Causal Contributions in Economics

Christopher Clarke*

Abstract

This chapter explores the idea of one variable making a causal contribution to another variable, and how this idea applies to economics. It also explores the related concept of what-if questions in economics. In particular, it contrasts the modular theory of causal contributions and what-if questions (advocated by interventionists) with the ceteris paribus theory (advocated by Jim Heckman and others). It notes a problem with the modular theory raised by Nancy Cartwright. And it notes how, according to the ceteris paribus theory, causal contributions and what-if questions are often indeterminate in economics.

1 Causal Contributions and What Ifs

It's not uncommon for economists to say that one variable made a causal contribution to another variable: the low price of lumber made a positive causal contribution to the high demand for lumber; large class size made a negative causal contribution to a child's educational attainment. Similarly, it's not uncommon for economists to answer "what if things had been different?" questions about hypothetical scenarios: what would the demand for lumber have been, if instead the price of lumber had been high? what would this child's educational attainment have been, if instead the size of her class had been low? This chapter is about these two concepts (causal contributions and what–if questions) and how they apply to economics.

The chapter will examine two theories of causal contributions: the modular theory (advocated by Judea Pearl and by interventionists such as Jim Woodward)

*This is the accepted version of a chapter that has been accepted for publication by Routledge in the forthcoming book *The Routledge Handbook of the Philosophy of Economics* edited by Julian Reiss and Conrad Heilmann due for publication in 2021. This work has received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme, under grant agreement no 715530.

and the ceteris paribus theory (defended by Jim Heckman and others). Since both these theories posit a tight connection between causal contributions and the answers to what-if questions, these theories also serve as theories of what-if questions as well. These two theories contrast with Nancy Cartwright's approach to causal contributions, according to which (a) no explicit theory of causal contributions is possible, and (b) there is a weaker connection between causal contributions and the answers to what-if questions (Cartwright 1989, 2007). Since Cartwright doesn't offer an explicit theory of causal contributions, I won't discuss her approach in this chapter. All I will say is that, the more problems one finds for the modular and ceteris paribus theories of causal contributions, the more attractive Cartwright's "no theory" approach to causal contributions becomes. I will also not discuss Hoover's (2001, 2011, 2013) theory of causation in economics, because Hoover's theory does not define causal contributions quantitatively, and because Hoover's theory does not give a recipe for answering what-if questions. Instead, Hoover's theory is qualitative: it's a theory of when one variable is a cause of another variable. Limitations of space prevent me from discussing Stephen LeRoy's theory (2016, n.d.), which I see as an important variant of Heckman's ceteris paribus theory. See also Julian Reiss (2012, 2009) for a discussion of what-if questions in the social sciences more broadly, and some alternative approaches to them not considered in this chapter.

This chapter will proceed as follows. Section 2 sets things up by distinguishing between direct causal contributions and overall causal contributions, and by defining the difference between an external variable and an internal variable. Section 3 lays out the modular theory of causal contributions. Section 4 notes some problems that the modular theory has when it is applied to economics. The most important problem is Cartwright's argument that modularity fails for complex social systems. Sections 5 and 6 develop my preferred version of the ceteris paribus theory.

I will illustrate these two theories by using two toy models, one of educational attainment and another of supply and demand. To keep an already complex discussion as simple as possible, I will not include any disturbance terms in these models. In other words, the models I am discussing do not have probability distributions attached to them. The discussion in this chapter can be extended to probabilistic econometric models, however, by treating these disturbance terms as additional external variables.

2 Key Concepts

To talk about causal contributions, economists find it useful to label some variables in their models as "external variables" and to label other variables as "internal variables". Some economists also find it useful to talk about the "direct causes" of a variable, and of the "direct causal contribution" that one variable makes to another. This section will illustrate these four concepts. To do so, I will present a model of educational attainment; I will then abstract away from this concrete example to give a precise definition of an external variable.

Direct Causal Contributions. The first equation that defines the model of educational attainment is $Y = \gamma_v V + \gamma_w W + \gamma_1 X_1$. This equation is to be interpreted as saying that, for any given child, there are three things that directly causally contributed to that child's educational attainment Y: the number of other children in that child's class V, the hours that child spends on extra-curricular activities W, and the educational policy X_1 enacted by the local / regional education authority responsible for that child. The coefficients γ_v, γ_w and γ_1 each denote an unknown constant—a positive or negative number that does not vary across the children in the population of children that we are studying. For example, γ_v is an unknown constant that describes the strength of the direct causal contribution that class size V made to educational attainment Y. Each extra member of a child's class directly contributed γ_v extra units to that child's educational attainment.

The second equation that defines the model is $V = \nu_1 X_1 + \nu_2 X_2$. This equation is to be interpreted as saying that, for any given child, there are two things that directly contributed to that child's class size V: regional education policy X_1, and the child's parental income X_2. Again ν_1 and ν_2 denote unknown constants that describe the strength of these direct causal contributions.

The final equation that defines the model is $W = \omega_2 X_2 + \omega_3 X_3$. This equation is to be interpreted as saying that, for any given child, there are two things that directly contributed to that child's extra-curricular activities W: the child's parental income X_2, and the child's attitude towards education X_3. Again ω_2 and ω_3 denote unknown constants that describe the strength of these direct causal contributions.

Direct Causes. I will call these equations *direct–causes equations* because they purport to describe direct causal contributions. One can depict what these equations say about direct causes by drawing a diagram. Specifically, whenever a variable makes a direct causal contribution to a second variable, one says that the first variable directly causes the second variable, and one draws an arrow starting at the first variable and ending at the second variable. See fig. 1.

Here class size V and extra-curricular activities W are each direct causes of attainment Y; parental income X_2 is an indirect cause of attainment Y, via both class size V and extra-curricular activities W as intermediaries; learning attitude

