CUADERNOS DE ECONOMÍA

75

UNIVERSIDAD
NACIONAL
DE COLOMBIA

CUADERNOS
DE ECONOMÍA

El contenido de los artículos y reseñas publicadas es responsabilidad de los autores y no refleja el punto de vista u opinión de la Escuela de Economía de la Facultad de Ciencias Económicas o de la Universidad Nacional de Colombia.

The content of all published articles and reviews does not reflect the official opinion of the Faculty of Economic Sciences at the School of Economics, or those of the Universidad Nacional de Colombia. Responsibility for the information and views expressed in the articles and reviews lies entirely with the author(s).

TOOLS FOR CAUSAL INFERENCE FROM CROSS-SECTIONAL INNOVATION SURVEYS WITH CONTINUOUS OR DISCRETE VARIABLES: THEORY AND APPLICATIONS

Alex Coad
Dominik Janzing
Paul Nightingale

A. Coad
CENTRUM Católica Graduate Business School (CCGBS), Lima, Perú, and Pontificia Universidad Católica del Perú (PUCP), Lima, Perú. acoad@pucp.edu.pe. Corresponding author.

D. Janzing
Causal Consulting, Rottenburg, Germany.

P. Nightingale
SPRU, University of Sussex, Brighton, UK.

Coad, A., Janzing, D., & Nightingale, P. (2018). Tools for causal inference from cross-sectional innovation surveys with continuous or discrete variables: Theory and applications. *Cuadernos de Economía*, *37*(75), 779-808.

This paper presents a new statistical toolkit by applying three techniques for data-driven causal inference from the machine learning community that are little-known among economists and innovation scholars: a conditional independence-based approach, additive noise models, and non-algorithmic inference by hand. We include three applications to CIS data to investigate public funding schemes for R&D investment, information sources for innovation, and innovation expenditures and firm growth. Preliminary results provide causal interpretations of some previously-observed correlations. Our statistical 'toolkit' could be a useful complement to existing techniques.

Coad, A., Janzing, D., & Nightingale, P. (2018). Herramientas para la inferencia causal de encuestas de innovación de corte transversal con variables continuas o discretas: Teoría y aplicaciones. *Cuadernos de Economía*, *37*(75), 779-808.

Este artículo presenta un nuevo conjunto de herramientas estadísticas al aplicar tres técnicas de inferencia causal basada en datos tomadas de la comunidad del aprendizaje automático (*maching learning*) y que son poco conocidas entre los economistas y los académicos de la innovación: un enfoque condicional basado en la independencia, modelos de ruido aditivo e inferencia no algorítmica a mano. Incluimos tres aplicaciones a los datos de la CIS —la encuesta de la comunidad sobre la innovación— para investigar los modelos de financiación pública para inversión en investigación y desarrollo, fuentes de información para la innovación, y gastos de innovación y crecimiento empresarial. Los resultados preliminares proporcionan interpretaciones causales de algunas correlaciones observadas previamente. Nuestro conjunto de herramientas estadísticas podría ser un complemento útil a las técnicas existentes.

Coad, A., Janzing, D., & Nightingale, P. (2018). Outils pour l'inférence causale d'enquêtes d'innovation de bilan transversal avec des variables continues ou discrètes : Théorie et applications. *Cuadernos de Economía*, *37*(75), 779-808.

Cet article présente un nouvel ensemble d'outils statistiques en appliquant trois techniques d'inférence causale basée sur des données prises de la communauté de l'apprentissage automatique (*maching learning*) et qui sont peu connues chez les économistes et les spécialistes de l'innovation : une approche conditionnelle basée sur l'indépendance, des modèles de bruit additif et inférence non algorythmique

manuelle. Nous incluons trois applications aux données de la CIS –l'enquête de la communauté sur l'innovation- pour étudier les modèles de financement public pour l'investissement en recherche et développement, sources d'information pour l'innovation, et dépenses d'innovation et de croissance entrepreneuriale. Les résultats préliminaires fournissent des interprétations causales de certaines corrélations observées antérieurement. Notre ensemble d'outils statistiques pourrait être un complément utile aux techniques existantes.

Mots-clefs: inférence causale, enquêtes d'innovation, apprentissage automatique (*machine learning*), modèles de bruit additif, graphes acycliques dirigés.
JEL: O30, C21.

Coad, A., Janzing, D., & Nightingale, P. (2018). Ferramentas para a inferência causal de pesquisas de inovação de corte transversal com variáveis contínuas ou discretas: teoria e aplicações. *Cuadernos de Economía*, **37(75), 779-808.**

Este artigo apresenta um novo conjunto de ferramentas estatísticas aplicando três técnicas de inferência causal baseadas em dados extraídos da comunidade de aprendizado automático (*maching learning*) e que são pouco conhecidas entre economistas e estudiosos da inovação: uma abordagem condicional baseada na independência, modelos aditivos de ruído e inferência não algorítmica à mão. Incluímos três aplicativos para os dados da CIS — a pesquisa da comunidade sobre inovação — para investigar os modelos de financiamento público para investimento em pesquisa e desenvolvimento, fontes de informação para inovação e gastos com inovação e crescimento de negócios. Os resultados preliminares fornecem interpretações causais de algumas correlações observadas anteriormente. Nosso conjunto de ferramentas estatísticas pode ser um complemento útil para as técnicas existentes.

Palavras-chave: inferência causal, pesquisas sobre inovação, aprendizado automático (*machine learning*), modelos de ruído aditivo, gráficos acíclicos dirigidos.

INTRODUCTION

The design of effective policy recommendations requires an understanding of not only the associations between key variables but also the causal relations governing the interactions of these variables (Spirtes, Glymour, & Scheines, 2000; Pearl, 2009; Peters, Janzing, & Schölkopf, 2017). However, a long-standing problem for innovation scholars is obtaining causal estimates from observational (i.e. non-experimental) datasets (Nichols, 2007; Cassiman & Veugelers, 2002; Heckman, 2010). For a long time, causal inference from cross-sectional surveys has been considered impossible. Nevertheless, advances in statistics and analysis of causality, combined with 'big data' and increases in computational power, have led to dramatic improvements in the ability of researchers to obtain causal estimates from observational datasets.

Hal Varian, Chief Economist at Google and Emeritus Professor at the University of California, Berkeley, commented on the value of machine learning techniques for econometricians:

> My standard advice to graduate students these days is go to the computer science department and take a class in machine learning. There have been very fruitful collaborations between computer scientists and statisticians in the last decade or so, and I expect collaborations between computer scientists and econometricians will also be productive in the future. Hal Varian (2014, p.3).

This paper seeks to transfer knowledge from computer science and machine learning communities into the economics of innovation and firm growth, by offering an accessible introduction to techniques for data-driven causal inference, as well as three applications to innovation survey datasets that are expected to have several implications for innovation policy.

The contribution of this paper is to introduce a variety of techniques (including very recent approaches) for causal inference to the toolbox of econometricians and innovation scholars: a conditional independence-based approach; additive noise models; and non-algorithmic inference by hand. These statistical tools are data-driven, rather than theory-driven, and can be useful alternatives to obtain causal estimates from observational data (i.e. instrumental variables techniques and regression discontinuity design). While several papers have previously introduced the conditional independence-based approach (Tool 1) in economic contexts such as monetary policy, macroeconomic SVAR (Structural Vector Autoregression) models, and corn price dynamics (e.g. Swanson & Granger, 1997; Moneta, 2008; Xu, 2017; see also Kwon & Bessler, 2011 for a survey), nevertheless the conditional independence-based approach has little been used in the context of the economics of innovation. Tool 2, and also Tool 3 (except for LiNGAM: see Moneta, Entner, Hoyer, & Coad, 2013 and Lanne, Meitz, & Saikkonen, 2017), are new to the field of economics. A further contribution is that these new techniques are applied to three con-

texts in the economics of innovation (i.e. funding for innovation, information sources for innovation, and innovation expenditures and firm growth) to obtain several interesting and policy-relevant results.

While most analyses of innovation datasets focus on reporting the statistical associations found in observational data, policy makers need causal evidence in order to understand if their interventions in a complex system of inter-related variables will have the expected outcomes. This paper, therefore, seeks to elucidate the causal relations between innovation variables using recent methodological advances in machine learning. While two recent survey papers in the Journal of Economic Perspectives have highlighted how machine learning techniques can provide interesting results regarding statistical associations (e.g. classification problems, regression trees, random forests, penalized regression, LASSO; see Varian, 2014 and Mullainathan & Spiess, 2017), we show how machine learning techniques offer interesting opportunities for causal inference.[1]

Section 2 presents the three tools, and Section 3 describes our CIS dataset. Section 4 contains the three empirical contexts: funding for innovation, information sources for innovation, and innovation expenditures and firm growth. Section 5 concludes.

