Causal inference and control

Week 7:

Instrumental variables

Søren Wengel Mogensen
Department of Automatic Control
soren.wengel_mogensen@control.lth.se

No exercise class on Wednesday, May 10.

Instrumental variables

Instrumental variable (IV) methods are a bag of similar tricks used in econometrics, causal inference and control theory (and epidemiology, statistics, ...) for *system identification* [Wright, 1928, Reiersøl, 1941, 1945, Sargan, 1958, Joseph et al., 1961, Wong, 1966, Wong and Polak, 1967, Hansen, 1982, Hall, 2005, Pearl, 2009, Thams et al., 2022]. They exploit *moment equations* to obtain consistent estimation of parameters of interest, e.g.,

$$E(g(Y_t, \theta)) = 0$$

where Y_t is observed and θ is unknown.

In the following, we assume random variables to have mean zero. We assume throughout that we have only *observational* data.

Simple IV

Consider an SCM with variables, I, X, Y, H,

$$\begin{aligned} I &= \varepsilon_I \\ H &= \varepsilon_H \\ X &= \alpha I + g_X(H, \varepsilon_X) \\ Y &= \beta X + g_Y(H, \varepsilon_Y), \end{aligned}$$

such that I, H, ε_X, and ε_Y are independent and H is unobserved (the error term of variable V's equation is denoted by ε_V instead of N_V). Note that

$$YI = \beta XI + g_Y(H, \varepsilon_Y)I$$

so if $E(XI) \neq 0$

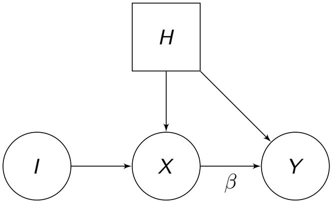
$$\beta = E(YI)/E(XI).$$

A simple IV example and its graph

Consider an SCM with variables, I, X, Y, H,

$$\begin{aligned} I &= \varepsilon_I \\ H &= \varepsilon_H \\ X &= \alpha I + g_X(H, \varepsilon_X) \\ Y &= \beta X + g_Y(H, \varepsilon_Y), \end{aligned}$$

such that I, H, ε_X, and ε_Y are independent and H is unobserved (square).



We have

$$YI = \beta XI + g_Y(H, \varepsilon_Y)I$$

so if $E(XI) \neq 0$

$$\beta = E(YI)/E(XI).$$

Instrumental variable, linear model

Assume $Y = \beta X + g_Y(H, \varepsilon_Y)$.

Definition (Instrumental variable, linear model)

We say that I is an *instrumental variable* if 1) I is independent of $g_Y(H, \varepsilon_Y)$, and 2) $cov(X, I) \neq 0$.

We will mostly consider linear models (linear SCMs), though some terms could be more general. We will say that the models are *essentially linear*.

We will mostly be interested in estimating a parameter β corresponding to $X \rightarrow Y$, i.e., when X is a parent of Y. One can also generalize this to estimate a total causal effect, i.e., when X is an ancestor but not a parent of Y [Pearl, 2009].

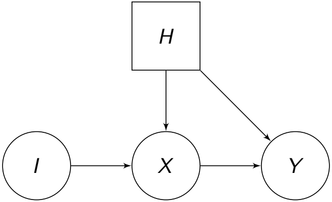
Graphical characterization of an instrumental variable

Definition (Instrumental variable, graphical definition)

We consider a DAG $\mathcal{G}$ which represents a SCM. We say that I is an *instrumental variable relative to the causal effect of X on Y* if 1) $(I \perp\!\!\!\perp_d Y)_{\bar{\mathcal{G}}}$ and 2) $cov(X, I) \neq 0$.

$\bar{\mathcal{G}}$ is the graph obtained by removing the edge $X \to Y$.

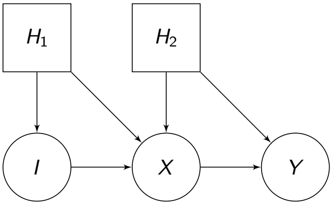
One should note that the classical IV graph (below) is just one example satisfying the graphical condition [Pearl, 2009].

