10

Reintroducing Dynamics into Static Causal Models

Naftali Weinberger

Recently developed graphical causal modeling techniques significantly downplay the role of time in causal inference. Time plays no role in the criteria specifying what it means for causal hypotheses to be observationally equivalent, and the probabilistic criteria used fail to distinguish among hypotheses that – given the assumption that causal variables precede effect variables – involve different time orderings among the variables. Additionally, the causal Markov condition – a central condition for choosing among causal hypotheses given a joint probability distribution – most straightforwardly applies to cases in which the variables are sampled from time-stationary distributions. Finally, it is commonplace to present models in which the variables are not explicitly indexed to times.

The lack of emphasis on time in causal models may suggest that the models are neutral with respect to the temporal relationships among the causal variables. Here I propose that, in fact, the lack of reference to time in many causal models is a legacy of the fact that causal models were initially designed to model the relationships in simultaneous equations models. Simultaneous equations represent the stable long-term relationships among the modeled variables rather than the shorter-term dynamic of systems in which variables have not reached stable values. Rather than being neutral with respect to the temporal relations among the variables, many apparently atemporal causal models in fact apply only under specific temporal assumptions about the timescale at which the system is being modeled.

After spelling out the interpretation of causal models as applying to systems that are treated as static, I consider ways that these models can be generalized to model the dynamics of systems that are away from equilibrium. First, I show how the dynamic causal models developed by Iwasaki and Simon (1994) generalize the causal ordering method from Simon (1977), which had been

designed for simultaneous equations. Simon's causal ordering serves as a basis for understanding the causal asymmetry of the structural equations that are still used to represent the dependence of effects on their causes. Second, I consider ways that the statistical methods used for causal modeling have primarily dealt with stationary distributions, and explain how certain non-stationary features of distributions correspond to the qualitative features in dynamic causal models. While I do not here provide a systematic treatment how to causally model time-series data, the present discussion contains both novel suggestions as to why such data present a challenge as well as suggestions for how dynamic causal models can be used as a bridge between contemporary graphical models and econometric time-series methods.

This chapter is organized as follows. Section 10.1 provides background regarding graphical causal models. Section 10.2 presents challenges to attributing times to the variables in these models, and explains how we can interpret the relationships among simultaneously measured variables causally. Section 10.3 discusses Iwasaki and Simon's dynamic causal models as a way of relaxing the presuppositions of simultaneous equations models. Section 10.4 draws connections between these dynamic causal models and methods employed in time-series analysis. Section 10.5 concludes.

10.1 Graphical Causal Models

The starting point of the present discussion is the graphical causal modeling methods developed by Pearl (2000) and Spirtes et al. (2000). While these works are relatively recent, their authors build on almost a century of contributions by psychologists, social scientists, economists, philosophers, and computer scientists. These methods provide a systematic and general account of how to use one's background causal assumptions to choose among competing causal hypotheses based on one's evidence, both in experimental and non-experimental contexts. For example, these methods enable one to determine which probability distributions over a set of variables are compatible with a particular causal hypothesis relating those variables. The methods also allow one to give a general account of when a causal quantity is *identifiable*, which means that it can be uniquely determined from the probability distribution, given one's causal assumptions.

In the graphical approach, causal models are typically represented using directed acyclic graphs, or DAGs. A DAG is a set of nodes connected by directed edges (i.e. arrows) such that one cannot get from a node back to itself via a set of connected edges all going in the same direction. When

DAGs are used for causal inference, the nodes are variables, the directed edges are direct causal relationships, and the lack of cycles corresponds to the assumption that no variable is either a direct or indirect cause of itself. The acyclicity requirement can sometimes be relaxed (Richardson, 1996; Park and Raskutti, 2016), although the general properties of cyclic graphs and the data-generating mechanisms they represent are less well understood. The term *causal model* typically refers to the combination of the graph and the joint probability distribution over the variables. As we will see, given a graph and distribution, it is sometimes possible to identify the functional relationship representing how an effect depends on its causes. As to what it means to interpret directed edges as direct causal relations, there are two general approaches. One is to treat the concept of direct cause as an undefined primitive whose content is determined holistically by the assumptions one uses to link DAGs to probability distributions. The other is to explicate X's being a direct cause of Y in terms of the possibility of changing Y via intervening on X while holding other causes of Y fixed.

I'll now provide further background regarding the use of DAGs for causal inference and identifiability, beginning with the latter. Let's begin by considering the simplest case in which a causal effect is *not* identifiable. Suppose that we were told that X causes Y, but these variables share an unknown common cause. The effect of X on Y would not be identifiable, since no matter how much we learned about the probabilistic relationship between X and Y, we could not know to what extent an inferred probabilistic dependency should be attributed to the causal relationship rather than to the common cause. In other words, one cannot uniquely determine the magnitude of the effect of X on Y from the DAG and probability distribution. In contrast, if we knew that X and Y shared a common cause Z, and that there were no other common causes of X and Y (other that causes of Z), then the effect of X on Y given particular values of Z could be identified by conditioning on different values of Z. Given the graph and the probability distribution, the effect would be identified by the expression $Pr(Y|X,Z)$.

A $path$ is a set of connected arrows. Where X and Y share an unmeasured common cause, the reason the effect is not identified is that in addition to the single-arrow path from X to Y there is also an additional path between X and Y via the common cause that 'transmits' probabilistic influence. Given an account of the conditions under which probabilistic influence is transmitted, we could provide conditions under which the effect of X on Y can be identified. We have already seen that paths containing a common cause transmit probabilistic influence unless the common cause is conditioned upon. In contrast,

causally independent causes of a common effect are uncorrelated *unless* one conditions on their common effect. So paths containing a common effect will not transmit probabilistic influence unless the effects are conditioned upon.

We can generalize these observations using the notion of *d-separation*:

Definition 10.1 d-Separation: A path is d-separated by variable set $\boldsymbol{Z}$ just in case:

(a) The path contains a triple i→m→j or i→m←j such that m is in $\boldsymbol{Z}$, or
(b) the path contains a collider i→m←j such that m is not in $\boldsymbol{Z}$ and no descendant of m is in $\boldsymbol{Z}$. [m is a *descendant* of $\boldsymbol{Z}$ if there is a path from m to $\boldsymbol{Z}$ consisting of directed arrows all going in the same direction.] (Pearl, 2000)

d-separation is a property of paths. Two variables are d-separated by $\boldsymbol{Z}$ if and only if they are d-separated by $\boldsymbol{Z}$ along all paths. Two variables are d-connected by $\boldsymbol{Z}$ if and only if they are not d-separated. To see how d-separation helps with identifiability, note the following sufficient condition. The effect of X on Y is identifiable given variable set $\boldsymbol{Z}$ if, in the DAG, X is d-separated from Y by $\boldsymbol{Z}$ along any "back-door" path that contains an arrow going into X. While this is just a sufficient condition, there exists a sound and complete procedure for determining whether a causal quantity is identifiable from a graph and probability distribution.

