All Causes in the Model and in Reality

Dionissi Aliprantis

April 11, 2022

Abstract: Causal effects are parameters of models. Using causal effects to accurately predict the consequences of actions requires assuming that all causes in the model capture all causes in reality, or in the Data Generating Process (DGP). Rubin Causal Models, Structural Causal Models, and Econometric Causal Models differ in their approaches to all causes assumptions (ACAs) about the DGP in the future. Explicitly stating ACAs helps to explain how disagreements can persist about the credibility of different causal methodologies and focuses attention on the assumptions necessary for prediction.

Keywords: Rubin Causal Model, Structural Causal Model, Econometric Causal Model, Data Generating Process, All Causes Assumption, The Problem of Induction

JEL Classification Codes: C00, B23

Federal Reserve Bank of Cleveland, *dionissi.aliprantis@clev.frb.org*.

I thank Elias Bareinboim, Greg Caetano, and Francisca G.-C. Richter for helpful comments.

This paper updates material previously circulated in a working paper under the title "A Distinction between Causal Effects in Structural and Rubin Causal Models."

1 Introduction

We can choose to act in many different ways. For example, we can choose whether to eat a high-salt diet and whether to emit carbon dioxide. Knowing the consequences of our actions allows us to make choices that create the type of world in which we would like to live.

Predicting the consequences of our actions is difficult because the world is constantly changing. Causal effects describe relationships that do not change across some combination of space, time, and actions. Therefore, causal effects are useful tools for predicting the consequences of our actions, which could be considered the goal of science.[1]

The desire to predict is why many social scientists are more interested in estimating causal effects than descriptive statistics.[2] The assumptions necessary to describe a distribution are typically much weaker than the assumptions necessary to undertake a causal analysis. We are willing to adopt stronger causal assumptions, knowing that they may be false, because doing so aids our ability to predict. In other words, the goal of prediction is what leads to our interest in causal statements made with lower confidence over descriptive statements made with higher confidence.

This note studies the use of causal effects to predict the consequences of actions. I build on Heckman (2005) and state three major tasks that this entails. The first task is representing reality with a model. The second task is judging when this representation is accurate, both in the past and in the future. The third task is actually making predictions.

The second task of relating the model to reality requires assumptions that the model accurately represents reality. A first subtask is judging when model parameters represent the outcomes of a past action. This requires that we observe a variable that mimicked an action manipulating treatment, inducing variation in treatment but no other cause in the all causes outcome equation. A second subtask is judging when model parameters will predict the outcomes of a future action. This requires that either (i) the outcome equation in reality remains stable between the time of the data used in estimation of a causal effect and the time for prediction, or (ii) changes to reality's outcome equation can be accurately modeled by the social scientists. Both of these subtasks can be stated as assumptions about the relationship between all causes outcome equations in the model and in reality.

Explicitly stating all causes assumptions (ACAs) helps to explain how disagreements can persist about the credibility of different causal methodologies and focuses attention on the assumptions necessary for prediction. Rubin Causal Models are designed to capture a Past ACA that an observed variable mimicked a past action, with predictions made under a Future ACA that the outcome equation is stable across settings. Structural Causal Models are designed to accommodate a Future ACA that incorporates a changing outcome equation, and Econometric Causal Models are designed to accommodate a Future ACA under such changes resulting from optimization and general equilibrium.

[1] Feynman (1999) notes that "science can be defined as a method for... trying to answer only questions which can be put into the form: If I do this, what will happen?"

[2] Chapter 1 of Angrist and Pischke (2008) elaborates on this point.

2 Tasks Necessary for Prediction with Causal Effects

One way of stating the goal of scientists is: We aim to use a model and past observations of reality to make accurate predictions about how reality will evolve in the future, especially under hypothetical actions. Here we are careful to distinguish between a model and reality. A model is a mental world where (mental) outcomes are generated by logically following a set of rules. Reality is the world we live in where (actual) outcomes are generated by a Data Generating Process (DGP).