METHODOLOGY

The basic assumption relating statistics and causality is Reichenbach's principle (Reichenbach, 1956), which states that every statistical dependence between two observed random variables X and Y indicate at least one of the following three alternatives is true: 1) X influences Y , 2) there is a common cause Z influencing X and Y, or, 3) Y influences X. In the second case, Reichenbach postulated that X and Y are conditionally independent, given Z, i.e., their probability densities satisfy the equation:

$$p(x, y|z) = p(x|z)p(y|z) \qquad (1)$$

for all x, y, z. Henceforth, we will denote this by X independent of Y, given Z.

The fact that all three cases can also occur together is an additional obstacle for causal inference. For this study, we will mostly assume that only one of the cases occurs and try to distinguish between them, subject to this assumption. We are aware of the fact that this oversimplifies many real-life situations. However, even if the cases interfere, one of the three types of causal links may be more significant than the others. It is also more valuable for practical purposes to focus on the main causal relations. After all, statements such as "every variable influences every other variable" are not especially helpful as guidance for future policies.

---

[1] George, Haas, and Pentland (2014) emphasize that big data techniques must move from investigating correlations to investigating causal effects.

Figure 1.
Directed Acyclic Graph.
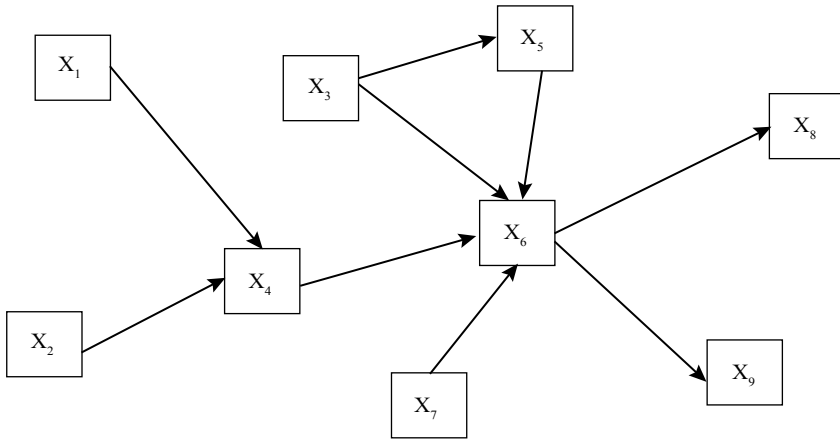
Source: the authors.

Our causal analysis involves the analysis of Directed Acyclic Graphs (or DAGs, see Figure 1). A graphical approach is useful for depicting causal relations between variables (Pearl, 2009). Arrows denote the direction of causality, and we subscribe to a "manipulation view" of causality (Kwon & Bessler, 2011, p.87) according to the (highly hypothetical) scenario whereby an intervention on one variable has an effect on another, while the remaining variables are kept at a fixed value. If we take the example of x_6 in Figure 1, then its 'children' are x_8 and x_9 while its 'parents' are x_3, x_4, x_5, and x_7. x_1 and x_2 have an indirect causal effect on x_6, operating via x_4, but if we control for x_4, then $[x_1, x_2]$ and x_6 are independent: i.e. $p(x_6|x_4, x_1, x_2) = p(x_6|x_4)$. The property that each variable is independent of its non-descendants – conditional on its parents – is known as the causal Markov condition (Spirtes, Glymour, & Scheines, 2000; Pearl, 2009). This condition implies that indirect (distant) causes become irrelevant when the direct (proximate) causes are known.

The density of the joint distribution $p(x_1, x_4, x_6)$, if it exists, can therefore be represented in equation form and factorized as follows:

$$p(x_1, x_4, x_6) = p(x_1).p(x_4|x_1).p(x_6|x_4) \qquad (2)$$

Another important assumption is known as "faithfulness", which allows us to infer dependences from the graph structure. The faithfulness assumption states that only those conditional independences occur that are implied by the graph structure. This implies, for instance, that two variables with a common cause will not be rendered statistically independent by structural parameters that – by chance, perhaps – are fine-tuned to exactly cancel each other out. This is conceptually similar to

the assumption that one object does not perfectly conceal a second object directly behind it that is eclipsed from the line of sight of a viewer located at a specific viewpoint (Pearl, 2009, p.48). In terms of Figure 1, faithfulness requires that the direct effect of x_3 on x_1 is not calibrated to be perfectly cancelled out by the indirect effect of x_3 on x_1 operating via x_5.

In keeping with the DAG perspective on causality, we use an arrow to denote a 'direct' causal influence, but the reader should keep in mind that the distinction between direct and indirect is only meant relative to the set of variables under consideration: 'direct' means that the influence is not mediated by any of the other variables in the DAG. Here we assume that an absolute distinction between 'direct' and 'indirect' influence is meaningless. This perspective is motivated by a physical picture of causality, according to which variables may refer to measurements in space and time: if X_i and X_j are variables measured at different locations, then every influence of X_i on X_j requires a physical signal propagating through space. Thus, we can replace the arrow $X_i \rightarrow X_j$ with an arbitrarily long chain of intermediate variables that refer to measurements along the way as the signal propagates.

Tool 1: Conditional Independence-based approach.

Unconditional independences

Insights into the causal relations between variables can be obtained by examining patterns of unconditional and conditional dependences between variables. For example, although correlation does not equal causation, no causation can be taken to imply no correlation (Kwon & Bessler, 2011, p.90), as implied by Reichenbach's principle.

Bryant, Bessler, and Haigh, (2009) and Kwon and Bessler (2011) show how the use of a third variable C can elucidate the causal relations between variables A and B by using three unconditional independences. Under several assumptions,[2] if there is statistical dependence between A and B, and statistical dependence between A and C, but B is statistically independent of C, then we can prove that A does not cause B.

If X and Y attain one-dimensional numeric values (regardless of whether they are continuous or discrete), they are independent if they are not causally related and thus uncorrelated: corr(X, Y) = 0. In principle, dependences could be only of higher order, i.e., X and Y could be dependent without being correlated, if there is non-linear dependence such as $X^2 + Y^2 = C$. We therefore also use a type of independence test that is able to detect higher-order dependences, namely the

---

[2] Bryant, Bessler, and Haigh, (2009) assume that Reichenbach's principle of common cause holds true. They assume causal faithfulness (i.e. two variables that share a common cause will not appear to be statistically independent by structural parameters that are 'fine-tuned' so as to precisely cancel each other out). They also assume that there are no causal cycles (such as A → B → C → A); however, they do not need to assume that all causally relevant variables are observed.

Hilbert Schmidt Independence Criterion (HSIC) by Gretton, Bousquet, Smola, and Schölkopf (2005) and Gretton, Herbrich, Smola, Bousquet, and Schölkopf (2005). HSIC thus measures dependence of random variables, such as a correlation coefficient, with the difference being that it accounts also for non-linear dependences.

Conditional independences

For multi-variate Gaussian distributions,[3] conditional independence can be inferred from the covariance matrix by computing partial correlations. Instead of using the covariance matrix, we describe the following more intuitive way to obtain partial correlations: let P(X, Y, Z) be Gaussian, then X independent of Y given Z is equivalent to:

$$corr(X - \alpha\, Z, Y - \beta\, Z) = 0 \tag{3}$$

where α and β are the structure coefficients obtained from least square regression when regressing X on Z and Y on Z, respectively. Explicitly, they are given by:

$$\alpha = Cov(X,Z)/Var(Z) \tag{4}$$

$$\beta = Cov(Y,Z)/Var(Z) \tag{5}$$

Note, however, that in non-Gaussian distributions, vanishing of the partial correlation on the left-hand side of (2) is neither necessary nor sufficient for X independent of Y given Z. On the one hand, there could be higher order dependences not detected by the correlations. On the other hand, the influence of Z on X and Y could be non-linear and, in this case, it would not entirely be screened off by a linear regression on Z. This is why using partial correlations instead of independence tests can introduce two types of errors: namely accepting independence even though it does not hold, or rejecting it even though it holds (even in the limit of infinite sample size). Conditional independence testing is a challenging problem, and, therefore, we always trust the results of unconditional tests more than those of conditional tests.

