

**Freedman, David**

**From association to causation via regression.** (English) Zbl 0873.90019  
*Adv. Appl. Math.* 18, No. 1, 59-110 (1997).

Summary: For nearly a century, investigators in the social sciences have used regression models to deduce cause-and-effect relationships from patterns of association. Path models and automated search procedures are more recent developments. In my view, this enterprise has not been successful. The models tend to neglect the difficulties in establishing causal relations, and the mathematical complexities tend to obscure rather than clarify the assumptions on which the analysis is based. Formal statistical inference is, by its nature, conditional. If maintained hypotheses  $A, B, C \dots$  hold, then  $H$  can be tested against the data. However, if  $A, B, C, \dots$ , remain in doubt, so must inferences about  $H$ . Careful scrutiny of maintained hypotheses should therefore be a critical part of empirical work – a principle honored more often in the breach than the observance. This paper focuses on modeling techniques that seem to convert association into causation. The object is to clarify the differences among the various uses of regression, as well as the source of the difficulty in making causal inferences by modeling.

**MSC:**

[91B84](#) Economic time series analysis  
[62P20](#) Applications of statistics to economics  
[91B82](#) Statistical methods; economic indices and measures

Cited in 7 Documents

**Keywords:**

[econometrics](#); [regression](#)

**Software:**

[TETRAD](#)

**Full Text:** [DOI](#)

**References:**

- [1] Bartels, L.M., Instrumental and “quasi-instrumental” variables, *Amer. J. Pol. sci.*, 35, 777-800, (1991)
- [2] Bartels, L.M.; Brady, H.E., The state of quantitative political methodology, *Political science: the state of the discipline II*, (1993), Amer. Pol. Sci. Assoc Washington
- [3] Blau, P.M.; Duncan, O.D., *The American occupational structure*, (1967), Wiley New York
- [4] Cairns, J., *Cancer: science and society*, (1978), Freeman San Francisco
- [5] Cartwright, N., *Nature’s capacities and their measurement*, (1989), Clarendon Press Oxford
- [6] C. C. Clogg, A. Haritou, 1994, *The Regression Method of Causal Inference and a Dilemma with This Method*, Department of Sociology, Pennsylvania State University
- [7] Cornfield, J.; Haenszel, W.; Hammond, E.C.; Lilienfeld, A.M.; Shimkin, M.B.; Wynder, E.L., Smoking and lung cancer: recent evidence and a discussion of some questions, *J. nat. cancer inst.*, 22, 173-203, (1959)
- [8] Cox, D.R.; Wermuth, N., Linear dependencies represented by chain graphs, *Statist. sci.*, 8, 204-283, (1993) · [Zbl 0955.62593](#)
- [9] R. Daggett, D. Freedman, 1985, *Econometrics and the law: A case study in the proof of antitrust damages*, Proceedings, Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, L. LeCamR. Olshen, I, 126, 175, Wadsworth, Belmont, CA · [Zbl 1391.62292](#)
- [10] Darroch, J.N.; Lauritzen, S.L.; Speed, T.P., Markov fields and log-linear interaction models for contingency tables, *Ann. statist.*, 8, 522-539, (1980) · [Zbl 0444.62064](#)
- [11] Desrosières, A., *La politique des grands nombres*, (1993), Éditions la Découverte
- [12] Duncan, O.D., *Introduction to structural equation models*, (1975), Academic Press New York
- [13] Ehrenberg, A.S.C.; Bound, J.A., Predictability and prediction, *J. roy. statist. soc. ser. A*, 156, 167-206, (1993)
- [14] Eitelberg, M.J., *Manpower for military occupations*, (1988), Office of the Assistant Secretary of Defense (Force Management and Personnel) Washington