Identifiability concerns what one can measure *given* causal knowledge and the probability distribution. But how do we get such knowledge in the first place? Here d-separation helps as well. We can determine which DAGs are compatible with a probability distribution using a criterion called the *causal Markov condition*, which states that d-separated variables will be probabilistically independent. More precisely:

Definition 10.2 Causal Markov Condition (CMC): given a graph G and a probability distribution P over V, for any sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ in V, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated conditional on $\boldsymbol{Z}$ in G, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent conditional on $\boldsymbol{Z}$ in P.

CMC will not hold generally if one considers a variable set V that omits common causes of variables in V. For ease of explication, in what follows I'll limit myself to considering *causally sufficient* variable sets that include any common cause of variables in the set. Causal sufficiency will not hold in general, and can be relaxed in many contexts. A further constraint is that the variables in V must be logically, conceptually, and metaphysically independent of each other. To see why, imagine two variables X and Y such that $X = A\&B$

and $Y = B$. Given this logical relationship, the probability of Y given X will be 1 even if the unconditional probability of Y is not 1. But this correlation reflects the logical relationship among the variables and does not by itself call for a causal explanation.

CMC specifies what it means for a DAG to be compatible with a probability distribution. The condition places only a weak constraint on the set of models compatible with a distribution and supplemental conditions are needed to choose amongst the hypotheses compatible with a given distribution. To see how weak the condition is, note that any complete graph in which all the variables are directly connected is compatible with any distribution (since no variables are d-separated conditional on any others). Nevertheless, to draw causal inferences from probabilistic knowledge *some* principle is required, and CMC's logical weakness gives it the virtue of being compatible with a wide range of accounts of causation. Moreover, given suitable restrictions on the variables considered (see above, as well as (Hitchcock, 2016, section 3.2)), it is a plausible candidate for a universally true generalization. The most serious potential counterexamples arise in quantum mechanics, and the interpretation of these cases remains controversial. I discuss some further alleged counterexamples in Section 10.4.

Using CMC combined with additional parsimony principles (Zhang, 2012; Forster et al., 2017) one can choose among causal DAGs based on a probability distribution. Yet there is a limit on how much one can infer from the distribution using these principles. Specifically, all DAGs that entail all and only the same conditional independencies according to CMC are considered to be observationally equivalent. For instance, one cannot distinguish between the three DAGs in Figure 10.1. As a result, additional background knowledge is required to distinguish among the DAGs, and a common plausible suggestion is that time will often play this role. Nevertheless, the fact that time is only invoked to choose among already specified models makes it clear that, in theory if not in practice, time plays a secondary role in graphical causal models. Notice, for instance, that if we were to know the time ordering of the variables X, Y, and Z in Figure 10.1, and to assume that causes preceded their effects, then the three DAGs would *not* be observationally equivalent.

The basis for using DAGs to establish identifiability is closely related to the metaphysical assumption underlying CMC. For example, blocking all back-door paths allows for identifiability because doing so blocks any way that the value of a cause is informationally relevant to the value of the effect variable, except via causal paths involving connected sets of unidirectional arrows from the cause to the effect. But this would not establish identifiability unless one believed that there must be some back-door path linking the cause to the effect

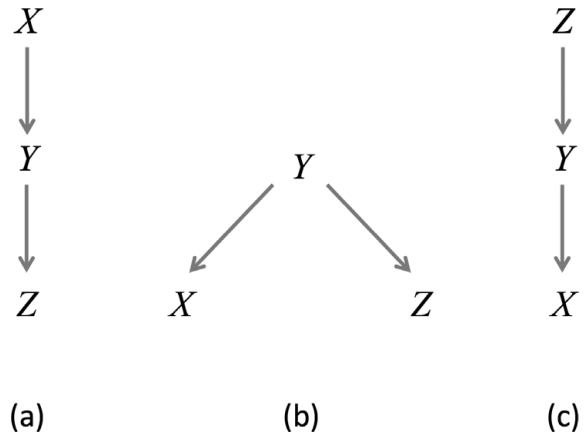

Figure 10.1 Three 'observationally equivalent' DAGs.

in order for one to provide information about the other that is not due to the causal path. In other words, one must accept CMC.

The concept of identifiability is crucial for understanding the relationship between causation and probabilities. The DAG framework does reduce causation to probabilities, but, given suitable causal assumptions, one can interpret certain conditional dependence statements as providing an unbiased measurement of the dependence of an effect on its causes. In such statements, the effect is the probabilistic consequent and its causes the antecedents in the probabilistic expression. When CMC holds, the joint probability distribution can be decomposed into a set of conditional independencies in which each variable is given as conditional only on its direct causes in the model. A useful way for thinking about the relationship between conditional dependence statements that can and cannot be interpreted causally is that only the former are invariant to interventions (Korb et al., 2004; Pearl, 2000). An intervention on a causal variable influences it in such away that any influence the intervention has on the intervened upon variable's effect is only via the intervened upon variable. The concept of an intervention allows for a systematic distinction between causal and non-causal probabilistic expressions, and has the added advantage of linking causal inference to experimental methodology, which also relies on interventions. Yet physically intervening is not necessary for gaining causal

knowledge. What matters for whether a probabilistic quantity may be causally interpreted is whether it is identifiable, with or without intervening.[1]

One of the main purposes of causal models is to aid causal inference. CMC is evidently not sufficient for inference, since further parsimony principles are needed to choose among the large set of Markovian models. Additionally, most variable sets do not obey CMC (due to omitted common causes, for instance). When it comes to the best methods for inferring causal models from different types of data, a healthy dose of pragmatism is warranted. One can test an inference method for reliability by running a simulation and seeing if the method recovers the data generating process. Such a method can be reliable even if it does not always produce models obeying CMC. Nevertheless, the assumption that, given suitable restrictions on acceptable variable sets, the true DAG over such variables will either satisfy CMC or can be embedded in a broader variable set that satisfies CMC is indispensible for causal inference. The reason is the close connection between CMC and identifiability. We need identifiability to distinguish between probabilistic expressions that do and do not provide unbiased effect measurements. Without some such notion, there is no theoretical link between causal hypotheses and probabilistic information. Such a theoretical link is necessary even for running simulations. To talk about causal models as a data-generating process, one needs to specify the rules by which causal models generate the data. This is so even if one adds distortions to the data-generating process such that the inferred distribution will not itself obey CMC.