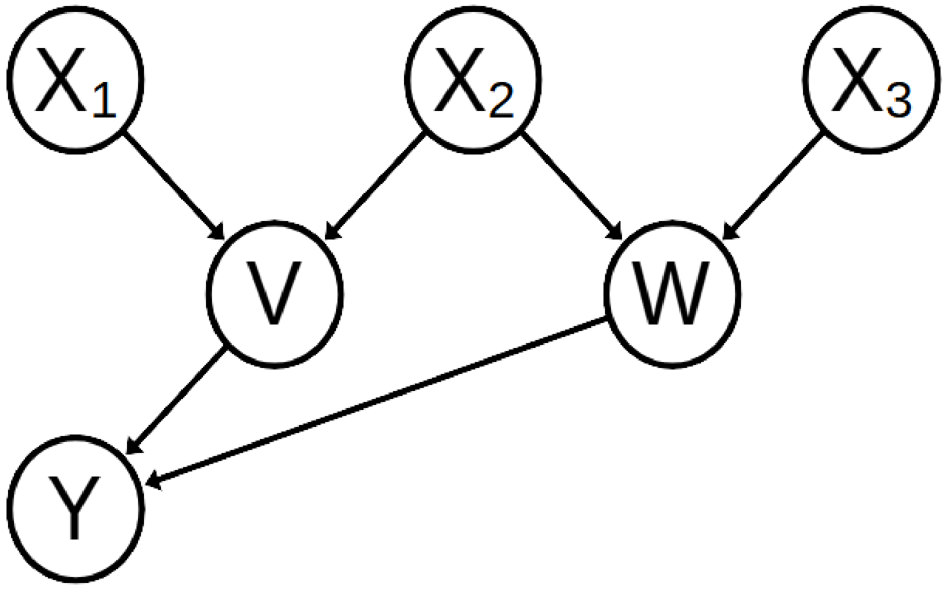

Figure 1: Direct Causes

X_3 is an indirect cause of attainment Y, via extra-curricular activities W as an intermediary; and education policy X_1 is a direct cause of attainment Y, and it also an indirect cause of attainment Y, via class size V as an intermediary. (Of course, the notion of a direct cause is relative to the variables that one includes in one's model. If, for example, one excluded V and W from our model, then X_2 and X_3 would become direct causes of Y.)

This notion of direct causation and of a direct causal contribution is meant to be intuitive, and many theorists would say that this notion cannot be reduced or defined in terms of any more fundamental notions. It's worth noting that one exception is Woodward (2003a) who provides the following definition of direct causation. Roughly. X_1 is a direct cause of Y if and only if: if X_1 were to take a different value (from the value it actually took), but all the other variables in the model (other than Y) were to take the values they actually took, then Y would take a different value (from the value Y actually took).

Hypothetical Differences in the Xs. The three equations that define the educational attainment model do not just describe the values that these variables $\{V, W, Y, X_1, X_2, X_3\}$ actually took in the population of children being studied. These three equations are to be interpreted as also describing the values, for any child in this population, that these variables $\{V, W, Y, X_1, X_2, X_3\}$ would have taken under any hypothetical scenario in which the X variables had differed (taking values different from the values that the X variables actually took). Thus they answer "what-if the Xs had differed?" questions. Take for example a hypothetical scenario in which education policy X_1 had taken value 5 instead, and parental income X_2 had taken value 8 instead, and learning attitude X_3 had taken value 2 instead. Under this hypothetical scenario the child's class size V would have been $5\nu_1 + 8\nu_2$, according to the equation $V = \nu_1 X_1 + \nu_2 X_2$, on this interpretation of the equation. In short, the equations that define the model not only make correct predictions for the scenario that actually occurred; they also make correct predictions about what would have occurred under hypothetical differences in the X variables in the model.

Note that one can substitute the first two equations that define the educational attainment model into the third equation to yield $Y = \gamma_v[\nu_1 X_1 + \nu_2 X_2] + \gamma_w[\omega_2 X_2 + \omega_3 X_3] + \gamma_1 X_1$. Tidying up gives us $Y = (\gamma_v \nu_1 + \gamma_1)X_1 + (\gamma_v \nu_2 + \gamma_w \omega_2)X_2 + \gamma_w \omega_3 X_3$. But, when a set of equations makes correct predictions under some hypothetical scenario, then any equation that is derived mathematically from those equations will also make correct predictions under that hypothetical scenario. Thus the following equations make correct predictions under hypothetical differences in the X variables in the model:

$$V = \nu_1 X_1 + \nu_2 X_2$$
$$W = \omega_2 X_2 + \omega_3 X_3$$
$$Y = (\gamma_v \nu_1 + \gamma_1)X_1 + (\gamma_v \nu_2 + \gamma_w \omega_2)X_2 + \gamma_w \omega_3 X_3.$$

External Variables versus Internal Variables. To divide the variables in one's economic model into external variables $\{X_1, X_2, X_3\}$ and internal variables $\{V, W, Y\}$ is to say something about how to interpret the equations that define one's economic model. It's to say:

> *External variables predict internal variables under hypothetical differences in the external variables*: for each internal variable in the model, one can derive (from the equations that define the model) an equation that expresses this internal variable purely as a function of one or more external variables. Any such equation derived from the model will make correct predictions (about the value this internal variable takes) under hypothetical differences in the external variables in the model.
>
> *External variables cause internal variables but not vice versa*: each of the external variables appearing in such an equation is a cause of the internal variable in question. But no internal variable is a cause of any external variable.
>
> *Variation freedom of external variables*: an external variable taking a given value doesn't preclude any other external variable from taking a given

value. More precisely: if x_1 denotes a possible value of external variable X_1, and if x_2 denotes a possible value of external variable X_2, and if x_3 denotes a possible value for external variable X_3, for example, then it is possible for $X_1 = x_1$ and $X_2 = x_2$and $X_3 = x_3$ to hold in any single case. By "possible" I mean both "consistent with the equations that define the model" and also something like "consistent with the system working as it normally does". In this respect, neither the model nor the system itself places strong restrictions on the value that an external variable can take, given the values of the other external variables.

The basic idea is that the model describes how the external variables causally determine the values of the internal variables, but it says nothing about the causes of the external variables themselves. Note that, as I've defined it here, the concept of an external / internal variable is a different concept from the concept of an exogenous / endogenous variable.[1]

3 The Modular Theory of Causal Contributions

With these concepts in hand, one can now describe the first theory of causal contributions and of what–if questions, which I call the "modular" theory. The modular theory is defended by Pearl (2009, 22–32, 70, 205–07) and Woodward (2003b). Woodward calls this theory the "interventionist" theory, but I find this metaphorical talk of "interventions" somewhat misleading so I will avoid it here. The modular theory is sometimes attributed to Haavelmo (1943), for example by Pearl (2009, 365). There is also a broad similarity between the modular theory and Simon's (1953) theory; for discussion see Cartwright (2007, 252).

The intuitive idea behind Modularity is that direct–causes equations are "modular": if one were to "intervene" in the system to "break" one of the direct–causes equations, such as $V = \nu_1 X_1 + \nu_2 X_2$ for example, this intervention would still leave all the other direct–causes equations intact. Thus if one were to intervene in the system to set V equal to 10 for example, this intervention would not change the value of any of the external X variables, nor would it change the other direct–causes equations $W = \omega_2 X_2 + \omega_3 X_3$ and $Y = \gamma_v V + \gamma_w W + \gamma_1 X_1$. Given this, one can calculate what would happen to Y, for example, if internal variable V were different.

[1] On the most common definition of exogeneity, exogeneity is a concept that applies to models in which each equation contains a "disturbance term", for example the U term in the equation $Y = \gamma X + U$. Exogeneity claims that the mathematical expectation $E(U|X)$ is equal to $E(U)$. See Engle, Hendry, and Richard (1983) for a classic discussion.