To partly overcome these limitations of conditional independence testing, we also used 'partial HSIC' (we are not aware of any example of it in the literature, but it is a straightforward replacement of partial correlation), that is, performing an HSIC test on the residuals X – α Z, Y – β Z. If their independence is accepted, then X independent of Y given Z necessarily holds. Hence, we have in the infinite sample limit only the risk of rejecting independence although it does hold, while the second type of error, namely accepting conditional independence although it does not hold, is only possible due to finite sampling, but not in the infinite sample limit.

---

[3] A vector-valued variable $(X_1, \ldots, X_d)$ is called multi-variate Gaussian if every linear combination $\Sigma_j\, c_j\, X_j$ is Gaussian distributed.

The conditional independence based approach can infer the causal direction between two variables A and B based on whether a third variable C has specific patterns of (in)dependency with A and B (Kwon & Bessler, 2011). Consider the case of two variables A and B, which are unconditionally independent, and then become dependent once conditioning on a third variable C. The only logical interpretation of such a statistical pattern in terms of causality (given that there are no hidden common causes) would be that C is caused by A and B (i.e. $A \rightarrow C \leftarrow B$, pattern which is known as a 'V-structure' or 'unshielded collider', represented for example by $X_1 \rightarrow X_3 \leftarrow X_2$ in Figure 1). Another illustration of how causal inference can be based on conditional and unconditional independence testing is provided by the example of a Y-structure in Box 1.

The conditional independence-based approach for causal identification seeks to apply logical rules to suggest how observed dependencies between variables should be causally oriented (see Pearl (2009) and Kwon & Bessler (2011) for surveys). The conditional independence-based approach has been used in several economic applications such as macroeconomic dynamics and vector autoregression models (Swanson & Granger, 1997; Demiralp & Hoover, 2003; Perez & Siegler, 2006; Moneta, 2008) as well as the analysis of corn price dynamics (Xu, 2017).

The conditional independence-based approach can help to "reduce the class of admissible causal structures among contemporaneous variables" (Moneta, 2008, p.276) by disproving certain specific causal relations in some cases (Bryant et al., 2009), although a drawback is that often it is not conclusive enough to deliver a unique set of causal orderings between variables (Moneta, 2008; Xu, 2017). Instead, ambiguities may remain and some causal relations will be unresolved. We therefore complement the conditional independence-based approach with other techniques: additive noise models, and non-algorithmic inference by hand. For an overview of these more recent techniques, see Peters, Janzing, and Schölkopf (2017), and also Mooij, Peters, Janzing, Zscheischler, and Schölkopf (2016) for extensive performance studies.

Box 1: Y-structures

Let us consider the following toy example of a pattern of conditional independences that admits inferring a definite causal influence from X on Y, despite possible unobserved common causes (i.e. in the case of Y-structures there is no need to assume causal sufficiency).

If the following four conditions are satisfied:

- Z_1 is independent of Z_2
- Z_1 and Z_2 become dependent when conditioning on X
- $\{Z_1, Z_2\}$ are dependent on Y without conditioning on X
- $\{Z_1, Z_2\}$ are independent of Y when conditioning on X

the figure below on the left ("Y-structure") is an example of a DAG entailing this pattern of conditional (in)dependences. Another example including hidden common causes (the grey nodes) is shown on the right-hand side. Both causal structures, however, coincide regarding the causal relation between X and Y and state that X is causing Y in an unconfounded way. In other words, the statistical dependence between X and Y is entirely due to the influence of X on Y without a hidden common cause, see Mani, Cooper, and Spirtes (2006) and Section 2.6 in Pearl (2009). Similar statements hold when the Y structure occurs as a subgraph of a larger DAG, and Z_1 and Z_2 become independent after conditioning on some additional set of variables. Scanning quadruples of variables in the search for independence patterns from Y-structures can aid causal inference.
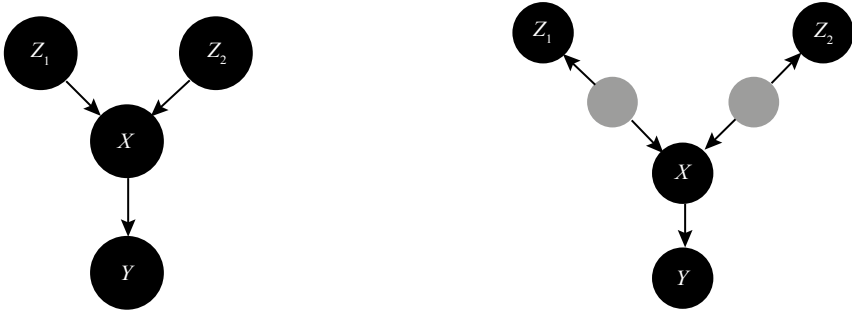
The figure on the left shows the simplest possible Y-structure. On the right, there is a causal structure involving latent variables (these unobserved variables are marked in grey), which entails the same conditional independences on the observed variables as the structure on the left.

Implementation

Since conditional independence testing is a difficult statistical problem, in particular when one conditions on a large number of variables, we focus on a subset of 5-8 variables. We first test all unconditional statistical independences between X and Y for all pairs (X, Y) of variables in this set. Then we test all conditional independences between X and Y, conditional on Z, for all possible triples (X, Y, Z). To avoid serious multi-testing issues and to increase the reliability of every single test, we do not perform tests for independences of the form X independent of Y conditional on $Z_1, Z_2, \ldots Z_n$, with n>1. We then construct an undirected graph where we connect each pair that is neither unconditionally nor conditionally independent. Whenever the number d of variables is larger than 3, it is possible that we obtain too many edges, because independence tests conditioning on more variables could render X and Y independent. We take this risk, however, for the above reasons. In some cases, the pattern of conditional independences also allows the direction of some of the edges to be inferred: whenever the resulting undirected graph contains the pattern X – Z – Y, where X and Y are non-adjacent, and we observe that X and Y are

independent but conditioning on Z renders them dependent, then Z must be the common effect of X and Y (i.e., we have a "v-structure" at Z, denoted as $X \rightarrow Z \leftarrow Y$). For this reason, we perform conditional independence tests also for pairs of variables that have already been verified to be unconditionally independent. From the point of view of constructing the skeleton, i.e., the DAG with undirected edges, the conditional independence tests would be redundant, but for orienting edges the conditional independence tests can be helpful. This argument, like the whole procedure above, assumes causal sufficiency, i.e., the absence of hidden common causes. It is therefore remarkable that the additive noise method below is in principle (under certain admittedly strong assumptions) able to detect the presence of hidden common causes, see Janzing et al. (2009).

Tool 2: Additive Noise Models (ANM)

Our second technique builds on insights that causal inference can exploit statistical information contained in the distribution of the error terms, and it focuses on two variables at a time. Causal inference based on additive noise models (ANM) complements the conditional independence-based approach outlined in the previous section because it can distinguish between possible causal directions between variables that have the same set of conditional independences. With additive noise models, inference proceeds by analysis of the patterns of noise between the variables (or, put differently, the distributions of the residuals).

Figure 2.
For y = f(x) + e, the 'width' of the noise is constant in one direction only, for non-linear f.
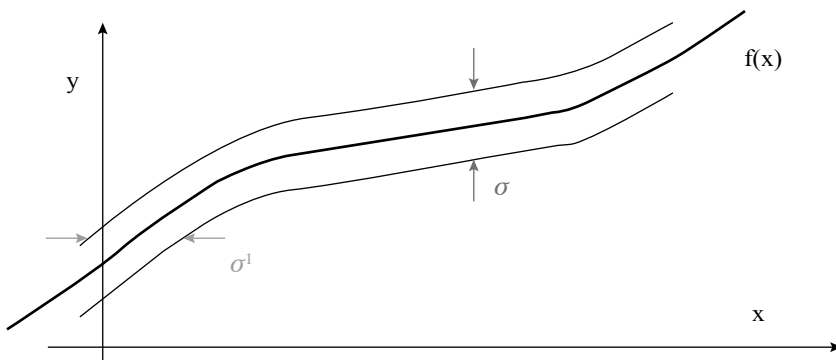
Source: Mooij, Peters, Janzing, Zscheischler, and Schölkopf (2016).

In particular, ANM is able to distinguish between $X \rightarrow Y$ and $Y \rightarrow X$ from the joint distribution $P_{X,Y}$ alone (Hoyer, Janzing, Mooij, Peters, & Schölkopf, 2009). ANMs can also be applied to discrete variables (Peters, Janzing, & Schölkopf, 2011) although at present there is no extensive evaluation of their performance.