For CMC to play a theoretical role in providing identifiability conditions, it is a virtue that it provides only weak constraints on which models it deems acceptable. Note that the concept of identification considered here is non-parametric. That is, the claim that effects of X and Y on Z are identified from the probability distribution $\Pr(Z|X,Y)$ does not indicate which probabilistic expression or structural equation is appropriate for quantifying the effect of X and Y on Z. Whether, for example, the structural equation giving Z as a function of X and Y involves an interaction term (e.g. δXY) is not determined by the DAG, but the identifiability of the $Pr(Z|X,Y)$ ensures that in principle whether X and Y interact could be established from knowledge of the true probability distribution. The fact that the form of identifiability that CMC underlies is silent with respect to, e.g., which correlation coefficient is best for quantifying causal strength, provides a reason for thinking that the notion

[1] If we focus on 'hard' interventions that fully determine the value of the intervened upon variable, the effect of X on Y is identifiable by intervening on X if $\Pr(Y|X)$ is identifiable in the DAG derived from breaking all arrows going into X.

of probabilistic independence invoked by CMC should also be given in a way that is independent of any way of measuring probabilistic independence. As we will see in Section 10.4, there are alleged counterexamples to CMC with variables that are by hypothesis not causally related, but which appear to be probabilistically dependent if one uses certain correlation coefficients. There I will side with those who claim that these correlations are insufficient for establishing probabilistic dependence, and that whether two variables are genuinely dependent for the purpose of applying CMC is something that is to be determined by statisticians using whichever measure of dependence is appropriate for the data.

This brief discussion of graphical causal models is sufficient for seeing how they represent causal relationships without explicit reference to time. The asymmetric relationships in the models are not explicated in terms of causes preceding their effects in time, but rather in terms of the different probabilistic consequences of the directions of the arrows in a DAG. In Section 10.3 we will see that the project of explicating causal directionality in fact precedes probabilistic approaches, and in Section 10.4 we will consider some subtle ways in which probabilistic assumptions relate to temporal ones. But before we get there, we need to ask an even more basic question. How do we make sense of causal models that do not specify the temporal relations among their variables?

10.2 Temporal Relations in Seemingly Atemporal Causal Models

In the previous section we reviewed graphical causal models and saw how the basic causal concepts are defined without explicit reference to time. Yet on reflection, it is somewhat puzzling as to how to understand the variables within the causal models without reference to time. In many causal models there is a single variable for each quantity. For instance, a model for an ideal gas system might have a variable V for volume, and not, for instance, a vector of variables $V_1, V_2, \ldots, V_n$ for the volumes at times 1 to n. In some cases, this does not lead to ambiguity. In the ideal gas case with a moveable piston, we know that equilibrium volume depends on pressure and temperature, and so we do not need to consider values away from equilibrium. But, more generally, simply to talk about the relationship between X and Y without saying when Y occurs (at least relative to X) leaves it ambiguous which relationship we are talking about. So we need to ask: which assumptions make it appropriate to compress potentially complex sets of causal relationships among time-series variables into simple representations involving single variables?

Some philosophers may take issue with my claim that the variables in a causal model in fact have a temporal ordering. Philosophers commonly treat variables as referring to properties, where the relationship between a variable and its values is akin to that between a determinable and its determinates (e.g. the relationship between 'being a shade of red' and 'being crimson'). On this way of talking, a variable can have different values at different times, since an object's properties can change over time. Yet this way of talking does not match statisticians' definition of a variable. Mathematically, a variable is a mutually exclusive and jointly exhaustive partition over a space of possibilities. Since it follows that a variable can have at most one value, to represent a quantity that changes over time we need multiple variables – typically, one corresponding to each observation. If we are not talking about a quantity that is changing in time, then we may define a variable without respect to time. But since we are here considering the role of time in causal models, it will help to eliminate ambiguity and to explicitly assign each variable a time.[2] Where we want to talk about a quantity changing over time, we will need a time-series of variables: $\{X_t, X_{t+1}, \ldots, X_{t+n}\}$.

It is natural to assume that once we make the time-indices of the variables in a causal model explicit, the temporal ordering should match the causal ordering. That is, causes must precede their effects. In the present volume, Livengood and Zwier's Chapter 9 seek to reconcile this plausible assumption with actual modeling practice. Here I propose an alternative explanation for why graphical causal models often do not include explicitly time-indexed variables. Specifically, all of the variables are indexed to the *same* time. This, I suggest, is a legacy of origin of certain aspects of the causal modeling framework in an econometrics tradition that uses so-called simultaneous equations. I will say a bit more about that tradition in the next section. Here I will simply assume that the variables in many causal models are simultaneous, and consider how we might interpret the relationships among such variables causally.

Talk of causal relations among simultaneous variables raises a puzzle, as causes are generally assumed to precede their effects. Some philosophers do allow for simultaneous causation, and equilibrium relationships – such as the effect of the current supply and demand of a good on its current (equilibrium) price – are plausible candidates for simultaneous relationships. Yet there are available causal interpretations of simultaneous equations, even assuming that causal relationships are genuinely diachronic. I consider two: the 'limit'

[2] See (Hitchcock, 2012) for a further defense of time-indexing variables.

interpretation and the 'steady-state' interpretation.[3] On the limit interpretation, the claim that X causes simultaneous variable Y should be interpreted as saying that X influences Y diachronically over arbitrarily small intervals. On the steady-state interpretation, which I adopt, simultaneous causal relationships among variables should not be understood as genuinely simultaneous, but rather as simultaneously observed. To interpret the relationship between X_t and Y_t causally one needs to assume that (a) X_t equals X_{t-1} and (b) Y is in a steady state at t. Condition (a) indicates that X_t itself is not the cause of Y_t but nevertheless enables one to infer the value of the genuine prior cause. Condition (b) indicates that Y at t has had sufficient time to reach a steady state in response to any changes in X. A variable is in steady state just in case its time-derivative is 0.