This talk of "interventions breaking equations" is metaphorical. So let me give a more rigorous description of the recipe that the modular theory suggests for answering what–if questions with respect to the educational attainment model. In the educational attainment model, the first step in the recipe is to write down the values $\{x_1, x_2, x_3, v, w, y\}$ that each of the variables $\{X_1, X_2, X_3, V, W, Y\}$ actually took in the case of a particular child. For example, let's take a child (called Menno) for whom:

$$x_1 = 10$$
$$x_2 = 3$$
$$x_3 = 4$$
$$v = 4x_1 + x_2 = 43$$
$$w = \tfrac{1}{2}x_2 + 2x_3 = 9.5$$
$$y = \tfrac{1}{2}v + 2w + 5x_1 = 90.5$$

(For ease of I illustration, I've filled in the unknown values of the constants $\nu_1, \nu_2, \omega_2, \omega_3, \gamma_v, \gamma_w, \gamma_1$ with some specific values, namely $4, 1, \frac{1}{2}, 2, \frac{1}{2}, 2, 5$.) The second step is to consider the hypothetical scenario in which the education policy X_1 in Menno's region had been 9 units for example (instead of its actual value of 10 units). How should one calculate the values of all the other variables under this hypothetical scenario in which X_1 differs? The modular theory endorses a principle called Modularity (Pearl 2009, 22–32, 69). Modularity makes precise the rough idea that "interventions" on a set of variables C will leave all the other external variables and direct–causes equations "intact":

> (Modularity) Imagine that you want to evaluate what would occur in any hypothetical scenario of the form: "if variable C_1 had instead taken value c'_1 (any value you like), and variable C_2 had instead taken value c'_2 (any value you like)". Here's how to do it:
>
> (a) For any external variable X (other than C_1 and C_2), the value that variable X would take in this hypothetical scenario is equal to the value that X took in the actual scenario;
>
> (b) For any internal variable Y (other than C_1 and C_2), the value that variable Y would take in this hypothetical scenario is correctly predicted by the equation that (in the actual scenario) specifies the direct causes of Y;
>
> (c) Variable C_1 takes the value c'_1 and variable C_2 takes the value c'_2 of course.

To understand what Modularity means, look how one can use it to calculate the values $\{x_1', x_2', x_3', v', w', y'\}$ that each of the variables $\{X_1, X_2, X_3, V, W, Y\}$ would have taken, in Menno's case, under the hypothetical scenario in which $X_1 = 9$:

$x_1' = 9$ according to (c) from Modularity

$x_2' = 3$ according to (a) from Modularity

$x_3' = 4$ according to (a) from Modularity

$v' = 4x_1' + x_2' = 39$ according to (b) from Modularity

$w' = \frac{1}{2}x_2' + 2x_3' = 9.5$ according to (b) from Modularity

$y' = \frac{1}{2}v' + 2w' + 5x_1' = 83.5$ according to (b) from Modularity

The modular theory then adds that

Causal contributions match hypothetical differences:

Whenever q denotes the value that variable Q actually took in a particular case, and p denotes the value that variable P actually took; and

Whenever q' denotes the value that Q would have taken under the hypothetical scenario in which P had taken the value p' instead;

Then P's taking value p (rather than taking value p') made an overall causal contribution to variable Q in the case in question of $q - q'$ units.

This tells us that the overall causal contribution that education policy X_1 taking value 10 (rather than taking value 9) made to this child's educational attainment is $y' - y = 90.5 - 83.5 = 7$ units. (Contrast this overall causal contribution with the direct causal contribution of education policy X_1 to educational attainment Y, namely 5 units. The overall contribution differs from the direct contribution, of course, because education policy X_1 also makes an indirect contribution to educational attainment Y, namely via class size V.)

This illustrates how the modular theory can be used to calculate the overall causal contribution that an external variable made to an internal variable. More controversially, the modular theory can also be used to calculate the overall causal contribution that any internal variable made to any other internal variable—and equally to calculate the value of any internal variable in any hypothetical scenarios in which one or more internal variables are hypothesized to differ. For example one can calculate the values $\{x_1'', x_2'', x_3'', v'', w'', y''\}$ that each of the variables $\{X_1, X_2, X_3, V, W, Y\}$ would have taken, in Menno's case, under the hypothetical scenario in which V is one unit less (than its actual value of 43):

$x_1'' = 10$ according to (a) from Modularity

$x_2'' = 3$ according to (a) from Modularity

$x_3'' = 4$ according to (a) from Modularity

$v'' = 43 - 1 = 42$ according to (c) from Modularity

$w'' = \frac{1}{2}x_2'' + 2x_3'' = 9.5$ according to (b) from Modularity

$y'' = \frac{1}{2}v'' + 2w'' + 5x_1'' = 90$ according to (b) from Modularity

This tells us that the overall causal contribution that class size being 43 (rather than 42) made to this child's educational attainment is $y' - y = 90.5 - 90 = .5$ units.

This is Pearl and Woodward's modular theory of overall causal contributions. Their theory allows one to calculate the overall causal contribution that any variable in the model made to any other variable in the model. Similarly, Modularity also allows one to predict the value that any variable would have taken under hypothetical differences to one or more of the other variables in the model. Note that the modular theory relies on direct–causal equations as an input. It presupposes that, for each internal variable, the model supplies us with exactly one equation that describes the direct causes of that variable.

4 Problems for the Modular Theory in Economics

The modular theory is popular in the sciences in general, but it is much less popular in economics in particular. To see why this is, consider the economic model of supply and demand:

Demand equation: $Q = \alpha P + \alpha_1 X_1 + \alpha_2 X_2$

Supply equation: $Q = \beta P + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4$

Imagine that the first equation (the demand equation) describes the quantity of lumber Q that is produced in a given period. (This model assumes that the market for lumber is in equilibrium: the quantity of lumber produced Q is the same as the quantity of lumber Q that is purchased in that period.) The demand equation relates this quantity Q to the price P at which purchasers can buy lumber during this period, the price X_1 at which purchasers can buy brick during this period, and X_2 the overall income of consumers in the economy. The second equation (the supply equation) relates quantity Q to the price P at which producers can sell lumber during this period, the price X_1 at which producers can sell brick during this period, and X_3 the technological conditions that determine how easy it is to produce lumber, and X_4 the technological conditions that determine how easy it is to produce brick. $\alpha, \alpha_1, \alpha_2, \beta, \beta_1, \beta_3$ and β_4 are unknown constants that don't vary across economies (in the population of market economies that we are studying).