Using causal effects to achieve the goal of prediction requires several distinct tasks. Table 1 describes these tasks. Task 1 is related to representing reality with a model and comes from Table 1 of Heckman (2005). Task 2 is related to judging whether one's model accurately characterizes how reality would respond under alternative future actions. Task 3 is the step of actually making predictions.

Table 1: Six Distinct Tasks Required for Prediction with Causal Effects

Task/Subtask	Description	Requirements
1	<u>Representing Reality with a Model</u>	
1.1	Defining the Set of Hypotheticals or Counterfactuals	A Scientific Theory (ie, A Model)
1.2	Identifying Model Parameters (Causal or Otherwise) from Hypothetical Population Data	Mathematical Analysis of Point or Set Identification
1.3	Identifying Model Parameters from Real Data	Estimation and Testing Theory
2	<u>Judging the Model's Relationship with Reality</u>	
2.1	Judging whether Model Parameters Represent Outcomes of Past Action	a) Comparison of Past Predictions with Past Observations b) Evidence and Reflection on Exclusion Restriction
2.2	Judging whether Model Parameters Will Predict Outcomes of Future Action	a) Comparison of Past Predictions with Past Observations b) Reflection on the Stability of the Model's Mechanisms in Future Environments
3	<u>Prediction with Causal Effects from the Model</u>	
3.1	Making Predictions about Outcomes in Reality under Future Actions	A Model and Data after Completing Tasks 1 and 2

Note: Tasks 1.1-1.3 are from Table 1 in Heckman (2005).

3 Representing Reality with Models

Suppose that our goal of prediction pertains to the variable Y in a DGP in which the "all causes" equation determining the outcome for individual i is $Y_i = g(D_i, \varepsilon_{Yi})$. There are unobserved factors ε_{Yi} driven by a group of non-mediators $L_{1i}, \ldots, L_{ji}$ and a group of mediators $M_{1i}, \ldots, M_{ki}$. An action A or a passively observed variable Z_i determines treatment D_i, possibly along with unobserved factors ε_{Di}. Note that all causes are illustrated in the figure below only for the outcome variable. All of the analysis below is considered for a single value of some fixed demographic variables $X_i = x$.