The steady-state interpretation of simultaneous equations is somewhat counterintuitive. While talk of simultaneity suggests that one is evaluating a system at a very short timescale, on the suggested interpretation simultaneous equations represent long-run relationships. To avoid confusion, it is important to distinguish between two inversely related rates that we might be referring to when talking about time in causal models. On the one hand, we might be talking about the timescale at which we evaluate a system, operationalized in terms of the rate at which the system is sampled. Here slower rates of sampling correspond to evaluating the system at a longer timescale. On the other hand, there is the rate at which a variable stabilizes in response to changes in its causes' values. One can think of simultaneous equations as evaluating the system at a sufficiently long timescale such that the rate at which the effects in the model respond to their causes is negligibly small at the sampling rate corresponding to that timescale.

Why interpret non-time-indexed causal models in the manner suggested here? Suppose that X causes Y, but it takes time for Y to stabilize. To make this a bit more concrete, imagine that if $Y = y_0$ at t and X is set to x at t, then Y will equal y_1 after 1 second and y_2 after 2 seconds, and will then remain at y_2 in the absence of further interventions. If we modeled this as $X \rightarrow Y$ it would be unclear at which time Y was being measured and we could not give a quantitative structural equation. There are two options: either X and Y are not represented as simultaneous or they are. If not, then the modeler is being sloppy, and the sloppiness can only be eliminated by specifying how much time has passed between X and Y. If X and Y are simultaneous, then we need to choose between the limit and the steady-state interpretation. On the limit interpretation the effect of X on Y corresponds to the instantaneous rate of

[3] This discussion is inspired by (Strotz and Wold, 1960, section 3).

change of Y given X at t: $\frac{dy}{dx}$. This fails to capture the longer-term effect of X on Y, and especially that Y is constant after t=2. In contrast, the steady-state interpretation clarifies that the variables are measured simultaneously, but does not require one to know precisely how long it takes Y to fully respond to the intervention on X. It indicates simply that the system is being evaluated at a timescale at which Y has had enough time to respond. So assuming that causal relationships take time and the modeler is not sloppy, treating the variables as simultaneous is the natural way to go.

The proposal that the variables in non-time-indexed DAGs are simultaneous illuminates some otherwise puzzling features of contemporary causal modeling. It would explain, for instance, why so much of causal inference is done on cross-sectional rather than longitudinal data, and also why social scientists working with time-series data often reach for alternative causal frameworks such as those relying on Granger causality (though see (Eichler, 2012) for a discussion of the use of DAGs for time-series analysis). This suggests that existing models have limitations regarding how to model time, though it is surprisingly unclear what those limitations are. The strategy of the present chapter is to begin with the working hypothesis that the limitation of causal models is due to their original application to simultaneous equations with the steady-state interpretation, and then to consider how they can be generalized to systems that are away from steady state. The following section presents Iwasaki and Simon's dynamic causal models, which provide such a generalization. These are, to be clear, only one type of dynamic causal model, and I here make no attempt to provide a survey of the options. Nevertheless, the following discussion will reveal them to be especially useful for thinking about the temporal assumptions of causal models.

10.3 Reintroducing Dynamics into Causal Ordering

The dynamic causal models in Iwasaki and Simon (1994) are a generalization of Simon's causal ordering method (Simon, 1977). While Simon does not make this explicit in the earlier paper, Iwasaki and Simon later write:

> [T]he theory's definition of causal relations is useful in generating a causal interpretation of the behavior of any system described by a system of simultaneous equations, including a physical device. Causal ordering was initially defined by Simon for a static system consisting of equilibrium equations. (1994, p. 145)

The progression in this section is from causal models with the steady-state interpretation to models that can represent systems away from steady state. I will begin by presenting Simon's original causal ordering method and explaining its relationship to contemporary modeling methods. The fact that his account can still be used to explain the asymmetry of the structural equations in causal models supports the working hypothesis that seemingly atemporal causal models contain simultaneous variables.

Simon asks what makes causation an asymmetric relationship, and he rejects the answer that it is temporal ordering. His strategy is to show how one can begin with a set of 'symmetric' equations in which one can freely move terms across the equals sign via applying the same operation to both sides, and to provide conditions under which such equations can be reorganized so that each variable is given as an effect of its causes. Systems that can be interpreted causally are those in which the values of certain variables can be solved for prior to others, and the causal ordering – that is, the partial ordering of the variables such that causes always precede their effects – is determined by the order in which the variables must be solved for.

As a simple example, consider an ideal gas system with the variables temperature (T), pressure (P), and volume (V). If the relevant system is a sealed container, we begin with the following symmetrical equations:

$$0 = f(P, V, T) \tag{10.1}$$

$$0 = f(T) \tag{10.2}$$

$$0 = f(V) \tag{10.3}$$

Equation (10.1) indicates that there is a function relating temperature, pressure, and volume, i.e. the ideal gas law. Equations (10.2) and (10.3) indicate that the values of temperature and volume are *exogenous* – that is, they do not depend on those of other variables in the system. Given these equations, one can solve for T and V first, and only then can one solve for P in equation (10.1). Accordingly, the equations can be rewritten as follows:

$$P = f(V, T) \qquad (i)$$

$$T = f(t) \qquad (ii)$$

$$V = f(v) \qquad (iii)$$

Equation (i) indicates that pressure is an effect of volume and temperature. Equations (ii) and (iii) indicate that T and V are assigned constant values. After reorganizing the equations, one can treat them as structural equations in which the variables on the left-hand side are effects of those on the right-hand side, where each variable appears on the left-hand side in a single equation.

Of course, not every set of symmetric equations can be causally ordered in this way. If one could not divide up the equations so that certain self-contained subsets could be solved first, no causal ordering could be identified given the equations.

In one sense, the causal ordering method does not provide causal information beyond that required for specifying the equations in the first place. We needed to specify which variables are lawfully related (in (10.1)) and also which variables were exogenous to find the causal ordering. If pressure, instead of volume, were exogenous, we would have to replace (10.3) with an equation (10.4):

$$0 = f(P) \tag{10.4}$$

These equations entail that T and P are causes of V. This causal ordering represents the relationships in a movable piston system, rather than a fixed-volume container. Like recent methods, the causal ordering method does not yield causal knowledge without causal assumptions. Nevertheless, the method reveals how causal asymmetries follow from symmetric laws and assumptions about exogeneity.