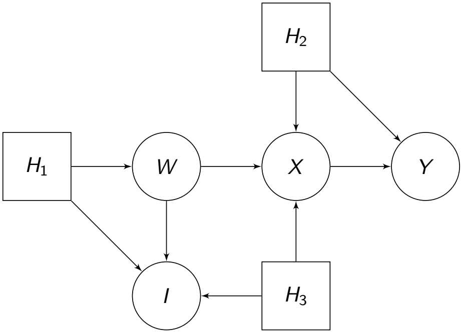


Example, Pearl [2009]

Definition (Instrumental variable, graphical definition)

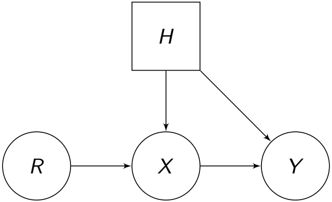
We consider a DAG $\mathcal{G}$ which represents a SCM. We say that I is an *instrumental variable relative to the causal effect of* X *on* Y if 1) $(I \perp\!\!\!\perp_d Y)_{\bar{\mathcal{G}}}$ and 2) $cov(X, I) \neq 0$.





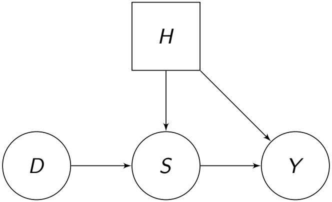
Example, noncompliance

Assume we have data from a randomized controlled trial, comparing a drug with placebo treatment. R is randomization (assigned treatment), X is treatment, Y is outcome, H is are unobserved variables. When R and T can be different, there is *noncompliance*.



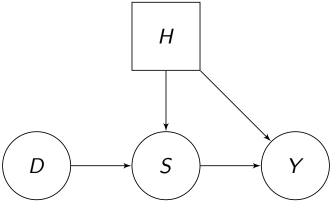
Example, Vietnam war draft [Angrist, 1990]

If we want to estimate the causal effect of military service, S, on lifetime income (or some other outcome), Y, *self-selection* is an issue. During the Vietnam war there was a draft lottery in the US.



Example, Vietnam war draft [Angrist, 1990]

If we want to estimate the causal effect of military service, S, on lifetime income (or some other outcome), Y, *self-selection* is an issue. During the Vietnam war there was a draft lottery in the US.



Does a SCM represented by the above graph seem reasonable?

Two-stage least squares

In linear models, *two-stage least squares* is a common approach to estimation using an instrumental variable. We have

$$\begin{aligned} X &= \alpha I + \gamma H + \varepsilon_X \\ Y &= \beta X + \delta H + \varepsilon_Y \\ &= \beta\alpha I + \beta(\gamma H + \varepsilon_X) + \delta H + \varepsilon_Y \end{aligned}$$

We know that I is independent of H (d-separation) and of ε_X, so we can consistently estimate α using regression of X on I. We can then regress Y on $\hat{\alpha} I$ (compare with $E(YI)/E(XI)$).

Two-stage least squares

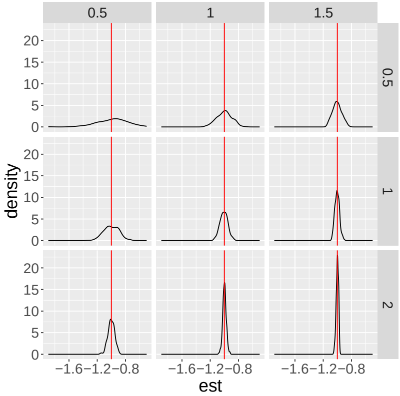
In linear models, *two-stage least squares* is a common approach to estimation using an instrumental variable. We have

$$
\begin{aligned}
X &= \alpha I + \gamma H + \varepsilon_X \\
Y &= \beta X + \delta H + \varepsilon_Y \\
&= \beta\alpha I + \beta(\gamma H + \varepsilon_X) + \delta H + \varepsilon_Y
\end{aligned}
$$

We know that I is independent of H (d-separation) and of ε_X, so we can consistently estimate α using regression of X on I. We can then regress Y on $\hat{\alpha}I$ (compare with $E(YI)/E(XI)$).

If $\alpha = 0$, this will not identify β.

Weak instruments



Densities of two-stage least square-estimates for varying variance of I (horizontal) and varying value of α (vertical). Red line is the true value.

Multivariate setting

The generalization to multivariate X, Y, I is straightforward (again, we assume zero mean). If

$$Y = AX + \varepsilon_Y$$

and I is independent of ε_Y, then

$$E(YI^T) = AE(XI^T).$$

If $E(XI^T)$ has full row rank (its rank equals the number of components in X), then this identifies A.