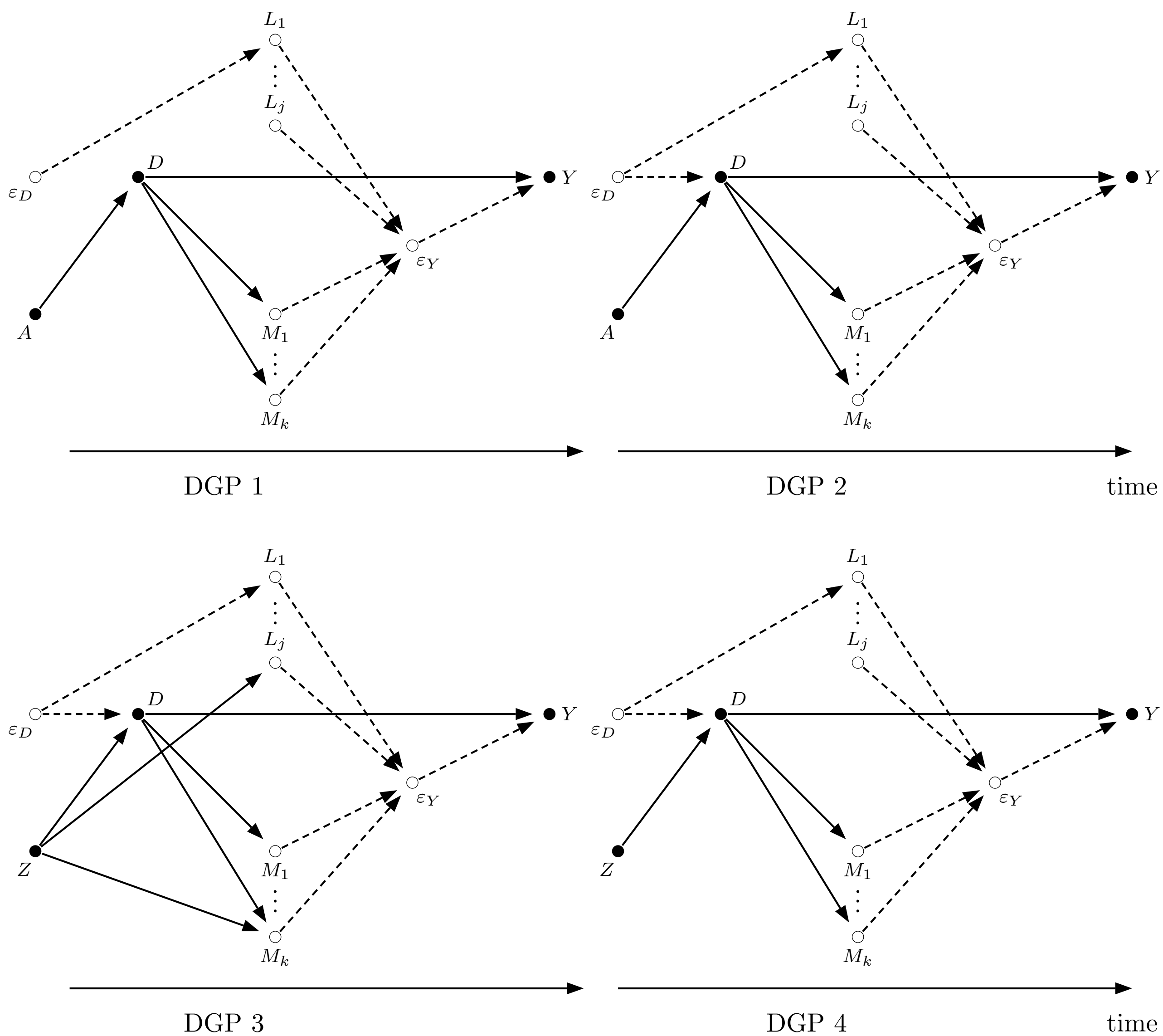

Figure 1: Four Possible Realities, or Data Generating Processes, for the Outcome Variable Y
Note: This figure follows the convention from Pearl (2009) of communicating that a variable is observed by drawing a solid marker and a solid line to its descendants and communicating that a variable is unobserved by drawing a hollow marker and dashed line to its descendants.

Beginning with Subtask 1.1, we want to represent in our minds how the outcome variable would react to actions we might take determining the treatment variable. Three approaches would be to use the Rubin Causal Model (RCM), a Structural Causal Model (SCM), or an Econometric Causal Model (ECM). The outcome equation in the RCM would be defined in terms of Potential Outcomes (POs) as $Y_i = Y_i(D_i)$. In an SCM or ECM the outcome equation, written as $Y_i = f_Y(D_i, U_{Yi})$, includes the variable U_{Yi} that helps determine the distribution of POs. A variable entering as an argument of $Y_i(\cdot)$ or $f_Y(\cdot)$ represents a variable human beings can act to control, even if only partially, that in the model would result in changes to the value of Y_i. Theory might push us to consider adding an intermediate variable, or mediator, to the outcome equation of the model as either $Y_i = Y_i(D_i, M_{1i})$ or $Y_i = f_Y(D_i, , M_{1i}, U_{Yi})$. We might also consider adding another parent of Y_i into the outcome equation as $Y_i = Y_i(D_i, L_{1i}, M_{1i})$ or $Y_i = f_Y(D_i, L_{1i}, M_{1i}, U_{Yi})$.