Although Simon's method does not rely on probabilities, the structural equations identified using the methods are the same as those identified using graphical causal models. To see how this could be, it helps to know that any variable set linked by equations with a deterministic component plus an independent error term for each variable automatically satisfies CMC. Accordingly, exogenous variables can be operationalized as those whose values depend on an independent error term alone. Nevertheless, we may be able to make judgments of independence even in cases where we cannot make probability attributions. In the ideal gas case, the judgment that temperature is exogenous is due to the ability of an experimenter to change the temperature at will by adjusting the heat bath in which the container is immersed, and the judgment that volume is exogenous reflects the assumption that the volume is either fixed by the experimental setup or can be adjusted by the experimenter. While modeling external sources of variation in a system as random variables enables one to infer causal relationships from probabilistic information, the type of independence that underlies causal attributions need not be identified with probabilistic independence.

I now consider Iwasaki and Simon's dynamic causal models, which generalize Simon's causal ordering method to systems in which at least some of the variables have not yet reached steady state. I will especially emphasize the models' use of time-derivatives to expand the representational power of the framework. This will help clarify the models' temporal assumptions, and will

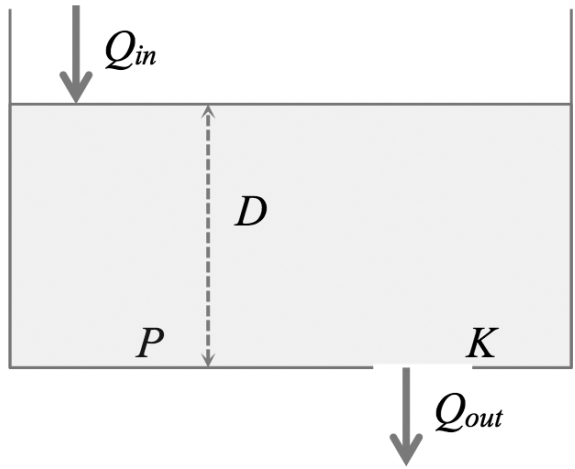

Figure 10.2 Iwasaki and Simon's bathtub example: Q_{in}: *rate of flow in*, Q_{out}: *rate of flow out*, D: *depth*, P: *pressure*, K: *drain size*.

further allow us to later compare Iwasaki and Simon's use of derivatives to the somewhat similar use of derivatives in time-series econometrics models.

Let's begin with an example (Figure 10.2). Consider a bathtub in which water flows in at a rate of Q_{in} and out at a rate of Q_{out}. The difference between these rates determines the rate of change of depth, which in turn influences depth, D. The pressure P at the bottom of the tub is proportional to depth. The rate of flow out depends on pressure and the size of the drain, K, which is treated as exogenous. The system is at steady state when Q_{in} equals Q_{out} and D no longer changes.

The symmetric equations for the steady state of this system are as follows.

$$Q_{out} = c_5 k P \tag{10.5}$$

$$D = c_6 P \tag{10.6}$$

$$Q_{in} = c_7 \tag{10.7}$$

$$K = c_8 \tag{10.8}$$

$$Q_{in} = Q_{out} \tag{10.9}$$

Equations (10.5) and (10.6) indicate that the rate of flow out is proportional to the size of the drain and pressure, and that depth is proportional to pressure. Equations (10.7) and (10.8) say that the rate of flow in and the size of the drain are exogenous. Equation (10.9) indicates that at equilibrium the rate of flow in equals the rate of flow out. This is what indicates that we are considering the

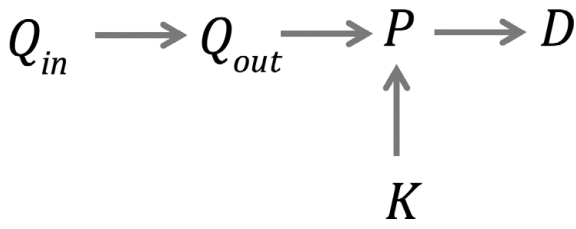

Figure 10.3 Equilibrium causal ordering for bathtub.

model at equilibrium. The resulting causal ordering is given by the graph in Figure 10.3. Since Q_{in} is exogenous, one can solve for Q_{out} via (10.9). From (10.5) we see that given K and Q_{out} we can solve for P, and then from (10.6) we derive D.

The causal model in Figure 10.3 is far from intuitive. To make sense of it, it is crucial to bear in mind that we are assuming that all variables have reached stable values in response to any perturbations of their causes. While changing the rate of flow in would have a short-term effect on depth and pressure, assuming that the system reaches steady state, the long-term rate of flow out will only depend on the new rate of flow in. Similarly, changing the size of the drain would alter the steady-state pressure and depth, but would not influence long-term Q_{in} and Q_{out}. This is not to say that the model is wholly satisfying. It is especially difficult to understand the causal ordering of pressure and depth. Within the causal ordering method, the way to understand this is that assuming that the system is at steady state, and that the drain is size k, the steady-state pressure must have a certain value p in order to achieve the required rate of flow out, and in order for pressure to have this value the water must be a certain depth. But this seems to describe teleological rather than causal relationships. In fact, Iwasaki and Simon bring this example in large measure to motivate the desirability of more dynamic representations of a system, which yield a more intuitive ordering. Whether or not the model is intuitive, it bears mentioning that Dash (2003) uses simulations to confirm that it is the model we would infer when sampling at a longer timescale.

Since Q_{in} and Q_{out} are rates of change, we could have represented them using time-derivatives. But we need not do so. While the traditional causal ordering method applies to variables that have reached stable values, these variables can themselves be derivatives, in which case the stable values are constant rates of change. In extending the framework to represent the dynamics, the innovation is not the use of time-derivatives per se, but the use of both variables and their time-derivatives within a single model. Since

variables and their time-derivatives are conceptually related, they are not fully independent. But, as we will now see, including both variables and their time-derivatives in a model enables us to reintroduce dynamics into a static set of equations.