Assume Y is a function of X up to an independent and identically distributed (IID) additive noise term that is statistically independent of X, i.e.,

$$Y = f_Y(X) + N_Y$$

where N_Y is independent of X. It can be shown that there is no additive noise model from Y to X in the 'generic case'[4], i.e., there is no function f_X such that,

$$X = f_X(Y) + N_X$$

with N_X independent of Y. Figure 2 visualizes the idea showing that the noise cannot be independent in both directions.

To see a real-world example, Figure 3 shows the first example from a database containing cause-effect variable pairs for which we believe to know the causal direction.[5] Up to some noise, Y is given by a function of X (which is close to linear apart from at low altitudes). Moreover, if we try to describe the altitude as a function of the temperature, the error term is not close to additive, but has a somewhat 'complex' structure, especially in the region of the y-axis corresponding to altitude zero (sea level). Phrased in terms of the language above, writing X as a function of Y yields a residual error term that is highly dependent on Y. On the other hand, writing Y as a function of X yields the noise term that is largely homogeneous along the x-axis. Hence, the noise is almost independent of X. Accordingly, additive noise based causal inference really infers altitude to be the cause of temperature (Mooij et al., 2016), which is certainly true: fixing a thermometer to a balloon would confirm that the temperature changes with the altitude, while heating a place would not change its altitude. Furthermore, this example of altitude causing temperature (rather than vice versa) highlights how, in a thought experiment of a cross-section of paired altitude-temperature datapoints, the causality runs from altitude to temperature even if our cross-section has no information on time lags. Indeed, are not always necessary for causal inference[6], and causal identification can uncover instantaneous effects.

The practical method for inferring causal directions works as follows:

(1) Perform a linear regression of Y on X, that is, find the function f_Y with

$$f_Y(x) := E[Y \,|x].$$

(2) compute the residual variable $N_Y := Y - f_Y(X)$, and

---

[4] The precise meaning of 'generic' here is complicated, see Hoyer, Janzing, Mooij, Peters, and Schölkopf, 2009; Peters, Janzing, and Schölkopf, 2017.

[5] Database with cause effect pairs: https://webdav.tuebingen.mpg.de/cause-effect/. Copyright for variable pairs can be found there.

[6] Granger causality is, under some conditions, also able to uncover instantaneous effects, see Figure 10.8b) and the corresponding explanations on page 207 in Peters, Janzing, and Schölkopf (2017).

(3) test whether N_Y is independent of X. Then do the same exchanging the roles of X and Y. If independence of the residual is accepted for one direction but not the other, the former is inferred to be the causal one. If independence is either accepted or rejected for both directions, nothing can be concluded. If a decision is enforced, one can just take the direction for which the p-value for the independence is larger.

Figure 3.
Scatter plot showing the relation between altitude (X) and temperature (Y) for places in Germany
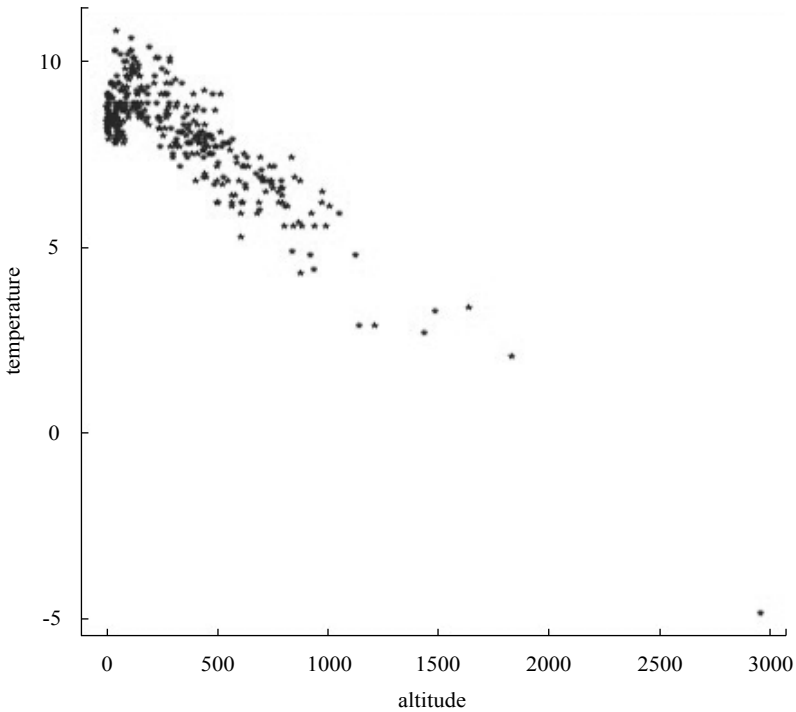
Source: Mooij et al. (2016). Example taken from the database of cause effect pairs at https://webdav.tuebingen.mpg.de/cause-effect/.

This, however, seems to yield performance that is only slightly above chance level (Mooij et al., 2016). Otherwise, setting the right confidence levels for the independence test is a difficult decision for which there is no general recommendation. Conservative decisions can yield rather reliable causal conclusions, as shown by extensive experiments in Mooij et al. (2016). It should be emphasized that additive noise based causal inference does not assume that every causal relation in real-life can be described by an additive noise model. Instead, it assumes that if there is an additive noise model in one direction, this is likely to be the causal one. Hence, causal inference via additive noise models may yield some interesting insights into

causal relations between variables although in many cases the results will probably be inconclusive.

For a justification of the reasoning behind the likely direction of causality in Additive Noise Models, we refer to Janzing and Steudel (2010). The idea is that a joint distribution $P_{X,Y}$ that admits an additive noise model from X to Y is unlikely to be generated by the causal structure $Y \rightarrow X$ because this requires atypical adjustments between P_Y and $P_{X|Y}$. To show this, Janzing and Steudel (2010) derive a differential equation that expresses the second derivative of the logarithm of p(y) in terms of derivatives of log p(x|y). Therefore, for a given conditional $P_{X|Y}$, only very specific choices of P_Y generate an additive noise model from X to Y.

Mooij et al. (2016) provide a recent extensive evaluation of additive noise based inference on real and simulated data. They also make a comparison with other causal inference methods that have been proposed during the past two decades.[7] Additionally, Peters et al. (2011) discuss additive noise models in the context of variables that are not continuous but also discrete. In this paper, we apply ANM-based causal inference only to discrete variables that attain at least four different values.

To our knowledge, the theory of additive noise models has only recently been developed in the machine learning literature (Hoyer et al., 2009; Janzing & Steudel, 2010; Peters et al., 2011, 2017; Mooij et al., 2016) and has not yet been introduced into economics or business research. However, given that these techniques are quite new, and their performance in economic contexts is still not well-known, our results should be seen as preliminary (especially in the case of ANMs on discrete rather than continuous variables).

Further novel techniques for distinguishing cause and effect are being developed. Bloebaum, Janzing, Washio, Shimizu, and Schölkopf (2018), for instance, infer the causal direction simply by comparing the size of the regression errors in least-squares regression and describe conditions under which this is justified. Extensive evaluations, however, are not yet available.

Tool 3: Non-algorithmic inference by hand

The approach introduced in this section is more of a 'meta-method' than a method, which introduces techniques that are not fully automated, but used on a case-by-case, manual basis.

Since the innovation survey data contains both continuous and discrete variables, we would require techniques and software that are able to infer causal directions when one variable is discrete and the other continuous. Unfortunately, there are no off-the-shelf methods available to do this. Sun et al. (2006) and Janzing et al. (2009) propose a method that has been applied to a very limited number of data

---

[7] The real-world data experiments refer to the benchmark data set http://webdav.tuebingen.mpg.de/cause-effect/

sets. In the absence of methods for automated causal discovery, we can try to get hints on the causal direction by using our intuition and arguments that rely on the Principle of Algorithmically Independent Conditionals (Janzing & Schölkopf, 2010; Lemeire & Janzing, 2012). For the special case of a simple bivariate causal relation with cause and effect, it states that the shortest description of the joint distribution $P_{cause,effect}$ is given by separate descriptions of P_{cause} and $P_{effect|cause}$. This implies, in particular, that describing $P_{cause,effect}$ in terms of P_{cause} and $P_{effect|cause}$ is 'simpler' than describing it in terms of P_{effect} and $P_{cause|effect}$[8]. To illustrate this principle, Janzing and Schölkopf (2010) and Lemeire and Janzing (2012) show the two toy examples presented in Figure 4. In both cases we have a joint distribution of the continuous variable Y and the binary variable X. On the left-hand side, P_Y is a mixture of two Gaussians, each of which can be assigned to the cases X = 0 and X = 1, respectively. This joint distribution $P_{X,Y}$ clearly indicates that X causes Y because this naturally explains why P_Y is a mixture of two Gaussians and why each component corresponds to a different value of X. When the same distribution is generated via the causal structure $Y \rightarrow X$ there is, at first, no explanation of why P_Y consists of two modes and, second, no explanation is provided of why each of the Gaussians corresponds to one value of X.[9] Moreover, the distribution on the right-hand side clearly indicates that Y causes X because the value of X is obtained by a simple thresholding mechanism, i.e., $P_{X|Y}$ is a 'machine' receiving continuous input Y and generating the output X = 0 or X = 1, depending on whether Y is above a certain threshold. To generate the same joint distribution of X and Y when X is the cause and Y is the effect involves a quite unusual mechanism for $P_{Y|X}$. Then, $P_{Y|X}$ would be a 'machine' with binary input X whose output is one of the two sides of a truncated Gaussian, depending on the input X.