Suppose that we were to model the DGP in two ways. In the first approach we would only include treatment as an argument of the outcome equation. In the second approach we would include the parents L_{1i} and M_{1i} in the outcome equation.[3] With RCMs this could be represented as:

Model 1

$$\begin{aligned} D_i &= D_i(Z_i) \\ Y_i &= Y_i(D_i) \end{aligned}$$

Model 2

$$\begin{aligned} D_i &= D_i(Z_i) \\ M_i &= M_i(D_i) \\ Y_i &= Y_i(D_i, L_{1i}, M_{1i}) \end{aligned}$$

Following Imbens and Rubin (2008), we can define each RCM in terms of its outcome equation and selection equation. The selection equations would be defined as $D_i(Z_i)$ in both models. The outcome equation would be $Y_i(D_i)$ in the first model and $Y_i(D_i, L_{1i}, M_{1i})$ in the second model.

With SCMs or ECMs this approach to modeling the DGP might be written as:

Model 1

$$\begin{aligned} Z_i &= f_Z^1(U_{Zi}^1) \\ D_i &= f_D^1(Z_i, U_{Di}^1) \\ Y_i &= f_Y^1(D_i, U_{Yi}^1) \end{aligned}$$

Model 2

$$\begin{aligned} X_{1i} &= f_{X1}^2(U_{X1i}^2) \\ Z_i &= f_Z^2(X_{1i}, U_{Zi}^2) \\ D_i &= f_D^2(Z_i, U_{Di}^2) \\ M_{1i} &= f_{M1}^2(D_i, U_{M1i}^2) \\ X_{2i} &= f_{X2}^2(U_{X2i}^2) \\ L_{1i} &= f_{L1}^2(X_{2i}, U_{L1i}^2) \\ Y_i &= f_Y^2(D_i, L_{1i}, M_{1i}, U_{Yi}^2) \end{aligned}$$

[3] $R_{\max}$, the assumed maximum R^2 of an outcome equation, is assumed to be 1 in Altonji et al. (2005) and allowed to be less than 1 in Oster (2019). One interpretation of $R_{\max}$ is the R^2 obtained when all variables L and M that are observable at a given scale of observation are included in the researcher's model and data set.

In terms of a fully specified model following Definition 7.1.1 in Pearl (2009) and Definition 1 in Pearl and Bareinboim (2014), we can define each SCM as (i) a set of background variables U^1 and U^2; (ii) a set of endogenous variables $V^1 = \{Z, D, Y\}$ and $V^2 = \{X_1, Z, D, M_1, X_2, L_1, Y\}$; (iii) the set of functions $F^1 = \{f^1_Z, f^1_D, f^1_Y\}$ and $F^2 = \{f^2_{X1}, f^2_Z, f^2_D, f^2_{M1}, f^2_{X2}, f^2_{L1}, f^2_Y\}$; and (iv) a joint probability distribution over the background variables, $P^1(U^1)$ and $P^2(U^2)$.

4 Judging the Model's Relationship with Reality

4.1 Representing the Result of a Past Action

We can imagine one reality in which we are able to act with complete control over treatment (DGP 1) and another reality in which our action influences treatment status in combination with some unobserved factors ε_D (DGP 2). We would like to be able to make predictions for each individual, but due to the fundamental problem of causal inference, being unable to observe individuals in each treatment state, we are forced to infer in DGP 1 the Average Treatment Effect (ATE, Holland (1986)). Alternatively, we might infer effects for some subpopulation in DGP 2 with a binary or discrete action using the Local Average Treatment Effect (LATE, Imbens and Angrist (1994)) or in DGP 2 with a continuous action using the Marginal Treatment Effect (MTE, Heckman and Vytlacil (2005)).

Identification of a LATE or MTE with an instrumental variable requires that the instrumental variable satisfies an independence assumption and an exclusion restriction. The exclusion restriction can be stated as an all causes assumption.

Past All Causes Assumption (ACA) : The variation in treatment caused by one observed variable Z mimicked an action taken to control that variable, in the sense that the observed variable influenced treatment D and no other cause in the all causes outcome equation.