Naturally enough, if we no longer assume that the system is at steady state, we need to eliminate the assumption that Q_{in} equals Q_{out}. We can replace equation (10.9) with the following:

$$D' = Q_{in} - Q_{out} \tag{10.10}$$

Equation (10.10) states that the rate of change of depth is equal to the difference between the rate of flow in and rate of flow out. When adding an equation with an explicit derivative into our model, we also need to add two other equations. One states that integrating D' yields D:

$$D = Int(D') \tag{10.11}$$

The other specifies an 'initial condition' for D at time 0:

$$D_0 = d \tag{10.12}$$

These equations enable us to represent the passing of time in the model. Given a variable's current value and its time-derivative, integration yields a discrete approximation of the variable's value at subsequent moments. Equations such as (10.12) are crucial for the causal ordering method, since they indicate that at a particular time-step we need to treat the value of the non-equilibrated as exogenously given. Given such a value, the model will tell us how the variables will evolve over time.

The dynamic model derived from combining (10.5) – (10.8) with (10.10) – (10.12) is given in Figure 10.4. The water flowing in increases depth, which increases pressure, which, given a drain of size k influences the rate at which the water flows out. The rates at which the water flows in and out in turn determine the rate at which the depth changes and thereby influence the future depth.

The graph in Figure 10.4 provides a subtle way of integrating activities occurring at several timescales into a single model. Variables linked by causal arrows are interpreted as they were in simultaneous equations models. That is, P, K, and Q_{out} respond quickly to changes in their causes and are measured at steady state. If desired, one could include additional derivatives for all variables in the model, but this would not substantially change the causal relationships

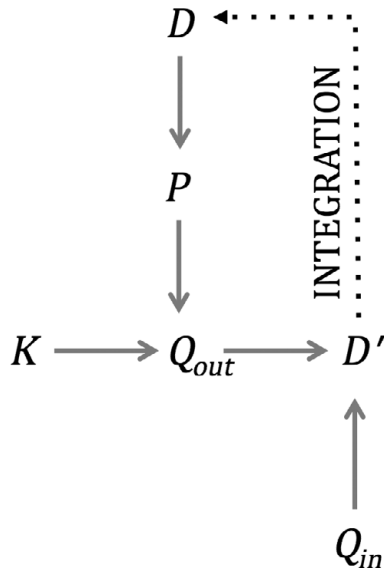

Figure 10.4 Dynamic causal ordering for bathtub.

among the variables.[4] However, the inclusion of D' along with the integration link indicates that we are sampling the system at a rate at which depth has not reached a stable value. Using this derivative, we represent how current depth influences its rate of change. It is via this form of self-regulation that the system gets moved towards equilibrium in the long run.

To get a feel for the dynamic graph, it helps to consider how we might represent the same causal relationships using a 'rolled-out' graph with time-indexed variables and no derivatives (Figure 10.5). While the rolled-out graph does not explicitly represent the way that the rate of change of D is a function of Q_{in} and Q_{out}, and that D' combines with D_0 to produce D_1, it otherwise preserves the causal ordering and differentiates between the synchronic and diachronic relationships. It helps to think of this graph in light of the CMC. Note, for example, that given D_1, P_1 is probabilistically independent of P_0. This reflects the sense in which P is 'memoryless' – to know its value at a time, we need not know its prior values. More generally, causal models for

[4] To illustrate, if we included P' in the model and added the required equations, the resulting graph would contain an arrow from D to P', an integration link from P' to P, and causal arrows from P to P' and from P to Q_{out}. This would indicate that it takes time for depth to influence pressure, but would not further change the relationships between D, P, and the other variables in the model.

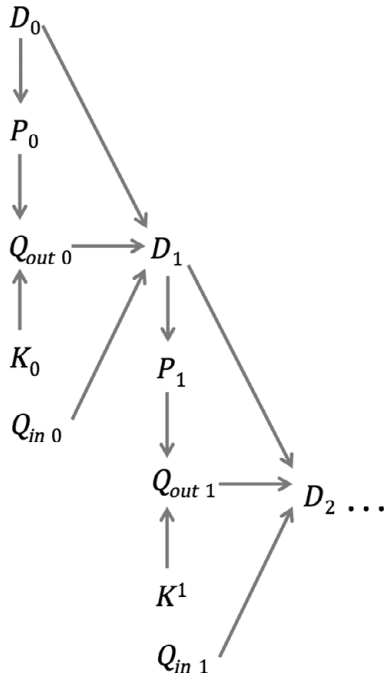

Figure 10.5 'Rolled-out' graph without time-derivatives.

static relationships are memoryless – which is why we needed to specify initial conditions only once we moved to the dynamic model. In contrast to pressure, depth is *not* memoryless when considering the system away from steady state. To know what the depth will be at the next moment you need to know its current value. As long as the system is away from equilibrium, the value of depth serves as a record of a past event – often the event that shocked the system out of equilibrium – and as an index for how far the system is from equilibrium.

The relationship between Figures 10.4 and 10.5 resembles that between causal models for simultaneous equations and time-indexed causal models more generally. The causal ordering method yields shorthand representations of the temporal development of a system without presupposing a characteristic timescale. In the static case, the graphs indicate that the variables are at steady state without needing to specify precisely how long it took them to stabilize. Similarly, causal models with derivatives convey information both about the relative rates at which variables influence one another and about the timescale at which the system is being evaluated, without providing specific

rates. From the inclusion of a variable along with its derivative, we can infer that the variable does not reach a stable state at the timescale at which the system is considered, and if there are other variables in the model without their derivatives, we know that they reach steady state at a comparatively faster rate.

Earlier I discussed the counterintuitive device of using simultaneous equations to model long-run relationships. The explanation I gave there similarly helps clarify how including diachronic causal relationships corresponds to evaluating the system at a shorter timescale. Specifically, by 'zooming in' and sampling the system at a faster rate, the slower relationships that appeared as simultaneous at the longer timescale now can be seen unfolding in time at the shorter timescale characterized by the faster sampling rate. While this is a lot to keep track of, the feature of the representation that is not counterintuitive is that synchronic relationships refer to faster causal interactions than those represented by diachronic relationships.

Iwasaki and Simon develop an 'equilibration' operator, which transforms a dynamic model into a model with the causal ordering one would get were one to wait for the equilibrated variable to reach steady state. To equilibrate D, delete equations (10.11) and (10.12) indicating the integration equation and the initial value of D, and replace D' with 0 in equation (10.10) ($D^{'} = Q_{in} - Q_{out}$). Then use the causal ordering method to derive the equilibrium mapping. Since replacing D' with 0 in (10.10) yields that $Q_{in} = Q_{out}$, we end up with the same equilibrium equations we started with and therefore get the equilibrium causal ordering. More generally, to equilibrate X (Dash, 2003):

1 Set all derivatives of X in the model to 0 and remove them from the model.
2 Delete all equations going into X or its derivatives.
3 Remap to get the new causal ordering.