To apply the modular theory to this model, we need direct–causes equations. But, as the supply and demand equations currently stand, there is a problem with interpreting them as direct–causes equations. Note that, as they stand, both equations have Q on the left-hand side. So, if we interpret each equation in the model as describing the direct causes of the left-hand side variable in the equation, the result is two contradictory stories about the direct causes of Q. The demand equation says that $\{P, X_1, X_2\}$ are the only direct causes of Q; the supply equation says that $\{P, X_1, X_3, X_4\}$ are the only direct causes of Q. Even worse, consider a hypothetical scenario in which P differs—for example, the hypothetical scenario in which P takes a value one unit less (than the value P actually took). These contradictory stories about the direct causes of Q lead to an incoherent story about the value that Q would take under this hypothetical scenario, one can show. Assuming that the demand equation correctly describes the direct causes of Q, Modularity says that the demand equation would hold under this hypothetical scenario. But Modularity also says that none of the X variables would differ under this hypothetical scenario. It follows that Q would be α units less under this hypothetical scenario (than the value that Q actually took). In contrast, however, if one assumes that the supply equation also correctly describes the direct causes of Q, then by the exact same logic we can conclude that Q would be β units less under this hypothetical scenario. So, unless $\alpha = \beta$, Modularity issues in an incoherent description of the value that Q would take under this hypothetical scenario.

This problem is easily fixed, however, by changing our interpretation of the equations in the supply–demand model. One option is to interpret the demand equation as describing the direct causes of Q, and to interpret the supply equation as describing the direct causes of P. To work with this interpretation, one re-expresses the supply equation in a mathematically equivalent form:

> There are some (unknown) values of constants $\alpha, \alpha_1, \alpha_2, \beta, \beta_1, \beta_3$ and β_4, such that for any economy (in the population being studied) the following two equations describe the direct causal contributions made to Q and P respectively:
>
> $Q = \alpha P + \alpha_1 X_1 + \alpha_2 X_2$
>
> $P = -(1/\beta)Q + (\beta_1/\beta)X_1 + (\beta_3/\beta)X_3 + (\beta_4/\beta)X_4$

This interpretation of the supply and demand equations clears up the issue of what directly causes what: $\{P, X_1, X_2\}$ are the direct causes of Q; and $\{Q, X_1, X_3, X_4\}$ are the direct causes of P. (Note the "mutual causation" in which

P causes Q and Q causes P. This mutual causation may sound weird, but at least it is not logically contradictory.)

This interpretation also allows the modular theory to issue in a coherent description of what would happen under the hypothetical scenario in which P, for example, had taken a value p' one unit less (than the value p that P actually took). Imagine for illustration that $x_1 = 8$, $x_2 = 3$, $x_3 = 4$ and $x_4 = 9$ are the values that the external variables actually took in the economy in question. Modularity says that the values $\{x'_1, x'_2, x'_3, x'_4, p', q'\}$ that the variables $\{X_1, X_2, X_3, X_4, P, Q\}$ would have taken, in the case of this particular economy, under the hypothetical scenario in which $P = p - 1$, are:

$x'_1 = 8$ according to (a) from Modularity

$x'_2 = 3$ according to (a) from Modularity

$x'_3 = 4$ according to (a) from Modularity

$x'_4 = 9$ according to (a) from Modularity

$q' = \alpha p' + \alpha_1 x'_1 + \alpha_2 x'_2$ according to the demand equation and (b) from Modularity

$p' = p - 1$ according to (c) from Modularity

Since the value of Q in the actual scenario is given by $q = \alpha p + \alpha_1 x_1 + \alpha_2 x_2$, as per the demand equation, it follows that $q - q' = \alpha$, one can calculate. Since the modular theory says that causal contributions match hypothetical differences, it follows that P taking value p (rather than $p-1$) made a causal contribution to Q in this particular economy of α units. In this respect, the causal contribution that price P makes to quantity Q can be "read off" the demand equation (in which α is the coefficient of the P variable).

Note that under the hypothetical scenario above, the supply equation is violated.[2] So under this hypothetical scenario, the supply side of the economy is not working as it normally does. In effect, this hypothetical scenario is the scenario in which the government has nationalised the production of lumber, has fixed the price of lumber at $p - 1$, and has guaranteed to produce as much lumber as was needed to keep up with demand. That is to say, this hypothetical scenario describes a government monopoly on lumber production.

However, an alternative (and equally well-motivated) interpretation of the supply and demand model is instead to interpret the demand equation as describing

[2] Since the supply equation does hold in the actual scenario we have $q = \beta p + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4$ and so $(q - \alpha) + \alpha = \beta + \beta(p - 1) + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4$. By substitution, we have $q' + \alpha = \beta + \beta p' + \beta_1 x'_1 + \beta_3 x'_3 + \beta_4 x'_4$, and so $q' \neq \beta p' + \beta_1 x'_1 + \beta_3 x'_3 + \beta_4 x'_4$, unless $\alpha = \beta$.

the direct causes of P, and to interpret the supply equation as describing the direct causes of Q. To work with this interpretation, one instead re-expresses the demand equation in a mathematically equivalent form:

> There are some (unknown) values of constants $\alpha, \alpha_1, \alpha_2, \beta, \beta_1, \beta_3$ and β_4, such that for any economy (in the population being studied) the following two equations describe the direct causal contributions made to P and Q respectively:
>
> $P = -(1/\alpha)Q + (\alpha_1/\alpha)X_1 + (\alpha_2/\alpha)X_2$
>
> $Q = \beta P + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4$

On this interpretation of the supply and demand equations, $\{P, X_1, X_3, X_4\}$ are the direct causes of Q; and $\{Q, X_1, X_2\}$ are the direct causes of P. This alternative interpretation issues in an alternative answer to the question of the value that Q would have taken in this particular economy, if P had taken a value p' one unit less (than the value p that P actually took). Modularity says that the values $\{x'_1, x'_2, x'_3, x'_4, p', q'\}$ that the variables $\{X_1, X_2, X_3, X_4, P, Q\}$ would have taken, in the case of this particular economy, under the hypothetical scenario in which $P = p - 1$, are:

> $x'_1 = 8$ according to (a) from Modularity
>
> $x'_2 = 3$ according to (a) from Modularity
>
> $x'_3 = 4$ according to (a) from Modularity
>
> $x'_4 = 9$ according to (a) from Modularity
>
> $p' = p - 1$ according to (c) from Modularity
>
> $q' = \beta p' + \beta_1 x'_1 + \beta_3 x'_3 + \beta_4 x'_4$ according to (b) from Modularity

Since the value of Q in the actual scenario is given by $q = \beta p + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4$, as per the supply equation, it follows that $q - q' = \beta$, one can calculate. So, on this alternative interpretation of what the supply and demand model means, the modular theory says that P taking value p (rather than $p - 1$) made a causal contribution to Q in this particular economy of β units. In this respect, the causal contribution that price P makes to quantity Q can be "read off" the supply equation (in which β is the coefficient of the P variable).