Identification of the ATEs for all individuals requires a strengthening of the exclusion restriction to the variable Z having full control over treatment while influencing no other cause in the all causes outcome equation. This stronger all causes assumption would allow us to identify $\mathbb{E}[Y|do(D = d)]$, or the effect of an ideal manipulation of D, which is an intervention setting the value of the variable(s) in question but making no other change to the system (Scheines (1997), Eberhardt and Scheines (2007)).

Finding an instrument, or a variable Z for which the Past ACA holds, represents a tremendous demand on the modeler in observational settings. Many researchers therefore focus on design-based approaches to ensuring the Past ACA holds, analyzing outcomes in settings with randomized and quasi-randomized treatments (Imbens (2010), Angrist and Pischke (2010)).

4.2 Predicting the Result of a Future Action

Of what use are the causal effects we have identified in data from the past? Having specified a model to represent reality and judged that model's relationship with reality, we now turn our attention to the future.

We consider Subtask 2.2 and Task 3 in terms of a decision problem: Should we take the action $A = 1$ over the status quo $A = 0$? We could frame this policy question in terms of prediction of the benefit to some outcome Y. If the cost of taking an action is $C(A = a)$, we might follow a decision rule of taking action $A = 1$ over the status quo $A = 0$ if the expected net benefit is greater under the action. Where $\mathbb{1}\{\cdot\}$ is an indicator function, we could write this decision rule as:

$$A = \mathbb{1}\Big\{ \ \ \mathbb{E}_i[Y_i|A=1] - C(A=1) \ \ > \ \ \mathbb{E}_i[Y_i|A=0] - C(A=0) \ \ \Big\}. \tag{1}$$

We will need to engage in Task 3, predicting $\mathbb{E}_i[Y_i|A = a]$.[4] For Subtask 2.2, now the question becomes judging the credibility of predictions generated by our estimated causal effects. These predictions would be

$$\begin{aligned}\underline{\text{Model 1}}: \quad \mathbb{E}_i[Y_i|A=a] =& \mathbb{E}_i[f_Y^1(D_i(a), U_{Yi}^1)] \\ =& \mathbb{E}_i[f_Y^1(f_D^1(a, U_{Di}^1), U_{Yi}^1)]\end{aligned}$$

$$\begin{aligned}\underline{\text{Model 2}}: \quad \mathbb{E}_i[Y_i|A=a] =& \mathbb{E}_i[f_Y^2(D_i, L_{1i}, M_{1i}, U_{Yi}^2)] \\ =& \mathbb{E}_i\Big[f_Y^2\Big(\ f_D^2(a, U_{Di}^2) \ , \ f_{L1}^2(X_{2i}, U_{L1i}^2) \ , \ f_{M1}^2\Big(f_D^2(a, U_{Di}^2) \ , \ U_{M1i}^2\Big), U_{Yi}^2\Big)\Big]\end{aligned}$$

If we have estimates of $\mathbb{E}_i[Y_i|A = 1]$ and $\mathbb{E}_i[Y_i|A = 0]$ for the action a in the past, we can just fill these estimates into Equation 1 under the following assumption:

Stable Future – All Causes Assumption (ACA) : All causes in the outcome equation are stable between the time of the data used in estimation and the time for which a prediction is made.

The Stable Future ACA is often made by finding a sample or context in which the outcomes equation is likely to remain stable, perhaps by matching or balancing on covariates (Nie et al. (2021); Olsen (2022)).

Alternatively, we might invoke the following assumption:

Modeled Future – All Causes Assumption (ACA) : All causes in the outcome equation are known and can be accurately modeled for the time at which a prediction is made.

The Modeled Future ACA allows for predictions to be made in the case that features of the outcome equation changes over time. Suppose, for example, it was known that mediator M1 would drop out of the outcome equation between the time of the past data and the future relevant for prediction, so that the DGP determined outcomes as if $M1$ was set at the fixed level m_1^*. In this case, one could not make predictions using estimates from Model 1, but could make predictions based on

[4]See Manski (2008) for a discussion of identification and decision problems.