The second step deletes both the integration equation and that stating X's initial conditions. Note that equilibration is always relative to a variable or set of variables. In models with multiple dynamic variables, certain variables sometimes must be equilibrated before others (Iwasaki and Simon, 1994, p. 166). Note also that equilibration does not always preserve the causal ordering among the variables, as can be seen from the reversal of the arrows linking D, P, and Q_{out} between the equilibrium and dynamic models.

The link between Iwasaki and Simon's dynamic causal models and the econometric time-series we will consider in the following section is that both use simultaneous equations to represent long-run equilibrium relationships and diachronic equations to represent non-equilibrium behavior, and both rely on derivatives (or difference equations) to do so. The econometric tradition is, in

general, much less systematic in identifying the causal assumptions of their models than the causal modeling tradition, but also more focused on issues of measurement, especially when it comes to the analysis of time-series data. So, the similarities just mentioned bode well for a better unification of causal theory with measurement methods.

10.4 Dynamic Causal Models and Econometric Time-Series Methods

While I suspect many philosophers and social scientists will be surprised by my claim that non-time-indexed models should be taken as representing steady-state relationships, there is one context in which philosophers have noticed that causal models are easier to apply in static scenarios. Specifically, for non-stationary time-series, more sophisticated statistical tools are needed to determine whether variables are probabilistically dependent, and thus, by CMC, d-connected in the true causal graph. This issue arises in discussions of Elliott Sober's proposed counterexample to CMC (Sober, 2001). The example involves the relationship between the sea level in Venice and British bread prices. These quantities are not causally related, but since both increase over time, the value of one is informative about that of the other. So, Sober claims, we have a case of a correlation without a causal explanation, violating CMC. I have deferred discussion of the counterexample until now, since it introduces time-series concepts that will help us understand the econometric methods that are the subject of the present section.

In my view, the correct response to Sober is that bread prices and sea levels are not in fact probabilistically dependent, and thus there is no dependency lacking a causal explanation. Hoover (2003) presents a helpful discussion of why the counterexample fails. In defining CMC in Section 10.1, we presumed knowledge of the true probability distribution over the modeled variables, and in particular knowledge of which variables are independent. But statistical independence is not directly observed; it must be inferred from a finite sample. Statistical inference requires a model of the distribution from which the data is sampled. One common assumption is that the random variables in a sequence are independent and identically distributed. If one uses the wrong probability model for statistical inference, one will make incorrect inferences regarding which variables are probabilistically independent, and thus regarding which causal models are ruled out by CMC.

In determining which probability model is apt for a set of observed random variables, it is crucial to consider whether the underlying distribution is

stationary or non-stationary with respect to time. Informally, a sequence is non-stationary if its variables or the relationships between them exhibit a systematic dependence on time. Sober's sequences are manifestly non-stationary, since the mean values of the variables in the sequences increase with time. If one models the two time-series for bread prices and sea levels using a correlation coefficient designed for stationary time-series, one will get a positive value, but this will not reflect a genuine probabilistic dependency. As Hoover notes, econometricians would model the time-series in Sober's example as being $I(1)$, indicating that although each time-series is non-stationary, the time-series derived from taking the first difference between time-steps – that is, subtracting each variable from its predecessor – is itself stationary. (Taking the difference in this manner corrects for autocorrelation.) The process of taking the difference between time-steps is called $integration$, the n number of times one must take the difference to achieve stationarity is called the *order of integration* $I(n)$. For two $I(1)$ series such as that in Sober's example, the test of whether they are probabilistically dependent is whether there is a function of the two differenced time-series that is itself stationary. This roughly corresponds to whether one time-series is informative about the other, even once one has taken the difference. In Sober's case, the answer is presumably no. Although the *level* of the sea is informative about the level of bread prices, the *changes* in one are not informative regarding the changes in the other.

While I view Hoover's response to Sober as successful, it is only a starting point for a broader discussion of causal inference from time-series. Defenders of Sober's counterexample (Reiss, 2007) have argued that Hoover's response severely limits the usefulness of CMC in inference by limiting the variable sets to which the condition applies and making causal inference overly dependent on the prior work of statisticians who process the data. Yet, as I argued above, while CMC can be relaxed and modified to deal with variable sets not satisfying CMC, its role in providing identification criteria specifying when a probabilistic relationship is causally interpretable is indispensible. For this purpose, what matters is that variables cannot be mutually informationally relevant in the absence of some causal connection (i.e. some d-connected path) and it is ok for the causal modeler to outsource the task of establishing informational relevance to the statistician. That said, CMC is only part of causal inference and we should not infer from Hoover's discussion that, as a general strategy, causal modelers can relegate the nuances of time-series inference to the statistician. As we have seen, causal models incorporate very specific temporal assumptions, and these assumptions are closely related to those studied by time-series econometricians.

Let's think further about stationary and non-stationary time-series. A simple example of a time-series gives each variable in the series as a function of its prior value plus an error term:

$$x_t = \rho x_{t-1} + \epsilon_t \tag{10.13}$$

When $|\rho| < 1$, the time-series given in (10.13) is stationary, and when $|\rho| = 1$ the time-series is non-stationary. This derives from the way that the shocks represented by the error terms accumulate. When $|\rho| < 1$, shocks from the past will continue to have an influence into the future, but the longer ago a shock is, the less influence it will have now. This is because although a shock now is incorporated into the present value of X, at each subsequent time-step the value of X gets multiplied by ρ. So the contribution of a shock now to the trend n time-steps later will be a function of ρ^n, which tends towards 0 as n increases as long as $|\rho| < 1$. In contrast, when $|\rho| = 1$ the force of each shock is never diminished. In stationary series, one can abstract away from the role of time in the sense that the relationship between the variables will only depend on their relative times, not absolute time. More specifically, in a stationary time-series, the mean and variance (and higher-order moments) are independent of time, and the covariation of variables at different times depends only on how far apart they are.

We see that even with stationary time-series a variable's current value reflects prior shocks, though as time passes the influence of prior shocks diminishes. Accordingly, ρ can be understood as indicating how much 'memory' a variable has of earlier events in the series. As we saw in our discussion of Iwasaki and Simon, considerations of how much memory a system has are not incidental to the content of causal models. While a system that has reached equilibrium retains no memory of the prior shocks that perturbed it from equilibrium, and can be represented using simultaneous equations, when a system is away from equilibrium a variable's value does contain information regarding prior shocks, and this must be incorporated into the model via relationships across time-steps. Whether one treats a stationary time-series as having memory depends both on the timescale at which one considers the system (i.e. the length of the lags) and a judgment regarding how small the influence of a prior shock needs to be before it counts as negligible enough to be ignored.