- [15] Engle, R.F.; Hendry, D.F.; Richard, J.F., Exogeneity, *Econometrica*, 51, 277-304, (1983) · [Zbl 0528.62093](#)
- [16] Floderus, B.; Cederlof, R.; Friberg, L., Smoking and mortality: A 21-year follow-up based on the swedish twin registry, *Internat. J. epidemiology*, 17, 332-340, (1988)
- [17] Freedman, D., A note on screening regression equations, *Amer. statist.*, 37, 152-155, (1983)
- [18] Freedman, D., As others see us: A case study in path analysis, *J. educ. statist.*, 12, (1987)
- [19] Freedman, D., *Statistical models and shoe leather*, ()
- [20] Freedman, D.; Lane, D., *Mathematical methods in statistics*, (1981), Norton New York
- [21] Gauss, C.F., *Theoria motus corporum coelestium*, (1809), Perthes and Besser Hamburg
- [22] D. Geiger, 1990, *Graphoids: A Qualitative Framework for Probabilistic Inference*, Department of Computer Science, UCLA
- [23] C. Glymour, 1993, *A review of recent work on the foundations of causal inference*
- [24] Glymour, C.; Scheines, R.; Spirtes, P.; Kelly, K., *Discovering causal structure*, (1987), Academic Press New York
- [25] Hakama, M.; Lehtinen, M.; Knekt, P.; Aromaa, A.; Leinikki, P.; Miettinen, A.; Paavonen, J.; Peto, R.; Teppo, L., Serum antibodies and subsequent cervical neoplasms: A prospective study with 12 years of follow-up, *Amer. J. epidemiology*, 137, 166-170, (1993)
- [26] Hausman, J., Specification tests in econometrics, *Econometrica*, 46, 1251-1271, (1978) · [Zbl 0397.62043](#)
- [27] Hofferth, S.L.; Moore, K.A., Early childbearing and later economic well-being, *Amer. soc. rev.*, 44, 784-815, (1979)
- [28] Holland, P., Statistics and causal inference, *J. amer. statist. assoc.*, 81, 945-960, (1986) · [Zbl 0607.62001](#)
- [29] Holland, P., Causal inference, path analysis, and recursive structural equations models, (), 449-484
- [30] Tobacco smoking, *Monographs on the evaluation of the carcinogenic risk of chemicals to humans*, 38, (1986), IARC Lyon
- [31] Kaprio, J.; Koskenvuo, M., Twins, smoking and mortality: A 12-year prospective study of smoking-discordant twin pairs, *Social sci. med.*, 29, 1083-1089, (1989)
- [32] Keynes, J.M., Professor Tinbergen's method, *Econ. J.*, 49, 558-570, (1939)
- [33] Keynes, J.M., Comment on Tinbergen's response, *Econ. J.*, 50, 154-156, (1940)
- [34] Kiiveri, H.; Speed, T., Structural analysis of multivariate data: A review, ()
- [35] Leamer, E.E., Vector autoregressions for causal inference, ()
- [36] Legendre, A.M., *Nouvelles méthodes pour la détermination des orbites des comètes*, (1805), Courcier
- [37] Liu, T.C., Under-identification, structural estimation, and forecasting, *Econometrica*, 28, 855-865, (1960)
- [38] Lucas, R.E., *Econometric policy evaluation: A critique*, (), 19-64
- [39] Maddala, G.S., *Introduction to econometrics*, (1992), McGraw-Hill New York · [Zbl 0385.62083](#)
- [40] Manski, C.F., Identification problems in the social sciences, (), 1-56
- [41] Meehl, P., *Clinical versus statistical prediction; A theoretical analysis and a review of the evidence*, (1954), University of Minnesota Press Minneapolis
- [42] Moore, K.A.; Hofferth, S.L., Factors affecting early family formation: A path model, *Popul. environ.*, 3, 73-98, (1980)
- [43] Pearl, J., Fusion, propagation and structuring in belief networks, *Artif. intell.*, 29, 241-288, (1986) · [Zbl 0624.68081](#)
- [44] Pearl, J., *Probabilistic reasoning in intelligent systems*, (1988), Morgan Kaufmann San Mateo
- [45] Pearl, J., Comment: graphical models, causality and intervention, *Statist. sci.*, 8, 266-273, (1993)
- [46] J. Pearl, 1994a, *On the Statistical Interpretation of Structural Equations*, Computer Science Department, UCLA
- [47] J. Pearl, 1994b, *On the Identification of Nonparametric Structural Equations*, Computer Science Department, UCLA
- [48] Pearl, J.; Geiger, D.; Verma, T., The logic of influence diagrams, (), 67-87
- [49] J. Pearl, T. Verma, 1991, A theory of inferred causation, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, J. A. Allen R. Fikes E. Sandewall, 441, 452, Morgan Kaufmann, San Mateo, CA · [Zbl 0765.68177](#)
- [50] J. Pearl, N. Wermuth, 1994, When can association graphs admit a causal explanation?, *Proceedings, Fourth International Workshop on Artificial Intelligence and Statistics, 1993*, P. Cheeseman W. Oldford, Artificial Intelligence and Statistics, 141, 150, Springer-Verlag, Berlin · [Zbl 0828.05060](#)
- [51] Peto, R.; zur Hausen, H., *Viral etiology of cervical cancer*, (1986), Cold Spring Harbor Laboratory
- [52] Pratt, J.; Schlaifer, R., On the nature and discovery of structure, *J. amer. statist. assoc.*, 79, 9-21, (1984) · [Zbl 0579.62095](#)
- [53] Pratt, J.; Schlaifer, R., On the interpretation and observation of laws, *J. econ.*, 39, 23-52, (1988)
- [54] Rao, C.R., *Linear statistical inference and its applications*, (1973), Wiley New York · [Zbl 0169.21302](#)
- [55] Rindfuss, R.R.; Bumpass, L.; St. John, C., Education and fertility: implications for the roles women occupy, *Amer. sociol. rev.*, 45, 431-447, (1980)
- [56] Rindfuss, R.R.; Bumpass, L.; St. John, C., Education and the timing of motherhood: disentangling causation, *J. marriage family*, 46, 981-984, (1984)
- [57] Seneta, E., Discussion, *J. educ. statist.*, 12, 198-201, (1987)