Model 2 as:

$$\underline{\text{Model 2}}: \quad \mathbb{E}_i[Y_i|A=a] = \mathbb{E}_i[f_Y^2(D_i, L_{1i}, m_{1i}^*, U_{Yi}^2)]$$
$$= \mathbb{E}_i\left[f_Y^2\left(\, f_D^2(a, U_{Di}^2)\,,\, f_{L1}^2(X_{2i}, U_{L1i}^2)\,,\, m_1^*\,, U_{Yi}^2\right)\right].$$

Obviously the Modeled Future ACA is a strong assumption. The alternative is to only make predictions in situations where the outcome equation remains stable over time and context.

Future ACAs can be cast in terms of the stability of the functional relationships F and the distribution of background variables $P(U)$ necessary for the model to accurately represent reality in the past. As Pearl (2009) writes, "It is the persistence across time of U and $f(x,u)$ that endows counterfactual expressions with predictive power; absent this persistence, the counterfactual loses its obvious predictive utility" (Section 7.2.2, p 219).

4.2.1 Example: Causal Effects of Neighborhoods

We consider Task 2 in terms of a practical example. Suppose that we are concerned about concentrated poverty in the United States. We are considering taking an action $A = 1$ over the status quo $A = 0$ that would change the subsidies and search supports offered through public housing vouchers in the US. The goal of such a policy would be to improve the socioeconomic status (SES) of program participants' neighborhoods, and we might suppose for the sake of exposition that we know the change in neighborhood SES the policy action would generate, as well as the cost of the action and the status quo (Collinson and Ganong (2018); Aliprantis et al. (2022, 2021); Bergman et al. (2020)). Let Y be income and recalling the decision rule

$$A = \mathbb{1}\Big\{\ \mathbb{E}_i[Y_i|A=1] - C(A=1)\ >\ \mathbb{E}_i[Y_i|A=0] - C(A=0)\ \Big\}, \tag{2}$$

consider whether we should take the action $A = 1$ over the status quo $A = 0$.

Assume that Task 1 and Subtask 2.1 have been accomplished using Model 1 to quasi-experimentally estimate effects of neighborhood SES (See Aliprantis and Richter (2020) or Pinto (2019).). That is, we have estimates of $\mathbb{E}_i[Y_i|A=1] = \mathbb{E}_i[f_Y^1(D_i(1, U_{Di}^1), U_{Yi}^1)]$ for some previous action a. In this case, we can just fill these estimates into Equation 2 under the Stable Future–ACA.

The parsimonious model specification in Model 1 that aids credibly addressing the Past ACA through design appears to make a Future ACA more difficult to adopt. There are many reasons that the Stable Future–ACA may be a strong and unrealistic assumption. For the neighborhood effects estimated in Model 1, D may only have been observed over a limited range (Aliprantis et al. (2021)), but general equilibrium effects could be an important part of f_Y^1 estimated on this limited range in past data (Agostinelli et al. (2020)). Likewise, the f_Y^1 and U_{Yi}^1 representing reality are likely to be unstable. *Any* un-modeled causes of Y like M_1 or L_1 affected by policy will change either the functional relationship f_Y^1 or the distribution of U_{Yi}^1 that accurately describes reality. Consider that several non-mediators L in the DGP have been the target of major policies, either around the time of the experimental estimates or afterwards. Some examples include welfare reform (Blank (2002)) and the Affordable Care Act. Consider also that several mediators M in the DGP have additionally been the target of major policy initiatives. Some examples include the Jobs-Plus Program providing job training and child care subsidies to low-income neighborhoods (de Souza Briggs et al. (2010)), and Enterprise Zones (EZs) aiming to increase labor demand in low-income neighborhoods (Neumark and Young (2021); Arefeva et al. (2021)).[5]

L in the DGP

1. welfare (social insurance)
2. childcare subsidies
3. health insurance
4. incarceration and related policies