The examples show that joint distributions of continuous and discrete variables may contain causal information in a particularly obvious manner. There are, however, no algorithms available that employ this kind of information apart from the preliminary tools mentioned above. We therefore rely on human judgements to infer the causal directions in such cases (i.e. human-assisted or "supervised" machine learning, as emphasized in Mullainathan and Spiess, 2017). Below, we will therefore visualize some particular bivariate joint distributions of binaries and continuous variables to get some, although quite limited, information on the causal directions. Although we cannot expect to find joint distributions of binaries and

---

[8] A recent proposal to implement this principle in practice can be found in Budhathoki, Vreeken, and Origo (2018).

[9] To understand the last argument the reader may verify that for two overlapping Gaussians it requires quite sophisticated tuning of the conditional P(X|Y) in order to achieve that both conditional distributions P(Y|X=0) and P(Y|X=1) become Gaussians.

continuous variables (in our real data) for which the causal directions are as obviously as for the cases in Figure 4, we will still try to get some hints.[10]

Figure 4.

Left: visualization of a joint distribution of a binary variable X and a continuous variable Y for which it is reasonably clear that the causal direction reads X → Y. Right: joint distribution for which it is reasonably clear that the causal direction is Y → X.
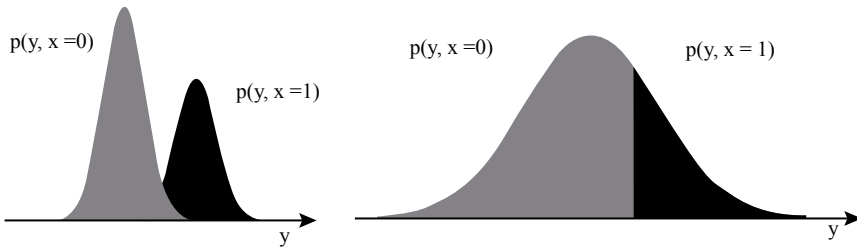
Figures are taken from Janzing and Schölkopf (2010), Janzing et al. (2009), and Lemeire and Janzing (2012).

Finally, another tool that could help causal inference in the case of continuous variables is the Linear Non-Gaussian Acyclic Model (LiNGAM) developed by Shimizu, Hoyer, Hyvarinen, and Kerminen, 2006 (see e.g. Shimizu, 2014 for an overview) and introduced into economics by Moneta et al. (2013) and Lanne et al. (2017). LiNGAM uses statistical information in the (necessarily non-Gaussian) distribution of the residuals to infer the likely direction of causality. LiNGAM analysis was pursued by Xu (2017) to help to orient the DAG's causal relations which had remained unresolved after an initial analysis using the conditional independence-based approach. LiNGAM will be applied 'manually' on a case-by-case basis to obtain further insights into causal relations where possible.

DATA

We analyse data taken from the Community Innovation Surveys (CIS), which are based on the OECD's Oslo Manual, and were administered in several European

---

[10]Although from a different context, the following example of causal relations between a binary and a continuous variable may be of interest. There is an obvious bimodal distribution in data on the relationship between height and sex, with an intuitively obvious causal connection; and there is a similar but much smaller bimodal relationship between sex and body temperature, particularly if there is a population of young women who are taking contraceptives or are pregnant. In contrast, Temperature-dependent sex determination (TSD), observed among reptiles and fish, occurs when the temperatures experienced during embryonic or larval development determine the sex of the offspring. In one instance, therefore, sex causes temperature, and in the other, temperature causes sex, which fits loosely with the two examples (although we do not claim that these gender-temperature distributions closely fit the distributions in Figure 4).

countries to gather information on the innovative activities of firms. The CIS questionnaire can be found online.[11]

CIS data is perhaps the best-known dataset on firm-level innovative activity; it has been extensively analysed and mined by economists and innovation scholars (Mairesse & Mohnen, 2010; Hall & Jaffe, 2012). While previous datasets on firm-level innovation focused on R&D expenditures and patent counts, CIS data has shed valuable light on other aspects of firm-level innovative activity although it also has a number of drawbacks, such as being cross-sectional in nature (thus impeding the investigation of lagged effects, or controlling for time-invariant firm-specific heterogeneity), and also having few variables that can serve as valid instrumental variables.

Mairesse and Mohnen (2010) write:

> "Basically innovation survey data are of a cross-sectional nature, and it is always problematic to address econometric endogeneity issues and make statements about directions of causality with cross-sectional data. ... we have very few exogenous or environmental variables that can serve as relevant and valid instruments." (p.1138)

Moreover, data confidentiality restrictions often prevent CIS data from being matched to other datasets or from matching the same firms across different CIS waves. In addition, at time of writing, the 2008 wave was already rather dated. Finally, another caveat is that many CIS questionnaire responses are evaluated subjectively, and there may be an individual-specific common cause that is correlated across a respondent's questionnaire responses, which could be a further obstacle to causal search.

Given these strengths and limitations, we consider the CIS data to be ideal for our current application, for several reasons:

- It is a very well-known dataset – hence the performance of our analytical tools will be widely appreciated

- It has been extensively analysed in previous work, but our new tools have the potential to provide new results, therefore enhancing our contribution over and above what has previously been reported

- Standard methods for estimating causal effects (e.g. instrumental variables, regression discontinuity design, panel data econometrics) are difficult or impossible to apply

- Most variables are not continuous but categorical or binary, which can be problematic for some estimators but not necessarily for our techniques

- Causal estimates based on CIS data will be valuable for innovation policy

---

[11]See http://ec.europa.eu/eurostat/web/microdata/community-innovation-survey [last accessed June 15, 2017].

To be precise, we focus on the 2008 wave of the CIS, with our raw data covering 16 countries: Bulgaria (BG), Cyprus (CY), Czech Republic (CZ), Germany (DE), Estonia (EE), Spain (ES), Hungary (HU), Ireland (IE), Italy (IT), Lithuania (LT), Latvia (LV), Norway (NO), Portugal (PT), Romania (RO), Slovenia (SI), and Slovakia (SK).

Our data have been deliberately noise-contaminated to anonymise the firms (Mairesse & Mohnen, 2010, p1148; see also Eurostat, 2009). This was done by capping the continuous variables relating to sales and R&D expenditure, and, for the largest values, the true values are not reported, but instead the largest values are approximated. These countries are pooled together to create a pan-European database. This reflects our interest in seeking broad characteristics of the behaviour of innovative firms, rather than focusing on possible local effects in particular countries or regions.

Observations are then randomly sampled. We do not try to have as many observations as possible in our data samples for two reasons. First, due to the computational burden (especially for additive noise models). Second, our analysis is primarily interested in effect sizes rather than statistical significance. We believe that in reality almost every variable pair contains a variable that influences the other (in at least one direction) when arbitrarily weak causal influences are taken into account. However, we are not interested in weak influences that only become statistically significant in sufficiently large sample sizes. Therefore, our data samples contain 2000 observations for our main analysis, and 200 observations for some robustness analysis.[12]

The CIS databases of the sixteen countries differ in terms of number of firms, hence the representativeness of the country's overall economy (in terms of representativeness of firms of different sizes, and firms in manufacturing vs. services sectors, etc.). There is slight variation across countries regarding which questions are asked and the order in which they appear in the questionnaire (Mairesse & Mohnen, 2010). Furthermore, the data does not accurately represent the proportions of innovative vs. non-innovative firms across European countries. We focus on firms with non-zero in-house R&D expenditure. We do not make specific efforts to distinguish between firms in different sectors for two reasons: previous research has emphasized the heterogeneity of innovation patterns within the same sector, and sector of activity has a low explanatory power in explaining firm-level innovation behaviour (Leiponen & Drejer, 2007; Srholec & Verspagen, 2012).[13] In keeping with the previous literature that applies the conditional independence-based approach (e.g. Swanson & Granger, 1997; Xu, 2017) and additive

---

[12]In the machine learning literature, it is not unusual to throw away observations in order to save computational time. Google throws away 99.9% of observations when it does analysis on its own data (see Varian, 2014, p4: "At Google, for example, I have found that random samples on the order of 0.1 percent work fine for analysis of business data.")