- [58] Sherman, K.J.; Daling, J.R.; Chu, J., Genital warts, other sexually transmitted diseases, and vulvar cancer, *Epidemiology*, 2, 257-262, (1991)
- [59] Simon, H., The meaning of causal ordering, (), 65-81
- [60] Smelser, N.J.; Gerstein, D.R., Behavioral and social science: fifty years of discovery, (1986), National Academy Press Washington
- [61] Speed, T.P.; Kiiveri, H.T., Gaussian Markov distributions over finite graphs, *Ann. statist.*, 14, 138-150, (1986) · [Zbl 0589.62033](#)
- [62] Spirtes, P.; Glymour, C.; Scheines, R., Causation, prediction and search, *Lecture notes in statistics*, 81, (1993), Springer-Verlag New York/Berlin
- [63] P. Spirtes, R. Scheines, C. Glymour, C. Meek, 1993, TETRAD II, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA
- [64] Stigler, S., The history of statistics, (1986), Harvard University Press Boston
- [65] Timberlake, M.; Williams, K., Dependence, political exclusion and government repression: some cross national evidence, *Amer. sociol. rev.*, 49, 141-146, (1984)
- [66] Tinbergen, J., Reply to Keynes, *Econ. J.*, 50, 141-154, (1940)
- [67] Verma, T.; Pearl, J., Causal networks: semantics and expressiveness, (), 69-76
- [68] J. R. Welsh, S. K. Kucinkas, L. T. Curran, 1990, Armed Services Vocational Battery (ASVAB): Integrative Review of Validity Studies
- [69] White, H., A heteroskedasticity-consistent estimator and a direct test for heteroskedasticity, *Econometrica*, 48, 817-838, (1980) · [Zbl 0459.62051](#)
- [70] White, H., Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25, (1982) · [Zbl 0478.62088](#)
- [71] Yule, G.U., An investigation into the causes of changes in pauperism in england, chiefly during the last two intercensal decades, *J. roy. statist. soc.*, 62, 249-295, (1989)
- [72] Carmelli, D.; Page, W.F., Twenty-four year mortality in world war II US male veteran twins discordant for cigarette smoking, *International journal of epidemiology*, 25, 554-559, (1996)
- [73] Humphreys, P.; Freedman, D., The grand leap, *Br. J. phi. sci.*, 47, 113-123, (1996)
- [74] Pearl, J., Causal diagrams for empirical research, *Biometrika*, 82, 669-710, (1995) · [Zbl 0860.62045](#)
- [75] Muñoz, N.; Bosch, F.X.; Shah, K.V.; Meheus, A., The epidemiology of human papillomavirus and cervical cancer, (1992), International Agency for Research on Cancer Lyon

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.