Multivariate setting

Definition (Instrumental variable, multivariate graphical definition)

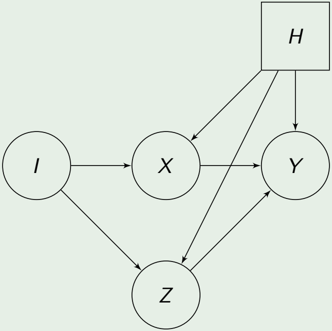
We consider a DAG $\mathcal{G}$ which represents a SCM. We say that I is an *instrumental variable relative to the total causal effect of X on Y* if 1) $(I \perp\!\!\!\perp_d Y)_{\bar{\mathcal{G}}}$ and 2) $cov(X, I)$ has full row rank.

When there are more components in I than in X, then we say there is *overidentification*.

Multivariate setting

It may be that one can find an instrument by including more variables in X [Thams et al., 2022]. When X is a random vector, we let d_X denote its dimension.

Example



$$Y = AX + BZ + CH + \varepsilon_Y$$

I is not (necessarily) independent of Z. However, if we write

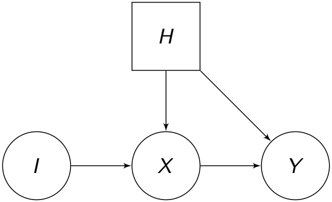
$$Y = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} X \\ Z \end{pmatrix} + CH + \varepsilon_Y$$

and

$$E\left(\begin{pmatrix} X \\ Z \end{pmatrix} I^T \right)$$

has full row rank, then we are back in the familiar setting ($d_I \geq d_X + d_Z$ is necessary).

Model with binary X, Y, I



Consider instead a model with binary I, X, and Y. H is not required to be binary. In this setting, the average causal effect (ACE) is

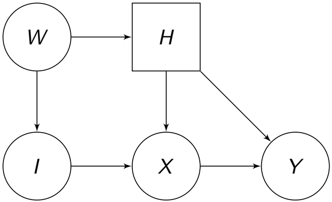
$$p^{do(X=1)}(Y=1) - p^{do(X=0)}(Y=1).$$

This is not identified when H is unobserved without making further assumptions. One can provide bounds for ACE [Balke and Pearl, 1997].

Conditional instrumental variables

Conditional IV

A slightly more involved model, $(I, X, Y, H, W)^T$ (we are again in the essentially linear models).



We have

$$Y = \beta X + g_Y(H, \varepsilon_Y)$$

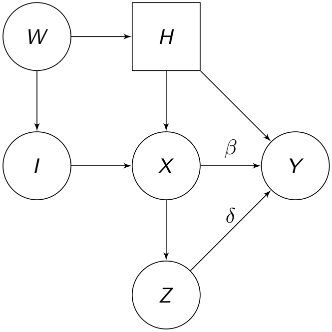
and

$$I = \gamma W + \varepsilon_I.$$

However, I is no longer independent of $g_Y(H, \varepsilon_Y)$. $I - \gamma W = I - E(I|W)$ is though!

Conditional IV, extended

We can add a further complication to the previous model. Consider $(I, X, Y, H, W, Z)^T$.



We have

$$Y = \beta X + \delta Z + g_Y(H, \varepsilon_Y).$$

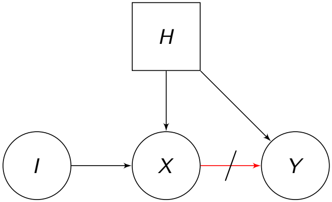
$\tilde{I} = I - E(I|W)$ is in a certain sense still valid,

$$E(Y\tilde{I}) = \beta E(X\tilde{I}) + \delta E(Z\tilde{I}),$$

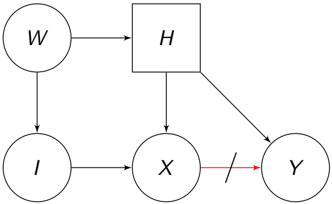
with an additional 'nuisance term'. However, we can estimate δ!

Conditional IVs

We can now give graphical definitions of (conditional) instrumental variables.



I is an instrument.



I is not an instrument.