M in the DGP

1. job training programs
2. local labor demand
3. public transportation

The cases above highlight why adopting the Stable Future–ACA could be less credible in some circumstances than adopting a Modeled Future–ACA. A Modeled Future–ACA could be made using Model 2 in which the parent L or mediator M affected by policy are arguments of the structural equation f_Y^2. Supposing that causal effects had been estimated (ie, Task 1 and Subtask 2.1 had been completed), we could predict the future response to our actions based on the causal effects from Model 2 as

$$\mathbb{E}_i[Y_i|A=a] = \mathbb{E}_i\left[f_Y^2\left(\ f_D^2\big(a, U_{Di}^2\big)\ ,\ f_{L1}^2\big(X_{2i}, U_{L1i}^2\big)\ ,\ f_{M1}^2\Big(f_D^2\big(a, U_{Di}^2\big)\ ,\ U_{M1i}^2\Big), U_{Yi}^2\right)\right]$$

after performing appropriate changes to changing features of the outcome equation. For example, looking at DGPs 1-4, one could more easily imagine that a policy could alter $f_{L1}^2(\cdot)$ or $f_{M1}^2(\cdot)$ without changing f_Y^2, U_Y^2, or $f_D^2\big(a, U_{Di}^2\big)$ than one could imagine that a policy could alter L_1 or M_1 without changing f_Y^1 or U_Y^1. Likewise, theory would be needed to make credible predictions about any function outside the support of the data observed in the past.

5 Conclusion

Approaches to the Future ACA appears to be where different causal methodologies diverge. There is no distinction between the causal effects expressed in Rubin and Structural Causal Models (Galles and Pearl (1998)).[6] There are some differences, however, with approaches to prediction. With RCMs, researchers typically consider estimates of causal effects from a range of circumstances, and then find those with F and $P(U)$ most like those believed to occur in the future to invoke a Stable Future–ACA, perhaps by matching or balancing on covariates (Nie et al. (2021); Olsen (2022)). With SCMs, a focus is on expanding the model to include all causal pathways of which the researcher is aware to credibly assume a Modeled Future–ACA. Like the expansion between Models

[5]This discussion is entirely focused on the outcome equation, but social interactions in the neighborhood sorting process itself may have changed (Sobel (2006); Manski (2013)).

[6]See related discussions in Markus (2021) and Weinberger (2021).

1 and 2, this modeling allows one to formalize the approach to judging the stability or changes in the F and $P(U)$ accurately representing reality over time (Pearl and Bareinboim (2014)). That is, adding claims about variables like L_1 and M_1 in Model 2 allows researchers to assess more specific claims about outcome equation stability in the future.

Another distinction emerges with the approach of Econometric Causal Models. This approach estimates causal effects within a model that is guided from the start with a focus on the Modeled Future–ACA (Hansen (2021), Keane (2010)); Heckman and Pinto (2022) argue that this focus is the distinguishing feature of ECMs.[7] For example, an ECM may be specified to incorporate general equilibrium effects that will be important for predicting future outcomes because they are either outside the support of data from the past or generated by a DGP where general equilibrium effects will interact in important ways with other mechanisms. The difficulty of using statistical evidence in favor of a Future ACA can help to explain the appeal of calibrating models, which one can interpret as an attempt to credibly adopt the Modeled Future–ACA by modeling critical features of the DGP that are dynamically changing or policy-relevant (Lucas (1976), Kydland and Prescott (1977)).

References

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020). It takes a village: The economics of parenting with neighborhood and peer effects. *NBER Working Paper 27050*.

Aliprantis, D., H. Martin, and D. C. Phillips (2022). Landlords and access to opportunity. *Journal of Urban Economics 129*.

Aliprantis, D., H. Martin, and K. Tauber (2021). What determines the success of housing mobility programs? *Mimeo., Federal Reserve Bank of Cleveland*.