Note that under the hypothetical scenario above, the demand equation is violated.[3] So under this hypothetical scenario, the demand side of the economy is

[3] Since the demand equation does hold in the actual scenario we have $q = \alpha p + \alpha_1 x_1 + \alpha_2 x_2$, and so $(q - \beta) + \beta = \alpha + \alpha(p-1) + \alpha_1 x_1 + \alpha_2 x_2$. By substitution we have $q' + \beta = \alpha + \alpha p' + \alpha_1 x'_1 + \alpha_2 x'_2$. And so $q' \neq \alpha p' + \alpha_1 x'_1 + \alpha_2 x'_2$, unless $\alpha = \beta$.

not working as it normally does. In effect, this hypothetical scenario is the scenario in which the government has banned the direct sale of lumber to lumber purchasers, and has instead insisted that the producers of lumber sell only to the government, and at a fixed price of $p - 1$, guaranteeing to buy as much as the producers are willing to produce. That is to say, this hypothetical scenario is a government monopsony on lumber purchase.

This illustrates how, when applying the modular theory, one needs to make a judgment call about what the direct causes of each internal variable are. And for the supply and demand system, there seem to be two equally well-motivated judgment calls here—with each judgment call issuing in a distinct measure of the causal contribution that P made to Q.

Having fixed this minor problem, however, a major problem remains: when economists ask of a particular market economy "what would value would quantity Q have taken, if the price P had taken value $p-1$ instead?", they often are not interested in hypothetical scenarios in which $P = p-1$ arose via a non-competitive arrangement, such as a government monopoly on lumber production, or a government monopsony on lumber sales. This is simply not the hypothetical scenario that most economists are interested in. Rather, most economists are interested in the normal workings of the market economy as a competitive system. So they are interested in a hypothetical scenario in which $P = p - 1$ arose from a competitive arrangement between buyers and between producers. Therefore the modular theory, when applied to the supply–demand model, is answering a question that economists are not usually interested in. (An analogy might help here: when asking what would have happened if the steering wheel in a car had been rotated right ninety degrees, one is usually interested in what would have happened if this had occurred with the car working as normal. One is not usually interested in what would have happened if this had occurred via someone detaching the steering wheel from the car, and then rotating the steering wheel by ninety degrees.)

This major problem for the modularity theory has been pressed most forcefully by Cartwright (2007). For Pearl's response see Pearl (2009, 106, 363–65, 374–78).

5 The General Supply–Demand Equation

To explore the supply and demand system in more depth, it will be useful to derive several consequences from the supply and demand equations. Do to this, multiply the demand equation by any constant λ you like, and multiply the supply equation by any constant μ you like. This yields

$$\lambda Q = \lambda\alpha P + \lambda\alpha_1 X_1 + \lambda\alpha_2 X_2$$
$$\mu Q = \mu\beta P + \mu\beta_1 X_1 + \mu\beta_3 X_3 + \mu\beta_4 X_4$$

Adding these two equations together gives

$$(\lambda + \mu)Q = (\lambda\alpha + \mu\beta)P + (\lambda\alpha_1 + \mu\beta_1)X_1 + \lambda\alpha_2 X_2 + \mu\beta_3 X_3 + \mu\beta_4 X_4$$

Dividing by $\lambda + \mu$ gives what I will call the *general supply–demand equation*, which combines the information from both the supply and demand equations:

$$Q = \frac{\lambda\alpha + \mu\beta}{\lambda + \mu}P + \frac{\lambda\alpha_1 + \mu\beta_1}{\lambda + \mu}X_1 + \frac{\lambda\alpha_2}{\lambda + \mu}X_2 + \frac{\mu\beta_3}{\lambda + \mu}X_3 + \frac{\mu\beta_4}{\lambda + \mu}X_4$$

One can derive various useful facts from this general supply-demand equation, by choosing particular values for λ and μ. Firstly, when you let $\mu = -\alpha$ and $\lambda = \beta$ in the general supply–demand equation, one derives the more specific equation:

$$Q = \frac{\beta\alpha_1 - \alpha\beta_1}{\beta - \alpha}X_1 + \frac{\beta\alpha_2}{\beta - \alpha}X_2 - \frac{\alpha\beta_3}{\beta - \alpha}X_3 - \frac{\alpha\beta_4}{\beta - \alpha}X_4$$

Since each of $\{X_1, X_2, X_3, X_4\}$ is an external variable, and the above function of Q is derived from our model, the definition of external variables from Section 2 says that each of $\{X_1, X_2, X_3, X_4\}$ is a cause of Q—although not necessarily a direct cause of Q. Secondly, since $\{X_1, X_2, X_3, X_4\}$ are each external variables, Q must be an internal variable, otherwise the variation-freedom condition on external variables would fail (again see Section 2).

Thirdly, when one lets $\lambda = -1$ and $\mu = 1$ (and divides by $\alpha - \beta$ rather than by $\lambda + \mu$), one derives the more specific equation:

$$P = \frac{\beta_1 - \alpha_1}{\alpha - \beta}X_1 - \frac{\alpha_2}{\alpha - \beta}X_2 + \frac{\beta_3}{\alpha - \beta}X_3 + \frac{\beta_4}{\alpha - \beta}X_4$$

It will be absolutely crucial for the discussion that follows to note from this equation that: given that the supply and demand equations hold, the values of $\{X_1, X_2, X_3, X_4\}$ together predict the value that P takes. Fourthly, it follows from this that P is an internal variable that is caused by each of $\{X_1, X_2, X_3, X_4\}$.

6 The Ceteris Paribus Theory

The problem for the modular theory that I discussed in Section 4 motivates the search for an alternative theory of overall causal contributions and of what–if hypotheticals. This section will develop an alternative theory, which I will call the *ceteris paribus theory*, and which I take to be in the spirit of Heckman's theory (Heckman 2000, 2005; Heckman and Vytlacil 2007), although there are a number of ambiguities in Heckman's own theory that make it unclear whether Heckman would endorse the ceteris paribus theory as I formulate it.

Heckman's theory doesn't require us to make any assumptions about direct causes. And so, unlike the previous sections, I will no longer assume anything about the direct causes of P and of Q. Instead I will assume only that: (i) the external variables in the supply–demand model are $\{X_1, X_2, X_3, X_4\}$, and (ii) the supply equation and the demand equation correctly predict the values of P and Q under hypothetical differences in these external variables. In virtue of this, the supply and demand equations are assumed to be "externally stable", to coin a phrase. It follows, as I pointed out in the last section, that $\{X_1, X_2, X_3, X_4\}$ are each causes of Q and are each causes of P. I will also assume, as most economists do, that: (iii) P is a cause of Q. After all, without this assumption, discussion of the causal contribution that P makes to Q is meaningless.