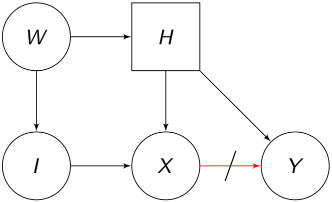
Conditional IVs

We let Y be a real-valued random variable below. We assume $Y = \beta X + \gamma Z + \varepsilon^Y$ where Z is some subset of variables in the SCM.

Definition

I is a *conditional instrumental variable* for $X \to Y$ given a set W (Pearl [2009], Thams et al. [2022]) if 1) I and Y are d-separated given W in $\bar{\mathcal{G}}$, 2) W is not descendants of X or of Y in $\mathcal{G}$, 3) $E(cov(X, I \mid W))$ has full row rank.

We have $cov(X, I \mid W) = E(XI \mid W) - E(X \mid W)E(I \mid W)$. $\bar{\mathcal{G}}$ is the graph where directed edges from X to Y have been removed.



I is not an instrument. However, I is a *conditional* instrument.

Conditional IVs

If conditions 1) and 2) are satisfied then

$$Y - \beta X \perp\!\!\!\perp I \mid W$$

and therefore

$$E(cov(Y - \beta X, I \mid W)) = 0.$$

Again, condition 3) allows us to solve for β.

Instrumental variables in time series models

Time series

We consider a time series $X_t = (X_t^1, \dots, X_t^n)^T$. If

$$X_t = \sum_{k=1}^{p} \Phi_k X_{t-k} + \varepsilon_t$$

then we say that X_t is a VAR(p)-model. We assume that ε_t is a sequence of independent and identically distributed random vectors with independent entries and zero mean.

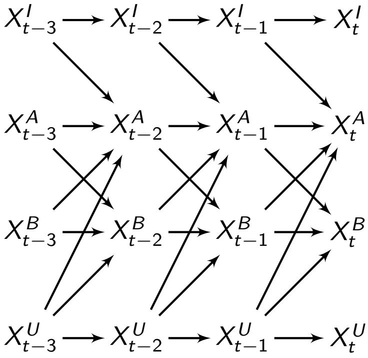
We assume stationarity of the time series.

Graphical representation, example

For simplicity say $p = 1$, $\Phi_1 = \Phi$,

$$X_t = \Phi X_{t-1} + \varepsilon_t$$

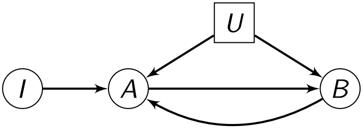
$$\Phi = \begin{bmatrix} \Phi_{II} & 0 & 0 & 0 \\ \Phi_{AI} & \Phi_{AA} & \Phi_{AB} & \Phi_{AU} \\ 0 & \Phi_{BA} & \Phi_{BB} & \Phi_{BU} \\ 0 & 0 & 0 & \Phi_{UU} \end{bmatrix}$$

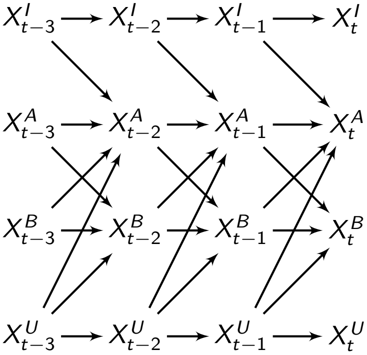


Graphical representation, example

Let $V = \{I, A, B, U\}$ in this example.

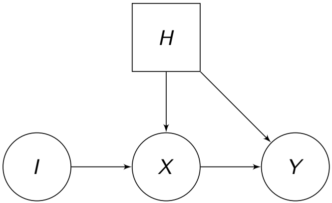
We can also use an alternative (and more compact) graph such that $Y \to Z$ if and only if there exist s, t such that $X_s^Y \to X_t^Z$ in the previous representation ($Y, Z \in V$). Self-edges are omitted.

