heterogeneous intermediate goods sector or as part of the production function in the final goods sector. However, it is clear that this approach does not restore the true underlying micro-level mechanisms. I therefore combine the ACE model with the DSGE model in two ways: to find the game equilibrium (complete or imperfect information, dynamic or static, rational or animal spirits or even the introduction of noisy traders) and to investigate suitable solution methods (This will be discussed in the next subsection). In recent years, some scholars have started to incorporate heterogeneous individuals such as monopolists and price followers into the unified model, found the corresponding theoretical closed solutions and then used realistic data to invert the parameters of the model such as the monopolist's makeup[Miller et al.(2021)[40],Nocke et al.(2018)[43]]. Obviously the solution to the ACE model does not have an explicit solution as is generally the case with traditional monopolistic vendor models, but it can be assumed that the solution to the ACE model obeys an unknown conditional distribution that can be measured. The results of the game equilibrium point can be combined with other sectors of the economy. Firstly, decisions by other economic sectors can be used as exogenous variables to influence the solution of the ACE model. Secondly, the solution of the ACE model is endogenous to the general equilibrium of the macroeconomy, making the stochastic shock analysis more micro-founded by controlling for the heterogeneity parameters of the manufacturers for analysis by impact response function(IRF) and forecast error variance decomposition(FEVD).

4.4 Empirical Applications

I have two working papers. One[54] is already uploaded to arXiv. The other is ”Local Government-led Monetary Reservoir Mechanism in China: An Empirical Perspective”, which uses Invariant Prediction and granular instrumental variables mentioned in the section 3 for causal analysis. These two papers are the theoretical model and the empirical analysis of my master's thesis, so the latter is still available in Chinese, and the English version of the abstract is attached at here for reference: This paper first investigates the causal relationship between the local government fiscal leverage, the deposit to loan ratio of local financial institutions and the real estate sector using the Invariant Prediction method of Peters et al. (2016)[47] with a panel data of over 280 cities in China from 2000-2014. The empirical results show that local governments influence the real estate sector through the expansion of infrastructure development and welfare spending. Furthermore, it affects the deposit to loan ratio of local financial institutions and results in the monetisation of fiscal deficits with the expansion of local government credit. We refer to this mechanism as the monetary reservoir. Then, we use the granular instrumental variables method Gabaix & Koijen's (2020)[18] and the systematic equations model to exclude errors caused by possible biases and inconsistent estimators and to explore the specific effects of the mechanism. The results show that the monetary reservoir mechanism is significant and has a remarkable short-term impact on economic growth, firms' profits, workers' employment choices and asset allocation through several channels, but it weakens the effectiveness of the monetary policy. Finally, monetary pools also decrease inputs for long-run growth, which is evident in the inverted U-shape of the effects of fiscal leverage or house prices on economic growth and corporate profits.

5 My Research Experiences

My undergraduate dissertation ”Short-term trend study of the SSE Index” examines whether the short-term movements of the SSE Index are random (efficient market) or artificially disturbed. I developed a theoretical model of optimal investor decision making, the final condition of which is a second-order differential equation with varying coefficients. In the empirical part I used the Autoregressive Integrated Moving Average(ARIMA) model for the regression and obtained significant results, which showed that the short-term(within a trading day) movements of the SSE Index in China are mainly influenced by investors' historical information and risk preferences. This article was included in the album of the most excellent 100 undergraduate theses of that year by Shenzhen University and compiled for publication. Although there were many imperfections and inaccuracies in the paper, it did stimulate my interest in academic research.