[13]Srholec and Verspagen (2012) summarize thus: "[h]eterogeneous, not sectoral or national, is the adjective that should be used to describe patterns of how firms innovate." (p.1247)

noise models (Mooij et al., 2016) and LiNGAM (Moneta et al., 2013), and in contrast to the usual linear regression approach, we do not include control variables in our analysis. This is for several reasons. First, the predominance of unexplained variance can be interpreted as a limit on how much omitted variable bias (OVB) can be reduced by including the available control variables because innovative activity is fundamentally difficult to predict.

Mairesse and Mohnen (2010) found the following:

> "the unexplained residual, that is, the measure of our ignorance in matters of innovation, is larger than the explained part of the share of total sales due to new products, even more in low tech than in high tech sectors." (p.1142)

Second, including control variables can either correct or spoil causal analysis depending on the positioning of these variables along the causal path, since conditioning on common effects generates undesired dependences (Pearl, 2009). Third, in any case, the CIS survey has only a few control variables that are not directly related to innovation (i.e. exporting status, sector and region dummies, and business group affiliation).

For ease of presentation, we do not report long tables of p-values (see instead Janzing, 2016), but report our results as DAGs.

Hence, we are not interested in international comparisons.[14] Nevertheless, we argue that this data is sufficient for our purposes of analysing causal relations between variables relating to innovation and firm growth in a sample of innovative firms.

ANALYSIS

In this section, we present the results that we consider to be the most interesting on theoretical and empirical grounds. The three tools described in Section 2 are used in combination to help to orient the causal arrows. Our results are presented in the form of (partially) Directed Acyclic Graphs (DAGs), following Pearl (2009) and Spirtes et al. (2000). (To be precise, we present partially directed acyclic graphs (PDAGs) because the causal directions are not all identified.) Random variables $X_1 \ldots X_n$ are the nodes, and an arrow from X_i to X_j indicates that interventions on X_i have an effect on X_j (assuming that the remaining variables in the DAG are adjusted to a fixed value). Arrows represent direct causal effects, but note that the distinction between direct and indirect effects depends on the set of variables included in the DAG. Here, we assume that there is no absolute distinction between 'direct' and 'indirect' influence. A line without an arrow represents an undirected relationship – i.e. a statistical association rather than a causal effect – where the direction of causality was not clearly resolved.

---

[14]See Mairesse and Mohnen (2010), "it is heroic to make international comparisons when the questionnaires differ in their content, the order of the questions and their formulations, and when the sampling of respondents differs across countries." (p.1140)

Case 0: sanity check

We begin with a 'sanity check' to verify that our data-driven analysis does not deliver results that are theoretically nonsensical. We investigate the causal relations between two variables where the true causal relationship is already known: i.e. that a firm's sales in 2006 cause a firm's sales in 2008 and not vice-versa. Indeed, the causal arrow is suggested to run from 2006 sales to 2008 sales, which is in line with expectations.[15]

Mooij et al. (2016, Appendix D) provide further sanity checks for simulated data, as well as real-world variable-pairs where the direction of causality is obvious, such as altitude → precipitation; latitude → temperature; age → wage per hour; day of the year → temperature; size of apartment → monthly rent; and age → relative spinal bone mineral density. They conclude that Additive Noise Models (ANM) that use HSIC perform reasonably well, provided that one decides only in cases where an additive noise model fits significantly better in one direction than the other.

Case 1: funding for innovation

A large literature in the economics of innovation has sought to evaluate the effectiveness of public schemes to provide funding for firms' innovative activity and, more specifically, R&D investments. While R&D investing firms are often associated with receipt of funding, the crucial question is whether funding causes R&D investment, or whether R&D investment causes receipt of funding. Standard econometric tools for causal inference, such as instrumental variables, or regression discontinuity design, are often problematic. The empirical literature has applied a variety of techniques to investigate this issue, and the debate rages on. Wallsten (2000) applies a three-stage least squares model and finds that R&D grants totally crowd out firm-financed R&D spending. Aerts and Schmidt (2008) reject the crowding out hypothesis, however, in their analysis of CIS data using both a non-parametric matching estimator and a conditional difference-in-differences estimator with repeated cross-sections (CDiDRCS). Hussinger (2008) finds that public R&D subsidies have a positive effect on treated firms' R&D intensity, using parametric and semiparametric two-step selection models. Howell (2017) applies a sharp regression discontinuity design (RDD) approach and observes that early-stage R&D grants have significant causal effects on firms' outcomes, while the performance of later stage R&D grants is rather disappointing.

Our analysis, in Figure 5, shows that in-house R&D causes EU-level funding, rather than vice versa. This suggests that EU-level funding has no additionality – instead funding is given as windfalls to firms that have already made their R&D investments. In-house R&D, and also total sales, are positively associated with government funding, but there is no evidence that it is funding that improves

---

[15]Details are in Janzing (2016, Section 6.5).

the performance of these firms rather than vice versa. Interestingly, and in line with previous research (see Hashi & Stojcic, 2013, p359, who analyse CIS4 data for sixteen European countries), unlike funding from European or national government sources, funding from regional authorities seems quite disconnected (and perhaps irrelevant) for firm size and innovative activity.

Figure 5.
Partially directed graph resulting from the independence pattern of rrdinx (inhouse R&D), turn08m (turnover in 2008), funeu (EU funding for innovation), fungmt (central government funding for innovation), funloc (local authority funding of innovation).
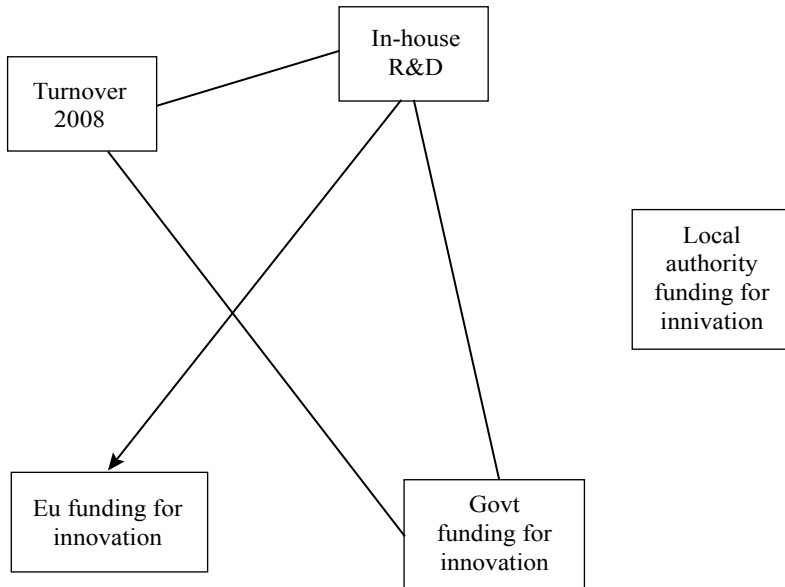
Source: Authors' own analysis.

## Case 2: information sources for innovation

Our second example considers how sources of information relate to firm performance. In the age of open innovation (Chesbrough, 2003), innovative activity is enhanced by drawing on information from diverse sources. However, the relationships between external information sources, R&D investment, and innovation, are complex and not well understood (Laursen & Salter, 2006). Previous research on this issue using CIS data has reported associations but not causal effects (Laursen & Salter, 2006; Vega-Jurado, Gutiérrez-Gracia, & Fernández-de-Lucio, 2009). One policy-relevant example relates to how policy initiatives might seek to encourage firms to join professional industry associations in order to obtain valuable information by networking with other firms. A German initiative requires

firms to join a German Chamber of Commerce (IHK), which provides support and advice to these firms,[16] perhaps with a view to trying to stimulate innovative activities or growth of these firms. However, our results suggest that joining an industry association is an outcome, rather than a causal determinant, of firm performance. Figure 6 shows that having professional and industry associations as a source of information is caused by sales growth, and is positively associated with R&D intensity. This is in contrast with Yam, Lo, Tang, and Lau (2011), who observe a statistical relationship between sources of innovation and R&D capability, and rely on theoretical assumptions to interpret this as evidence that it is the source of information that causes R&D capability.

Conferences, as a source of information, have a causal effect on treating scientific journals or professional associations as information sources.

Figure 6.