With these assumptions in hand, let's now consider a hypothetical scenario under which P takes a value one unit less (than P took in the actual scenario). One might then be tempted to reason (naively) as follows. "Note the variables $\{P, X_1, X_2\}$ on the right-hand side of the demand equation. These variables are each a cause of the variable Q on the left-hand side of the demand equation. But one might assume that, under a hypothetical scenario in which P differs, each of the other causes in this set $\{P, X_1, X_2\}$ would take the same value (as it took in the actual scenario). That is to say, one might assume that X_1 and X_2 don't differ under this hypothetical scenario. If one also assumes that the demand equation correctly predicts what would happen under this hypothetical scenario, it follows: under this hypothetical scenario, Q would have taken the value $q' = \alpha(p-1) + \alpha_1 x_1 + \alpha_2 x_2$. But, since Q actually took the value $q = \alpha p + \alpha_1 x_1 + \alpha_2 x_2$, it follows that $q - q' = \alpha$. And, since causal contributions match hypothetical differences, one assumes, α is the causal contribution that an extra unit of P made to Q in this particular economy."

I will call the general idea here the *extremely naive ceteris paribus theory*:

> Whenever (I) $Q = q(P, O)$ expresses Q as an externally stable function of P and some other variables O; and
>
> Whenever (II) P and O are each causes of Q; and
>
> Whenever o denotes the values that these other variables O actually took;
>
> Then, under the hypothetical scenario in which P instead had taken the value p', O would have taken the value o, and Q would have taken the value $q' = q(p', o)$.
>
> Since causal contributions match hypothetical differences, P's taking value p (rather than p') made an overall causal contribution to variable

Q in the case in question of $q - q'$ units.

I use the label "ceteris paribus"—all else being equal—to mark the fact that this theory uses hypothetical scenarios in which these other causes O take the values that they actually take. (It's also worth noting that, according to the ceteris paribus theory, facts about what what would happen in a hypothetical scenario in which P is different are relative to the variable Q that one chooses to focus on as an outcome variable.)

Why is this theory extremely naive? Consider the case in which a body-builder is trying to gain muscle mass. Imagine that muscle mass is caused by gym activity and by protein consumed, in accordance with the equation Mass = 3Gym + 4Protein. Imagine also that gym activity is also a cause of protein consumption, in accordance with the equation Protein = 2Gym. Note that the extremely naive ceteris paribus theory mistakenly entails that the overall contribution that an extra unit of Gym makes to Mass is 3 units. But this is incorrect: 3 is the merely the direct causal contribution that Gym makes to Mass; there is also the indirect contribution that Gym makes to Mass through Protein consumption as an intermediary. To calculate the overall causal contribution, note that Mass = 3Gym + 4(2Gym) = 11Gym.

Fixing this problem is easy, of course. One improves the ceteris paribus theory by adding to it the condition that the ceteris paribus theory only applies:

> Whenever (III) the cause in question (for example P or Gym) is not a cause of any of the other variables O

This improved theory fixes the problem. Mass = 3Gym + 4Protein fails the condition of application *III* of the improved theory, because Gym is a cause of Protein.

What does this improved theory say about the supply and demand system? Since P is an internal variable, our definition of internal variables in Section 2 tells us that P is not a cause of X_1 or X_2. And so the improved theory can be applied to the demand equation $Q = \alpha P + \alpha_1 X_1 + \alpha_2 X_2$ to calculate that an extra unit of P makes an overall contribution of α units to Q. (But, since X_1 and X_2 are causes of P, condition *III* says that the improved theory cannot be applied to the demand equation. Thus condition *III* prevents one from drawing the conclusion that an extra unit of X_1 makes an overall contribution of α_1 units to Q.)

I call this improved theory the *somewhat naive ceteris paribus theory*. This is because this ceteris paribus theory still faces a major problem. To see this problem, recall the general supply–demand equation from Section 5.

$$Q = \frac{\lambda\alpha + \mu\beta}{\lambda + \mu}P + \frac{\lambda\alpha_1 + \mu\beta_1}{\lambda + \mu}X_1 + \frac{\lambda\alpha_2}{\lambda + \mu}X_2 + \frac{\mu\beta_3}{\lambda + \mu}X_3 + \frac{\mu\beta_4}{\lambda + \mu}X_4$$

This equation is externally stable, because it follows from two externally stable equations. In virtue of this, this equation satisfies condition *I* from the somewhat naive ceteris paribus theory. But recall that the last section established that these other variables $O = \{X_1, X_2, X_3, X_4\}$ are each causes of Q. And recall that we are assuming that P is a cause of Q. And so, in virtue of this, condition *II* is satisfied too. But, P is an internal variable, as shown in Section 5. And it follows from our definition of an internal variable that P doesn't cause any of $O = \{X_1, X_2, X_3, X_4\}$, since they are all external variables. So, in virtue of this, condition *III* is satisfied too. So the general supply–demand equation satisfies conditions *I–III* for applying the somewhat naive ceteris paribus theory. The result is that $(\lambda\alpha + \mu\beta)/(\lambda + \mu)$ is a correct description of the causal contribution that each extra unit of P makes to Q. But note that λ and μ can take any values that you like, and conditions *I–III* are still satisfied. And so $(\lambda\alpha + \mu\beta)/(\lambda + \mu)$ can be any value that you like. Therefore the causal contribution of P to Q is correctly described by any number that you like, positive or negative, according to the somewhat naive ceteris paribus theory. It is maximally indeterminate, is the unwelcome conclusion.

How might one further improve the ceteris paribus theory to avoid this unwelcome conclusion? Let's say that, whenever both the supply and the demand equations hold, the economy is "working normally" as a competitive economy. Given this, one might improve the ceteris paribus theory by stipulating that the theory only applies to hypothetical scenarios in which the economy is working normally. Namely, one might add the following condition:

> Whenever (IV) there is a logically possible hypothetical scenario in which P takes value p', and the other variables O takes value o, and the system is "working normally"

This "normal workings" condition can be motivated independently from the desire to avoid the unwelcome conclusion above: at the end of the Section 4, I suggested that economists are typically interested in hypothetical scenarios in which there is competition both between buyers and between producers; when modeling competitive economies, economists are not typically interested in hypothetical scenarios in which the government has imposed a monopoly on lumber production, or a monopsony on lumber purchases, for example. That is to say, they are interested only in hypothetical cases in which the economy is working normally.

To see the implications of this "normal workings" condition, remember from Section 5 that, given that the supply and demand equations hold, the values of the other variables $O = \{X_1, X_2, X_3, X_4\}$ predict the value that P takes. So there is no logically possible hypothetical scenario in which (a) P is one unit less (than in the actual scenario), (b) all the other variables $O = \{X_1, X_2, X_3, X_4\}$ took the same values (as they took in the actual scenario), and (c) the system is working normally.