These considerations suggest that extending causal models from the simultaneous case to the dynamic one will require the incorporation of more sophisticated time-series methods into causal modeling. Even if causal modelers continue outsourcing the data analysis to statisticians, they cannot ignore that

models for different timescales correspond to different time-series with distinct statistical properties.

Of course, the very simple time-series represented by (10.13) does not involve causal relationships. Let's now turn to a causal example. The following example comes from Hoover's illuminating discussion (Hoover, 2015) of when we might interpret the relationships in econometric time-series models causally. The example concerns a drunk who, after leaving the bar, is followed by a concerned friend. The friend following the drunk keeps a safe distance in order to avoid detection, but will speed up to get closer to the drunk whenever the distance between them gets too large.

The drunk's trajectory is a random walk. That is, at each time-step the drunk is equally likely to move in any direction. The equation giving the movement of the drunk is that in (10.13) with $\rho = 1$, meaning that the time-series for the drunk's trajectory will be non-stationary. The friend's movement is also represented by a non-stationary time-series. Nevertheless, the time-series for the friend and the drunk are first-order cointegrated. This cointegration relationship indicates that in the long term, the friend and the drunk will never drift more than a certain amount apart. Since the friend is following the drunk, it is clear that the position of the drunk causes that of the friend. But this relationship will only be evident when considering the system at the right timescale. If one observes the system at too short a timescale, one will miss the fact that the friend and the drunk never drift too far apart. If one observes it at too long a timescale, one will discover the long-term cointegration relationship, but miss that it is the drunk influencing the friend and not vice versa.

The causal ordering framework does a nice job representing this case (Figure 10.6). X_d is the position of the drunk and X_f is that of the friend. In the dynamic graph on the left, the distance between the drunk and the

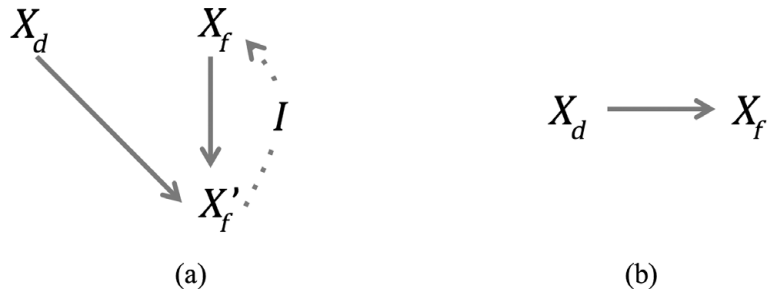

Figure 10.6 Dynamic graph for cointegration example X_d: *position of drunk*, X_f: *position of friend*.

friend determines the position of the friend at the next time-step. The use of a derivative/integration link pair indicates that the friend reduces the distance only after some time has passed. (Note that the co*integration* relationship between the time-series is represented using an integration link in the causal model.) Applying the equilibration operation to X_f yields the simpler graph on the right, which indicates that the position of the drunk is causing that of the friend.

The present example could be represented in econometrics using an error correction model (Engle and Granger, 1987; Wooldridge, 2015). In the background, we assume that the time-series for the friend and drunk are I(1), and that there is a long-run relationship between them as follows:

$$x_{f(t)} = \alpha + \beta x_{d(t)} \tag{10.14}$$

We represent the equation for the dependence of the position of the friend on that of the drunk as follows:

$$\Delta x_{f(t)} = \delta \Delta x_{d(t)} - \gamma (x_{f(t-1)} - \alpha - \beta x_{d(t-1)}) + u_t \tag{10.15}$$

Equation (10.15) takes the first difference for both the friend and the drunk to render the non-stationary time-series stationary. δ indicates the short-term effect of the change in the position of the drunk on the change of position of the friend. In our example δ will be 0, as the friend does not respond immediately to the drunk's movements. The term in parentheses represents the longer-term cointegrating relationship. Given the long-run equilibrium relationship from (10.14), any difference between the first term and the difference between the second and third term indicates that the system is away from equilibrium. The coefficient γ determines how quickly the system will return to equilibrium.

It is straightforward to see how Equation (10.15) with $\delta = 0$ neatly corresponds to the dynamic graph in Figure 10.6(a), and how equation (10.14) corresponds to the static model in Figure 10.6(b). The example chosen was exceedingly simple, so it remains to be seen how generally causal models can be linked to econometric time-series models. Nevertheless, by linking the dynamic causal model to a well-understood econometric model, we see how the mathematical operations that econometricians use to correct for autocorrelation map on to the relationships in dynamic causal models.

I conclude this section with a reflection of the relationship between causal and statistical inference when it comes to more complex time-series. I've argued that CMC is a general rule of causal inference and that the question of which variables are probabilistically independent should be left to the statisticians. Relatedly, if variables are probabilistically dependent, CMC does not differentiate among the inferences one can draw from different types of

dependencies. But this just means that we need principles in addition to CMC. This is generally acknowledged even in the stationary case, where CMC must be supplemented by additional principles to choose among Markovian models. When we consider more complex time-series, we need to make additional distinctions among ways that variables can be probabilistically dependent, and whether we treat certain vectors of variables as dependent or independent will sometimes depend on the timescale at which we consider the system. This is the lesson of Iwasaki and Simon's dynamic causal models, and the present section provides preliminary confirmation that inferring these models from time-series data will require an appeal to time-series techniques from econometrics.

10.5 Conclusion

While graphical causal models have revolutionized causal inference and analysis, it remains surprisingly difficult to specify their domain of application. There is work to be done in determining how graphical methods can be used for inferring causal hypotheses based on time-series data. Yet in order to figure out how the framework can be expanded, we need to first get clear on its current limitations. Here I have suggested that far from being neutral regarding the temporal relations among causal variables, existing models often implicitly come with strong assumptions about the variables' temporal stability. This diagnosis indicates that generalizing the framework is a matter of expanding it to represent cases in which at least some variables are away from their steady states. I have shown how Iwasaki and Simon's dynamic causal models provide such a generalization, and further illustrated the relationship between the generalized representation and concepts from econometric time-series methods.