Partially directed graph resulting primarily from the independence pattern of rdint (R&D intensity), gr_sales (sales growth), scon (sources of information: conferences, trade fairs, exhibitions), sjou (sources of information: scientific journals and trade/technical publications), spro (professional and industry associations). The edge scon-sjou has been directed via discrete ANM.
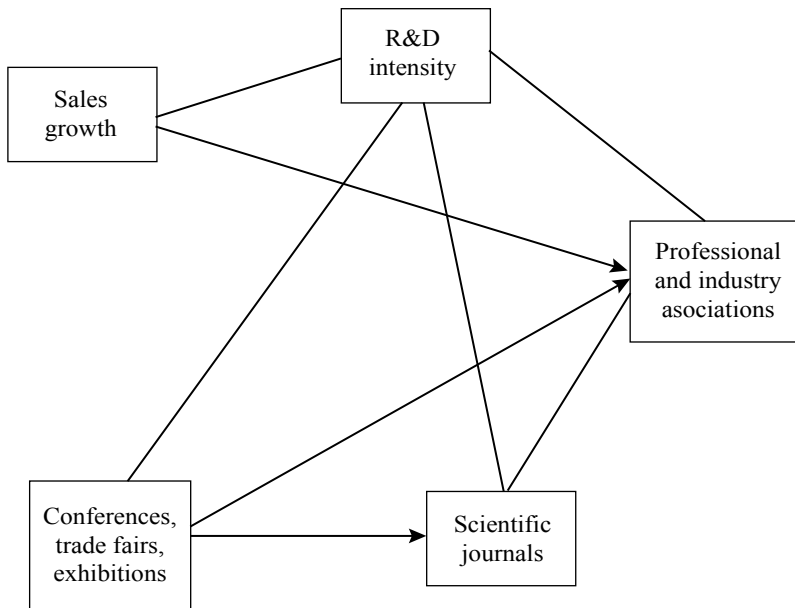
Source: Authors' own analysis.

---

[16]All German companies registered in Germany, with the exception of handicraft businesses, the free professions, and farms, are required by law to join a chamber of commerce. See https://www. dihk.de/en (last accessed June 20th, 2017).

Case 3: innovation expenditures

Although R&D investment is often the first choice of indicator of innovative activity, only a small subset of firms will have positive R&D expenditure, which has led scholars to consider other useful indicators of innovation such as acquisition of machinery/equipment/software, and training (Hall & Jaffe, 2012). In this example, we take a closer look at the different types of innovation expenditure, to investigate how innovative activity might be stimulated more effectively. Previous research has shown that suppliers of machinery, equipment, and software are associated with innovative activity in low- and medium-tech sectors (Heidenreich, 2009). Indeed, acquisition of machinery, equipment, and software plays an important role in firm-level innovation, accounting for between 30% and 90% of innovation expenditures across sectors (see Hughes & Mina, 2012, p5 for UK evidence). However, the Open Innovation paradigm suggests that innovative activity is stimulated by external R&D and external knowledge acquisition (Chesbrough, 2003). The following question therefore arises: should firms be encouraged to acquire external knowledge or machinery/equipment/software? Our results suggest the former. Acquisition of external knowledge has knock-on effects on acquisition of machinery/equipment/software, as well as on training; and training in turn has an impact on acquisition of machinery/equipment/software. Furthermore, external R&D and market introduction of innovations both have causal effects on acquisition of machinery/equipment/software, but this latter has no causal effect on the other variables investigated in this case. Hence, attempts to stimulate expenditures on machinery/equipment/software would not be an effective policy, because these expenditures are stimulated by other innovation expenditures anyway, and because they have no further impacts on other variables.

CONCLUSION

For a long time, causal inference from cross-sectional innovation surveys has been considered impossible. This article introduced a toolkit to innovation scholars by applying techniques from the machine learning community, which includes some recent methods. In particular, three approaches were described and applied: a conditional independence-based approach, additive noise models, and non-algorithmic inference by hand. These techniques were then applied to very well-known data on firm-level innovation: the EU Community Innovation Survey (CIS) data in order to obtain new insights. Three applications are discussed: funding for innovation, information sources for innovation, and innovation expenditures and firm growth. Our results – although preliminary – complement existing findings by offering causal interpretations of previously-observed correlations. Regarding funding for innovation, our results suggest that in-house R&D is a cause, rather than an effect, of receiving EU funding. Regarding information sources, we found that interest in professional & industry associations is caused by sales growth and conferences / trade fairs / exhibitions (and this latter is a cause of interest in scientific

journals). Regarding innovation expenditures, we find a number of results, in particular that acquisition of machinery / equipment / software occurs towards the end of the causal ordering, being causally influenced by several other dimensions of innovation expenditure.

Figure 7.
Partially directed graph resulting from the independence pattern of rrdex (external R&D), rmac (acquisition of machinery, equipment, software), roek (acquisition of external knowledge), rtr (training for innovative activities), rmar (market introduction of innovations), gr_sales. Inference was also undertaken using discrete ANM.
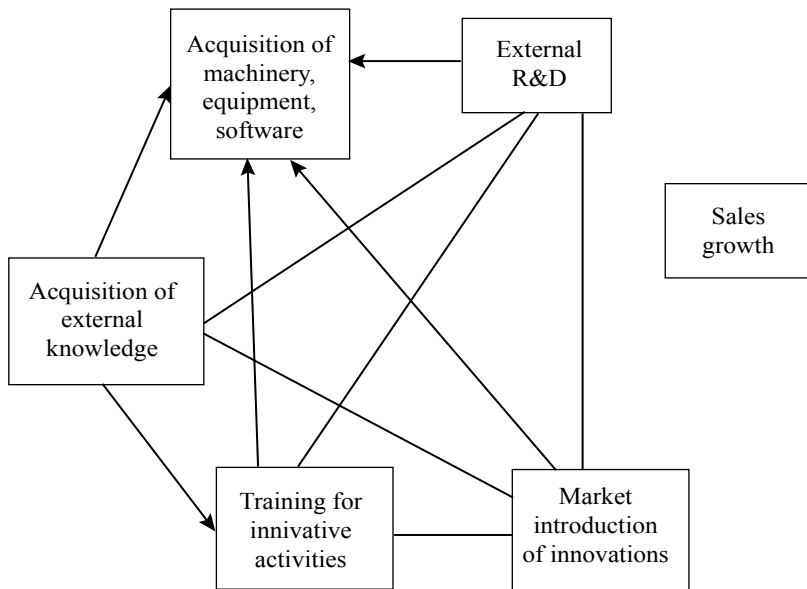
Source: Authors' own analysis.

Future work could extend these techniques from cross-sectional data to panel data. This will presumably be a relatively trivial extension – considering that the most challenging task when identifying a panel regression model such as a structural vector autoregression is first identifying the matrix of instantaneous causal effects in the cross-section (Hyvarinen, Shimizu, & Hoyer, 2008; Moneta et al., 2013). Future work could also investigate which of the three particular tools discussed above works best in which particular context.

Our analysis has a number of limitations, chief among which is that most of our results are not significant. In most cases, it was not possible, given our conservative thresholds for statistical significance, to provide a conclusive estimate of what is causing what (a problem also faced in previous work, e.g. Moneta, 2008; Xu, 2017).

Nevertheless, we maintain that the techniques introduced here are a useful complement to existing research. We consider that even if we only discover one causal relation, our efforts will be worthwhile.[17] Another limitation is that more work needs to be done to validate these techniques (as emphasized also by Mooij et al., 2016), to better understand their reliability. Other limitations, that constitute areas for further research, would be to investigate whether our results are sensitive to our choice of sample (in particular, our focus on R&D investors) or whether our results vary across sectors or across countries.

This paper sought to introduce innovation scholars to an interesting research trajectory regarding data-driven causal inference in cross-sectional survey data. Our aim is to draw attention to these techniques, in the hope that they will be further applied and developed, as another tool in the econometrician´s toolbox. Given the perceived crisis in modern science concerning lack of trust in published research and lack of replicability of research findings, there is a need for a cautious and humble cross-triangulation across research techniques. We hope to contribute to this process, also by being explicit about the fact that inferring causal relations from observational data is extremely challenging. We should in particular emphasize that we have also used methods for which no extensive performance studies exist yet.

REFERENCES

1. Aerts, K., & Schmidt, T. (2008). Two for the price of one?: Additionality effects of R&D subsidies: A comparison between Flanders and Germany. *Research Policy, 37*(5), 806-822.