And so the hypothetical scenario in which P differs, but in which all the other variables $O = \{X_1, X_2, X_3, X_4\}$ remain the same, fails the "working normally" condition. So the set of other variables $O = \{X_1, X_2, X_3, X_4\}$ fails condition *IV*. And so, unlike the somewhat naive ceteris paribus theory, the "normal workings" ceteris paribus theory does not issue in the unwelcome conclusion that—for any value of λ and μ that you like—the causal contribution that each extra unit of P makes to Q is correctly described by $(\lambda\alpha + \mu\beta)/(\lambda + \mu)$. That is to say, it does not issue in the conclusion that this causal contribution is maximally indeterminate.

Having said that, consider the more specific equation that results when you let $\mu = -\alpha_1$ and $\lambda = \beta_1$ in the general supply–demand equation:

$$Q = \frac{\beta_1\alpha - \alpha_1\beta}{\beta_1 - \alpha_1}P + \frac{\beta_1\alpha_2}{\beta_1 - \alpha_1}X_2 - \frac{\alpha_1\beta_3}{\beta_1 - \alpha_1}X_3 - \frac{\alpha_1\beta_4}{\beta_1 - \alpha_1}X_4$$

This specific choice of μ and λ ensures that the coefficient for the X_1 term in the general supply–demand equation is zero, and so it eliminates X_1 from the general supply–demand equation. So the other variables O in this more specific equation are $O = \{X_2, X_3, X_4\}$. But it is possible for P to differ (from its actual value) and for these other variables $O = \{X_2, X_3, X_4\}$ to take the same values (as they actually took) and for the economy to work as normal. Therefore, the "normal workings" ceteris paribus theory applies to this more specific equation, an equation in which the other variables are $O = \{X_2, X_3, X_4\}$. The theory says that $(\beta_1\alpha - \alpha_1\beta)/(\beta_1 - \alpha_1)$ is a correct description of the overall causal contribution that an extra unit of P made to Q.

Similarly, if you let $\mu = 0$, the more specific equation that results from the general supply–demand equation is just the demand equation itself $Q = \alpha P + \alpha_1 X_1 + \alpha_2 X_2$. This specific choice of μ ensures that coefficients for both the X_3 and the X_4 term in the general supply–demand equation is zero, and so it eliminates X_3 and X_4 from the general supply–demand equation. So the other variables O in this more specific equation are $O = \{X_1, X_2\}$. But it is possible for P to differ (from its actual value) and for the other variables $O = \{X_1, X_2\}$ to take the same values (as they actually took) and for the economy to work as normal. Therefore, the "normal workings" ceteris paribus theory applies to this more specific equation, namely the demand equation, an equation for which the other variables are $O = \{X_1, X_2\}$. The theory says that α is also a correct description of the overall causal contribution that one unit of P made to Q.

Similarly, if you let $\lambda = 0$, the more specific equation that results from the general supply–demand equation is just the supply equation itself $Q = \beta P + \beta_1 X_1 + \beta_3 X_3 + \beta_4 X_4$. This specific choice of λ ensures that the coefficient for the X_2 terms in the general supply–demand equation is zero, and so it eliminates X_2 from the general supply–demand equation. So the other variables O in this more specific equation are $O = \{X_1, X_3, X_4\}$. But it is possible for P to differ (from its actual value) and for these other variables $O = \{X_1, X_3, X_4\}$ to take the same values (as they actually too) and for the economy to work as normal. Therefore, the "normal workings" ceteris paribus theory applies to this more specific equation, namely the supply equation, an equation for which $O = \{X_1, X_3, X_4\}$. The theory says that β is also a correct description of the overall causal contribution that one unit of P made to Q.

Further inspection shows that there are no other choices of λ and μ that eliminate any of the Xs from the general supply–demand equation. However, only when one finds a function for Q in which one has eliminated one of the Xs from the right-hand side of the function, is it possible for P to differ while the other variables O on the right hand side of the function take the same values. And so there are only three ways to apply the normal workings ceteris paribus theory to the supply and demand system. The first way is to apply it with $O = \{X_2, X_3, X_4\}$ to show that $(\beta_1\alpha - \alpha_1\beta)/(\beta_1 - \alpha_1)$ is a correct description of the overall causal contribution that one unit of P made to Q. The second way is to apply it with $O = \{X_1, X_2\}$ to show that α is also a correct description of this overall causal contribution too. The third way is to apply it with $O = \{X_1, X_3, X_4\}$ to show that β is also a correct description. Thus, for the "normal workings" ceteris paribus theory, causal contributions are relative to the choice of O. (Compare and contrast the modular theory which said that there is only one correct descriptions, namely β or α, depending on what you judge the direct causes of Q to be.)

In sum, I've explained and motivated the *normal ceteris paribus theory*:

> Whenever (I) $Q = q(P, O)$ expresses Q as an externally stable function of P and some other variables O; and
>
> Whenever (II) P and O are each causes of Q; and
>
> Whenever (III) P is not a cause of any of the other variables O; and
>
> Whenever o denotes the values that these other variables O actually took; and
>
> Whenever (IV) there is a logically possible hypothetical scenario in which P takes value p', and O takes value o, and the system is "working normally"
>
> Then, under the hypothetical scenario in which P instead had taken the value p', O would have taken the value o, and Q would have taken the value $q' = q(p', o)$. This answer is relative to one's choice of O and Q.
>
> Since causal contributions match hypothetical differences, P's taking value p (rather than p') made an overall causal contribution to variable

Q in the case in question of $q - q'$ units. This answer is relative to one's choice of O and Q.

Thus, when examining hypothetical scenarios in which P differs, the normal ceteris paribus theory takes a different approach from the modular theory. On the one hand, the normal ceteris paribus theory considers hypothetical scenarios in which the system is working normally, scenarios for example in which both the supply and demand equations hold. The modular theory, one the other hand, considers situations in which one of these two equations fails, and so the economy is not a fully competitive market economy. We've already seen how this leads to different conclusions: the normal ceteris paribus theory says that $(\beta_1\alpha - \alpha_1\beta)/(\beta_1 - \alpha_1)$ is a correct description of the causal contribution of P to Q, whereas the modular theory denies this.