After I became a postgraduate student, my supervisor, Professor Xiao Shuguang, gave me a lot of guidance and involved me in his research. Professor Xiao's main research interests are game theory and industrial organisation theory. He is a PhD supervisor in Quantitative Economics and a ”Min River Scholar” in Fujian Province. He had published some articles in famous journal in China such as ”China Industrial Economics”, ”Nankai Business Review” and ”Accounting Research”. This year, he is leading a National Social Science Foundation project named ”Research on the Reform of New Mixed Ownership Enterprises and the Drive for High-Quality Development” (Grant No. 21BJL010), and I am fortunate to be involved in his research project. I have now completed a paper on ”China's Easily Overlooked Monetary Transmission Mechanism: Moneytary Reservoir”[54](Prof. Xiao is the corresponding author and the first co-author). In the process of learning from Professor Xiao, I have also made full use of my own strengths. I had worked as a data analyst in an internet company, so I am quite proficient in using Python. on the other hand, I had enrolled in the entrance exam of academic master programme of Finance at Renmin University of China, which requires a high level of mathematical economics and econometrics knowledge, and I eventually ranked first among all candidates in the same programme in my economics subject (but unfortunately my results in other subjects were not as good), so I had a relatively solid grounding in mathematical economics and econometric. Given these two reasons, I chose the DSGE model, which is able to combine programming and mathematical analysis, as my main research tool. As the system of knowledge in DSGE models is very complex, in addition to the classic DSGE textbook., I have read "Econometric anaysis" by William H. Greene, "Advanced Macroeconomics" by David Romer, "Microeconomic Theory" (mainly game theory and information economics) by Mas-Colell, Whinston & Greenand and "Bayesian Econometrics" by Gary Koop. While reading articles in the world-renowned economics journal such as "American Economics Review", I noticed that computational economics methods were appearing more frequently each year, so I read "Agent- Based Modelling in Economics" by Lynne Hamill & Nigel Gilbert to gain some insight into the ACE models. At the same time, I also noticed a trend in recent years of combining machine learning methods with classical statistical theory in prestigious journals such as "Econometrica" and the "Journal of the American Statistical Association". I realised that the three kinds of models might be able to complement each other to form a more effective and practical framework for theoretical analysis and empirical research. With this in mind, I have taken courses on machine learning and deep learning by Andrew Ng, "Deep Learning" by Ian Goodfellow, Yoshua Bengio & Aaron Courville and "Machine Learning" by Zhihua Zhou. After reading these books, I became more interested in generative neural networks and reviewed more related literature, so I started to study statistic and computational models combining generating adversarial networks, agent-based economics models and dynamic stochastic general equilibrium models(GAN-ACE-DSGE) for policy evaluation.

Although at this stage I am mainly concerned with Chinese economic issues, economic theory does not exist at national borders. For example, the problems of government debt and monetary overdraft that I studied in my master's degree in fact occurred in Europe, the United States and Japan as well. Their essential differences are determined by their respective national polities and histories, while their essential convergence implies the universality of economic theories. Personally, I believe that many current problems in economics (both traditional themes and various new branches) are actually causal mechanism studies and the estimation of causal effect. If economists make mistakes in understanding the causal mechanisms behind economic problems, wrong policies will be proposed and implemented, resulting in wasted resources or even irreversible consequences. Therefore, I would like to do theoretical research by studying causal inference models to upgrade methodological models. This is the source of my motivation and enthusiasm to apply for the PHD program.

References

[1] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

[2] Raghavendra Addanki and Shiva Prasad Kasiviswanathan. Collaborative causal discovery with atomic interventions, 2021.

[3] Sina Akbari, Ehsan Mokhtarian, AmirEmad Ghassami, and Negar Kiyavash. Recursive causal structure learning in the presence of latent variables and selection bias, 2021.

[4] Muhammad Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852, 2018.

[5] Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics*. Princeton university press, 2008.

[6] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[7] Oriana Bandiera, Iwan Barankay, and Imran Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

[8] Andrew Bennett, Nathan Kallus, and Tobias Schnabel. Deep generalized method of moments for instrumental variable analysis. *ArXiv*, abs/1905.12495, 2019.

[9] Olivier Blanchard. Do dsge models have a future. *DSGE Models in the Conduct of Policy: Use as intended*, page 93, 2016.

[10] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

[11] Peter Bühlmann, J. Peters, and Jan Ernest. Cam: Causal additive models, high-dimensional order search and penalized regression. *ArXiv*, abs/1310.1533, 2013.

[12] Leonardo Bursztyn, Florian Ederer, Bruno Ferman, and Noam Yuchtman. Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301, 2014.

[13] Steve Cicala. Imperfect markets versus imperfect regulation in us electricity generation. Technical report, National Bureau of Economic Research, 2017.

[14] Thomas D Cook. "waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2):636–654, 2008.

[15] Antonio Coppola, Rishabh D Guha, and Dong Ryeol Lee. Granular instrumental variables. 2019.

[16] Ryan Engstrom, Jonathan Samuel Hersh, and David Locke Newhouse. Poverty from space: using high-resolution satellite imagery for estimating economic well-being. *World Bank Policy Research Working Paper*, (8284), 2017.

[17] Joshua M Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up.* Brookings Institution Press, 1996.

[18] Xavier Gabaix and Ralph SJ Koijen. Granular instrumental variables. Technical report, National Bureau of Economic Research, 2020.

[19] Jordi Galí. Some scattered thoughts on dsge models. *DSGE Models in the Conduct of Policy; Use as Intended*, pages 86–92, 2017.

[20] Mauro Gallegati and Matteo G Richiardi. Agent based models in economics and complexity., 2009.

[21] Shan Ge. How do financial constraints affect product pricing? evidence from weather and life insurance premiums. *The Journal of Finance*, 77(1):449–503, 2022.