Aliprantis, D. and F. G.-C. Richter (2020). Evidence of neighborhood effects from Moving to Opportunity: LATEs of neighborhood quality. *The Review of Economics and Statistics 102*(4), 633–647. DOI: 10.1162/rest_a_00933.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). An evaluation of Instrumental Variable strategies for estimating the effects of Catholic schooling. *Journal of Human Resources 40*(4), 791–821.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

[7] There can be overlap in approaches; White and Chalak (2009) incorporates features helpful for assuming the Modeled Future–ACA from ECMs into SCMs.

Arefeva, A., M. A. Davis, A. C. Ghent, and M. Park (2021). Job growth from Opportunity Zones. *SSRN Working Paper 3645507*.

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz, and C. Palmer (2020). Creating Moves to Opportunity: Experimental evidence on barriers to neighborhood choice. *NBER Working Paper 26164*.

Blank, R. M. (2002). Evaluating welfare reform in the United States. *Journal of Economic Literature 40*(4), 1105–1166.

Collinson, R. and P. Ganong (2018). How do changes in housing voucher design affect rent and neighborhood quality? *American Economic Journal: Economic Policy 10*(2), 62–89.

de Souza Briggs, X., S. J. Popkin, and J. Goering (2010). *Moving to Opportunity: The story of an American experiment to fight ghetto poverty*. Oxford University Press.

Eberhardt, F. and R. Scheines (2007). Interventions and causal inference. *Philosophy of Science 74*(5), 981–995.

Feynman, R. P. (1999). *The Pleasure of Finding Things Out*. Basic.

Galles, D. and J. Pearl (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science 3*(1), 151–182.

Hansen, B. E. (2021). *Econometrics*. Princeton University Press.

Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology 35*(1), 1–97.

Heckman, J. J. and R. Pinto (2022). Causality and econometrics. *NBER Working Paper 29787*.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association 81*(396), 945–960.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature 48*(2), 399–423.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Imbens, G. W. and D. B. Rubin (2008). Rubin Causal Model. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). London: Palgrave Macmillan.

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics 156*(1), 3–20.

Kydland, F. E. and E. C. Prescott (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy 85*(3), 473–491.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Markets*, Volume 1, pp. 19–46. Carnegie-Rochester Conference Series on Public Policy.

Manski, C. F. (2008). *Identification for Prediction and Decision*. Harvard University Press.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal 16*(1), S1–S23.

Markus, K. A. (2021). Causal effects and counterfactual conditionals: Contrasting Rubin, Lewis and Pearl. *Economics & Philosophy*, 1–21.

Neumark, D. and T. Young (2021). Heterogeneous effects of state Enterprise Zone programs in the shorter run and longer run. *Economic Development Quarterly 35*(2), 91–107.

Nie, X., G. Imbens, and S. Wager (2021). Covariate balancing sensitivity analysis for extrapolating randomized trials across locations. *arXiv preprint arXiv:2112.04723*.

Olsen, E. T. R. B. (2022). *Enhancing the Generalizability of Impact Studies in Education* (NCEE 2022-003 ed.). Washington, DC: US Department of Education, Institute of Education Science.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

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

Pearl, J. and E. Bareinboim (2014). External validity: From do-calculus to transportability across populations. *Statistical Science 29*(4), 579–595.

Pinto, R. (2019). Noncompliance as a rational choice: A framework that exploits compromises in social experiments to identify causal effects. *Mimeo., UCLA*.

Scheines, R. (1997). An introduction to causal inference. In V. McKim and S. Turner (Eds.), *Causality In Crisis?*, pp. 185–199. University of Notre Dame Press.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association 101*(476), 1398–1407.

Weinberger, N. (2021). Comparing Rubin and Pearl's causal modeling frameworks: A commentary on Markus (2021). *Mimeo., University of Pittsburgh*.

White, H. and K. Chalak (2009). Settable systems: An extension of Pearl's Causal Model with optimization, equilibrium, and learning. *Journal of Machine Learning Research 10*(8).