To further understand the differences between these two theories, let's return to the educational attainment model, and let's say that the educational system is working normally if and only if the three equations in the model all hold:

$$V = \nu_1 X_1 + \nu_2 X_2$$
$$W = \omega_2 X_2 + \omega_3 X_3$$
$$Y = \gamma_v V + \gamma_w W + \gamma_1 X_1$$

Now, it follows from the first equation here that that $\lambda V - \lambda\nu_1 X_1 - \lambda\nu_2 X_2 = 0$ holds for any value of λ you like. But we've already established that $Y = (\gamma_v\nu_1 + \gamma_1)X_1 + (\gamma_v\nu_2 + \gamma_w\omega_2)X_2 + \gamma_w\omega_3 X_3$ follows from these equations. Adding the former equation to the latter gives us, for any value of λ you like, the general equation:

$$Y = \lambda V + ([\gamma_v - \lambda]\nu_1 + \gamma_1)X_1 + ([\gamma_v - \lambda]\nu_2 + \gamma_w\omega_2)X_2 + \gamma_w\omega_3 X_3.$$

If one chooses $\lambda = \gamma_v + \gamma_1/\nu_1$, then the more specific equation that results from this general equation is one in which the X_1 term is eliminated:

$$Y = (\gamma_v + \frac{\gamma_1}{\nu_1})V + (\gamma_w\omega_2 - \frac{\gamma_1\nu_2}{\nu_1})X_2 + \gamma_w\omega_3 X_3.$$

This equation is externally stable, because it follows from two externally stable equations. In virtue of this, this equation satisfies condition *I* from the normal ceteris paribus theory. But $\{V, X_2, X_3\}$ are each causes of Y, and so condition *II* is satisfied also. But, since V is an internal variable, it follows from our definition of an internal variable that V doesn't cause any of the external variables $\{X_1, X_2, X_3\}$. In virtue of V not causing X_2 or X_3, condition *III* is satisfied also. But there is a hypothetical scenario in which X_2 and X_3 each takes the same value (as it took in the actual scenario), but in which V takes a value one unit lower (than it took in the actual scenario), and in which the system works normally. (Under this hypothetical scenario, X_1 takes a value $1/\nu_1$ units lower than it actually took.) In virtue of this, condition *IV* is satisfied also. So the normal ceteris paribus theory applies to this equation. The result is that $\gamma_v + \gamma_1/\nu_1$ is a correct description of the overall causal contribution that an extra unit of V makes to Y. That is to say, relative to $O = \{X_2, X_3\}$.

However, if instead one chooses $\lambda = \gamma_v + \gamma_w\omega_2/\nu_2$, then the more specific equation that results from this general equation is one in which the X_2 term is eliminated:

$$Y = (\gamma_v + \frac{\gamma_w\omega_2}{\nu_2})V + (\gamma_1 - \frac{\gamma_w\omega_2\nu_1}{\nu_2})X_1 + \gamma_w\omega_3X_3.$$

But there is a hypothetical scenario in which X_1 and X_3 each takes the same value (as it took in the actual scenario), but in which V takes a value one unit lower (than it took in the actual scenario) and in which the system works normally. (Under this hypothetical scenario, X_2 takes a value $1/\nu_2$ units lower than it actually took.) So, by the same logic, the normal ceteris paribus theory applies to this equation. The theory says that $\lambda = \gamma_v + \gamma_w\omega_2/\nu_2$ is also a correct description of the overall causal contribution that an extra unit of V makes to Y. That is to say, relative to $O = \{X_1, X_3\}$.

However, what if instead one considers our original equation $Y = \gamma_v V + \gamma_w W + \gamma_1 X_1$? Note that there is a hypothetical scenario in which in which W and X_1 each takes the same value (as it took in the actual scenario), but in which V takes a value one unit lower (than it took in the actual scenario), and in which the system is working normally. (Under this hypothetical scenario, X_2 takes a value $1/\nu_2$ units lower than it actually took, and X_3 takes a value $\omega_2/\omega_3\nu_2$ units higher than it actually took.) In virtue of this, condition *IV* is satisfied. Let's also assume that V doesn't cause W. In virtue of this, condition *III* is satisfied also. Let's also assume that W is a cause of Y. In virtue of this, condition *II* is satisfied also. So the normal ceteris paribus theory applies to this equation. The theory says that the causal contribution that an extra unit of V made to Y was γ_v units. That is to say, relative to $O = \{W, X_1\}$.

Standing back from this, this illustrates how the normal ceteris paribus theory differs from the modular theory, which says that γ_v is the only correct description of this causal contribution.

7 Conclusion

This chapter has contrasted two theories of causal contributions and of what-if questions, the modular theory and the ceteris paribus theory. The modular theory requires information about direct causal contributions as an input, and it faces Cartwright's objection that it (arguably) it interprets causal contributions and what-if questions in a way that makes them uninteresting to economists. The ceteris paribus theory only requires information about which variables are external variables (and which equations are stable under differences in the external variables). But it entails that causal contributions are often relative to one's choice of "other variables", the variables that one imagines "holding fixed" while one imagines the cause in question varying.

7 References

Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. Oxford University Press.

———. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge; New York: Cambridge University Press.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard. 1983. "Exogeneity." *Econometrica* 51: 277–304. https://doi.org/10.2307/1911990.

Haavelmo, Trygve. 1943. "The Statistical Implications of a System of Simultaneous Equations." *Econometrica* 11: 1–12. https://doi.org/10.2307/1905714.

Heckman, James J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Centuryretrospective." *The Quarterly Journal of Economics* 115: 45–97.

———. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35: 1–97. https://doi.org/10.1111/j.0081-1750.2006.00164.x.

Heckman, James J., and Edward J. Vytlacil. 2007. "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causalmodels, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*, 6:4779–4874. Elsevier. https://doi.org/10.1016/s1573-4412(07)06070-9.

Hoover, Kevin D. 2001. *Causality in Macroeconomics*. Cambridge University Press.

———. 2011. "Counterfactuals and Causal Structure." In *Causality in the Sciences*, edited by Phyllis Mckay Illari, Federica Russo, and Jon Williamson. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.001.0001.

———. 2013. "Identity, Structure, and Causal Representation in Scientific Models." In *Mechanism and Causality in Biology and Economics*, edited by Hsiang-ke Chao, Szu-ting Chen, and Roberta L. Millstein, 35–57. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2454-9_3.

LeRoy, Stephen F. 2016. "Implementation-Neutral Causation." *Economics and Philosophy* 32: 121–42. https://doi.org/10.1017/S0266267115000280.

LeRoy, Stephen F. n.d. "Implementation-Neutral Causation in Structural Models," 22.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. New York Ny Usa: Cambridge University Press.

Reiss, Julian. 2009. "Counterfactuals, Thought Experiments, and Singular Causal Analysisin History." *Philosophy of Science* 76: 712–23.

———. 2012. "Counterfactuals." In *The Oxford Handbook of Philosophy of Social Science*, edited by Harold Kincaid. Oxford ; New York: Oxford University Press.

Simon, Herbert A. 1953. "Causal Ordering and Identifiability." In *Studies in Econometric Method. Cowles Commission for Research Ineconomics, Monograph No. 14*, edited by W. C. Hood and T. C. Koopmans, 4974. John Wiley; Sons.

Woodward, James. 2003a. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

———. 2003b. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press. https://doi.org/10.1093/0195155270.001.0001/acprof-9780195155273.