CAUSAL INFERENCE AND STATISTICAL MODELS
IN MODERN SOCIAL SCIENCES

Hans-Peter Blossfeld

1. Introduction

The empirical investigation of causal relationships is an important but difficult scientific endeavor. In the social sciences, two understandings of causation have guided the empirical analysis of causal relationships: (1) “Causation as robust dependence” and (2) “causation as consequential manipulation.” Both approaches clearly have strengths and weaknesses for the social sciences which will be described in detail in this chapter. Based on this discussion, a third understanding of “causation as generative process,” proposed by David Cox, is then further developed. This idea seems to be particularly valuable for modern social sciences because it can easily be combined with a narrative in terms of actor’s objectives, knowledge, reasoning, and decisions (methodological individualism). Using event history models, this approach will then be applied to the causal analysis of an interdependent dynamic social system. In doing so, we first describe parallel processes and time-dependent covariates, the latter of which are often used to include the sample path of parallel processes in transition rate models. The widely used ‘system’ and ‘causal’ approach are contrasted, with the latter proposed as a more appropriate method from an analytical point of view and that it provides straightforward solutions to simultaneity problems, time lags and varying temporal shapes of effects. Based on separate applications in West and East Germany, Canada, Latvia, and the Netherlands, the usefulness of the approach of “causation as generative process” is demonstrated by analyzing two highly interdependent family processes: entry into marriage (for individuals in a consensual union) as the dependent process and first pregnancy/childbirth as the explaining one. After potential statistical reasons for the time-dependent effects are described, we move to more substantive explanations, including the importance of actors, probabilistic causal relations, preferences and negotiation, observed and unobserved decisions and the problem of conditioning on future events.
2. Models of Causal Inference

The goal to find scientifically based evidence for causal relationships leads to design questions, such as which inference model is appropriate to specify the relationship between cause and effect and which statistical procedures can be used to determine the strength of that relationship (Schneider et al. 2007). Two different models of causal inference have dominated the work of practitioners in the social sciences over the last decades: (1) “Causation as robust dependence” and (2) “causation as consequential manipulation.” The former approach—which in multiple regression or path analysis is known as the “control variable” or “parti ll ing” approach (Duncan 1966; Kerlinger and Pedhazer 1973; Blalock 1970) and in the econometric analysis of time-series as “Granger causation” (Granger 1969; Johnston 1972)—starts from the presumption that correlation does not necessarily imply causation but causation must in some way or the other imply correlation. In this view, the key problem of causal inference is to determine whether an observed correlation of variable X with variable Y, where X is temporally prior to Y, can be established as a “genuine causal relationship.”

The advocates of the “causation as robust dependence” approach call X a “genuine” cause of Y in so far as the dependence of Y on X cannot be eliminated through additional variables being introduced into the statistical analysis. Thus, in this approach causation is established essentially through the elimination of spurious (or non-causal) influences. Although this approach has dominated the social sciences for several decades, sociologists consider it as a too limited approach. First, they think that causal inference should not be limited entirely to a matter of statistical predictability but should include predictability in accordance with theory (Goldthorpe 2001: 3). Second, since scientists rarely know all of the causes of observed effects or how they relate to one another, it is impossible to be sure that all other important variables have in fact been controlled for (Shadish, Cook and Campbell 2002). A variable X can therefore never be regarded as having causal significance for Y in anything more than a provisional sense: “At any point, further information might be produced that would show that the dependence of Y on X is not robust after all or, in other words, that the apparent causal force of X is, at least to some extent, spurious.” (Goldthorpe 2001: 5)

The second understanding of “causation as consequential manipulation” seems to have emerged as a reaction to the limitations of “cau-