[22] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

[23] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[24] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

[25] Stephen Hansen, Michael McMahon, and Andrea Prat. Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870, 2018.

[26] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, pages 1414–1423. PMLR, 2017.

[27] Jason S. Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep iv: A flexible approach for counterfactual prediction. In *ICML*, 2017.

[28] Christina Heinze-Deml, J. Peters, and Nicolai Meinshausen. Invariant causal prediction for nonlinear models. *Journal of Causal Inference*, 6, 2018.

[29] Kevin D Hoover. Causality in economics and econometrics. 2006.

[30] Patrik O. Hoyer, Shohei Shimizu, Antti J. Kerminen, and Markus Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *Int. J. Approx. Reason.*, 49:362–378, 2008.

[31] Jean Kaddour, Yuchen Zhu, Qi Liu, Matt J Kusner, and Ricardo Silva. Causal effect inference for structured treatments. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[32] Nobuhiro Kiyotaki and John Moore. Credit cycles. *Journal of political economy*, 105(2):211–248, 1997.

[33] Gary Koop, M Hashem Pesaran, and Simon M Potter. Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147, 1996.

[34] Anton Korinek. 7. thoughts on dsge macroeconomics: Matching the moment, but missing the point? In *Toward a Just Society*, pages 159–173. Columbia University Press, 2018.

[35] Finn E Kydland and Edward C Prescott. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370, 1982.

[36] Markku Lanne and Henri Nyberg. Generalized forecast error variance decomposition for linear and nonlinear multivariate models. *Oxford Bulletin of Economics and Statistics*, 78(4):595–603, 2016.

[37] Sanghack Lee and Elias Bareinboim. Causal identification with matrix equations. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[38] Michael P Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293, 2022.

[39] Ramon Marimon, Ellen McGrattan, and Thomas J Sargent. Money as a medium of exchange in an economy with artificially intelligent agents. *Journal of Economic dynamics and control*, 14(2):329–373, 1990.

[40] Nathan H Miller, Gloria Sheu, and Matthew C Weinberg. Oligopolistic price leadership and mergers: The united states beer industry. *American Economic Review*, 111(10):3123–59, 2021.

[41] Joris M. Mooij, J. Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: Methods and benchmarks. *ArXiv*, abs/1412.3773, 2016.

[42] Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

[43] Volker Nocke and Nicolas Schutz. Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica*, 86(2):523–557, 2018.

[44] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[45] Judea Pearl et al. Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19:2, 2000.

[46] Peters. Causal discovery with continuous additive noise models. 2014.

[47] J. Peters, Peter Buhlmann, and Nicolai Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. *arXiv: Methodology*, 2015.

[48] J. Peters, Dominik Janzing, and Bernhard Schölkopf. Elements of causal inference: Foundations and learning algorithms. 2017.

[49] Raanan Y. Rohekar, Shami Nisimov, Yaniv Gurwicz, and Gal Novik. Iterative causal discovery in the possible presence of latent confounders and selection bias, 2021.

[50] Paul Romer. The trouble with macroeconomics. —-, 2016.

[51] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[52] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[53] Thomas C Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186, 1971.

[54] Xiao Shuguang and Lai Xinglin. China's easily overlooked monetary transmission mechanism:real estate monetary reservoi, 2021.

[55] Rahul Singh, Maneesh Sahani, and Arthur Gretton. Kernel instrumental variable regression. In *NeurIPS*, 2019.

[56] Joseph E Stiglitz. Where modern macroeconomics went wrong. *Oxford Review of Economic Policy*, 34(1-2):70–106, 2018.

[57] Xinwei Sun, Botong Wu, Xiangyu Zheng, Chang Liu, Wei Chen, Tao Qin, and Tie-Yan Liu. Recovering latent causal factor for generalization to distributional shifts. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021.

[58] Gabriele Tedeschi, Giulia Iori, and Mauro Gallegati. Herding effects in order driven markets: The rise and fall of gurus. *Journal of Economic Behavior & Organization*, 81(1):82–96, 2012.

[59] Donald L Thistlethwaite and Donald T Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, 51(6):309, 1960.

[60] Hal R Varian. Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, 113(27):7310–7315, 2016.

[61] Haiyuan Wan. The unintended consequences of urban community infrastructure investment on entrepreneurship in china. *Economic Research Journal*, 9:39–55, 2021.

[62] Coady Wing, Kosali Simon, and Ricardo A Bello-Gomez. Designing difference in difference studies: best practices for public health policy research. *Annual review of public health*, 39, 2018.

[63] Matej Zečević, Devendra Singh Dhami, Athresh Karanam, Sriraam Natarajan, and Kristian Kersting. Interventional sum-product networks: Causal inference with tractable probabilistic models, 2021.