2. Bryant, H. L., Bessler, D. A., & Haigh, M. S. (2009). Disproving causal relationships using observational data. *Oxford Bulletin of Economics and Statistics, 71*(3), 357-374.

3. Bloebaum, P., Janzing, D., Washio, T., Shimizu, S., & Schölkopf, B. (2018). *Cause-Effect Inference by Comparing Regression Errors*. Presented at AISTATS. For an extended version, see https://arxiv.org/abs/1802.06698.

4. Cassiman B., & Veugelers, R., (2002). R&D cooperation and spillovers: Some empirical evidence from Belgium. *American Economic Review, 92*(4), 1169-1184.

5. Budhathoki, K., & Vreeken, J. O. (2018). Causal inference by compression. *Knowledge and Information Systems, 56*(2), Springer. (IF 2.247).

6. Cattaruzzo, S. (2016). Novel tools for causal inference: A critical application to Spanish innovation studies. Supervisor: Alessio Moneta. University of Pisa/Sant'Anna School of Advanced Studies; Master's Degree Thesis in Economics, November 2016.

---

[17]This idea was expressed long ago by Democritus (460-370 BCE): "I would rather discover one causal law than be King of Persia."

7. Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Cambridge, MA: Harvard Business Press.

8. Demiralp, S., & Hoover, K. (2003). Searching for the causal structure of a vector autoregression. *Oxford Bulletin of Economics and Statistics, 65*, 745-767.

9. Eurostat, (2009). Work Session on Statistical Data Confidentiality, Manchester, 17-19 December 20. Office for Official Publications of the European Communities, Luxembourg, Retrieved April 12th, 2016. http://ec.europa. eu/eurostat/ en/web/products-statistical-working-papers/-/KS-78-09-723

10. George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal, 57*(2), 321-326.

11. Gretton, A., Bousquet, O., Smola, A., & Schölkopf, B. (2005a). Measuring statistical dependence with Hilbert-Schmidt norms. In *Proceedings of the 16th Conference on Algorithmic Learning Theory*, pages 63-77, Berlin: Springer-Verlag.

12. Gretton, A., Herbrich, R., Smola, A., Bousquet, O., & Schölkopf, B. (2005b). Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6, 2075-2129.

13. Hall, B. H., & Jaffe A. B. (2012). *Measuring science, technology, and innovation: A review*. (Report prepared for the Panel on Developing Science, Technology, and Innovation Indicators for the Future, National Academies of Science. May 2012).

14. Hashi, I., & Stojčić, N. (2013). The impact of innovation activities on firm performance using a multi-stage model: Evidence from the Community Innovation Survey 4. *Research Policy, 42*(2), 353-366.

15. Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature, 48*(2), 356–398.

16. Heidenreich, M. (2009). Innovation patterns and location of European low- and medium-technology industries. *Research Policy, 38*(3), 483-494.

17. Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review, 107*(4), 1136-1164.

18. Hoyer, P., Janzing, D., Mooij, J., Peters, J., & Schölkopf, B. (2008). Nonlinear causal discovery with additive noise models. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Proceedings of the conference Neural Information Processing Systems (NIPS) 2008*, Vancouver, Canada: MIT Press.

19. Hughes, A., & Mina, A. (2012). The UK R&D landscape. *UK-IRC Report for the Enhancing Value Task Force*, (March 2012).

20. Hussinger, K. (2008). R&D and subsidies at the firm level: An application of parametric and semiparametric two-step selection models. *Journal of Applied Econometrics*, 23, 729-747.

21. Hyvarinen, A., Shimizu, S., & Hoyer, P. O. (2008). Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-Gaussianity. Presented in *Proceedings of the 25th International Conference on Machine Learning (ICML2008)*, Helsinki, Finland (July 05 - 09, 2008).

22. Janzing, D., Peters, J., Mooij, J., & Schölkopf, B. (2009). *Identifying confounders using additive noise models* (Montreal, Quebec, Canada), in Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (pp. 249-257). Arlington, Virginia, United States: AUAI Press.

23. Janzing, D., Sun, X., & Schölkopf, B. (2006). Causal inference by choosing graphs with most plausible Markov kernels. In *Proceedings of the 9th International Symposium on Artificial Intelligence and Mathematics*, pages 1–11, Fort Lauderdale, FL: Max-Planck-Gesellschaft.

24. Janzing, D., Sun, X., & Schölkopf, B. (2009). Distinguishing cause and effect via second order exponential models. Retreived from http://arxiv.org/abs/0910.5561.

25. Janzing, D., & Schölkopf, B. (2010). Causal inference using the algorithmic Markov condition. *IEEE Transactions on Information Theory*, *56*(10), 5168–5194,

26. Janzing, D. (2016). *Study on: Tools for causal inference from cross-sectional innovation surveys with continuous or discrete variables*. European Commission – Joint Research Center. Available at: http://iri.jrc.ec.europa.eu/research-collaborations.html

27. Janzing, D., & Steudel, B. (2010). Justifying additive-noise-based causal discovery via algorithmic information theory. *Open Systems and Information Dynamics, 17*(2), 189-212.

28. Kwon, D. H., & Bessler, D. A. (2011). Graphical methods, inductive causal inference, and econometrics: A literature review. *Computational Economics, 38*(1), 85-106.

29. Lanne, M., Meitz, M., & Saikkonen, P. (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics, 196*(2), 288-304.

30. Laursen, K., & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal, 27*(2), 131-150.

31. Leiponen A., & Drejer I. (2007). What exactly are technological regimes? Intra-industry heterogeneity in the organization of innovation activities. *Research Policy*, *36*, 1221-1238.

32. Lemeire, J., & Janzing, D. (2013). Replacing causal faithfulness with algorithmic independence of conditionals. *Minds and Machines*, *23*(2), 227-249.

33. Mairesse, J., & Mohnen, P. (2010). Using innovation surveys for econometric analysis. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the Economics of Innovation* (Vol. 2, pp. 1129-1155), Amsterdam: North Holland.

34. Mani S., Cooper, G. F., & Spirtes, P. (2006). A theoretical study of Y structures for causal discovery. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence* (UAI), pp. 314-323.

35. Moneta, A. (2008). Graphical causal models and VARs: An empirical assessment of the real business cycles hypothesis. *Empirical Economics,* 35, 275-300.

36. Moneta, A., Entner, D., Hoyer, P., & Coad, A. (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics, 75*(5), 705-730.

37. Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research, 17*(32), 1-102.

38. Mullainathan S., & Spiess J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106.

39. Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge: Cambridge University Press.

40. Perez, S., & Siegler, M. (2006). Agricultural and monetary shocks before the great depression: A graph-theoretic causal investigation. *Journal of Macroeconomics, 28*(4), 720-736.

41. Peters, J., Janzing, D., & Schölkopf, B. (2011). Causal inference on discrete data using additive noise models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(12), 2436-2450.

42. Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*, Cambridge, MA: MIT press.

43. Reichenbach, H. (1956). The direction of time. Berkeley: University of California Press.

44. Schimel, J. (2012). *Writing science: how to write papers that get cited and proposals that get funded*. Oxford, UK: Oxford University Press.

45. Shimizu, S., Hoyer, P., Hyvarinen, A., & Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, 2003-2030.

46. Shimizu S., (2014). LiNGAM: Non-Gaussian methods for estimating causal structures. *Behaviormetrika, 41*(1), 65-98.

47. Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT press.

48. Srholec, M., & Verspagen, B. (2012). The Voyage of the Beagle into innovation: explorations on heterogeneity, selection, and sectors. *Industrial and Corporate Change, 21*(5): 1221-1253.

49. Swanson, N. R., & Granger, C. W. J. (1997). Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions. *Journal of the American Statistical Association, 92*(437), 357-367.

50. Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3-28.

51. Vega-Jurado, J., Gutiérrez-Gracia, A., & Fernández-de-Lucio, I. (2009). Does external knowledge sourcing matter for innovation? Evidence from the Spanish manufacturing industry. *Industrial and Corporate Change, 18*(4), 637-670.

52. Wallsten, S. J. (2000). The effects of government-industry R&D programs on private R&D: The case of the Small Business Innovation Research program. *Rand Journal of Economics, 31*(1), 82-100.

53. Xu, X. (2017). Contemporaneous causal orderings of US corn cash prices through directed acyclic graphs. *Empirical Economics, 52*(2), 731-758.

54. Yam, R. C., Lo, W., Tang, E. P., & Lau, A. K. (2011). Analysis of sources of innovation, technological innovation capabilities, and performance: An empirical study of Hong Kong manufacturing industries. *Research Policy, 40*(3), 391-402.

CUADERNOS DE ECONOMÍA

75

CONTENIDO