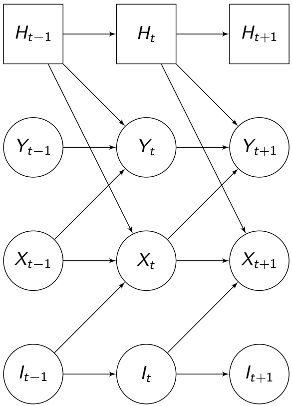




Time series, simple IV

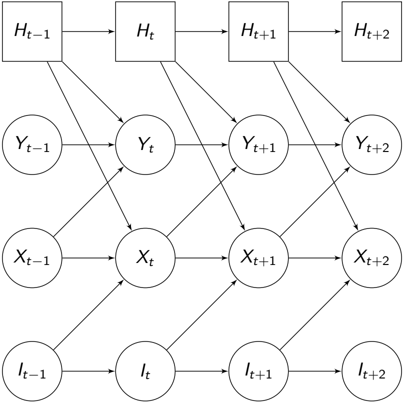
IV methods use the sparsity of the system to provide consistent estimators. This is of course also possible in time series models.





Time series, simple IV

The first thing to note is that I_t is no longer an instrumental variable for $X_{t+1} \to Y_{t+2}$. However, $(I_{t-1}, I_t)^T$ is an instrument for $(X_{t+1}, Y_{t+1})^T \to Y_{t+2}$ (if the rank condition holds).



Policy instruments

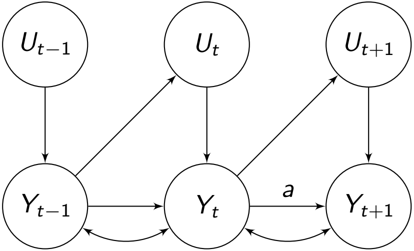
In the instrumental variable example that we already looked at the central assumption is the existence of an observed, *exogenous* random variable which influences the outcome, Y, only through the treatment X.

In time series, a *policy process* is a process in which each variable has an observed parent set. This occurs in applications, when we are monitoring a process and providing feedback, e.g., in economics or control theory.

Controlled time series (correlated noise)

We may think of some processes as *control input*. Let us identify $Y_t \xrightarrow{a} Y_{t+1}$. Look at

$$Y_{t+1} = aY_t + cU_{t+1} + \varepsilon^Y_{t+1} = aY_t + cU_{t+1} + \phi^Y_t + \phi^Y_{t+1}$$

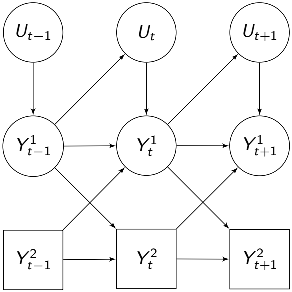


U_t is a valid instrument (uncorrelated with ε^Y_t), allowing for the 'nuisance term' cU_{t+1}. If the correlated noise has higher order (i.e., $\varepsilon^Y_t = \sum_{i=0}^k \phi^Y_{t-i}$, $k > 1$), we may use a U further in the past.

Controlled time series (partial observation)

Let us identify $Y_t^1 \xrightarrow{a} Y_{t+1}^1$. Look at

$$Y_{t+1}^1 = aY_t^1 + bY_t^2 + cU_{t+1} + \varepsilon_t^{Y^1}$$



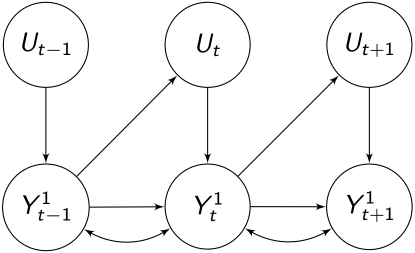
We can use a conditional IV again, $U_t - E(U_t | Y_{t-1}^1)$ (there is a 'nuisance term' [Thams et al., 2022]).

In large-scale systems, this allows *distributed* system identification.

Quality of an instrument

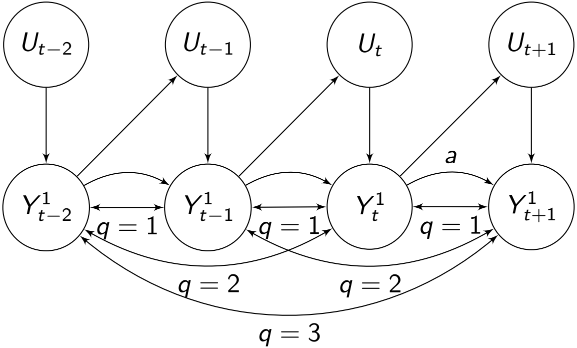
We need an instrument which is strongly correlated with X when estimating $X \xrightarrow{a} Y$. Back to this example,

$$Y_{t+1} = aY_t^1 + cU_{t+1} + \varepsilon_t^Y.$$

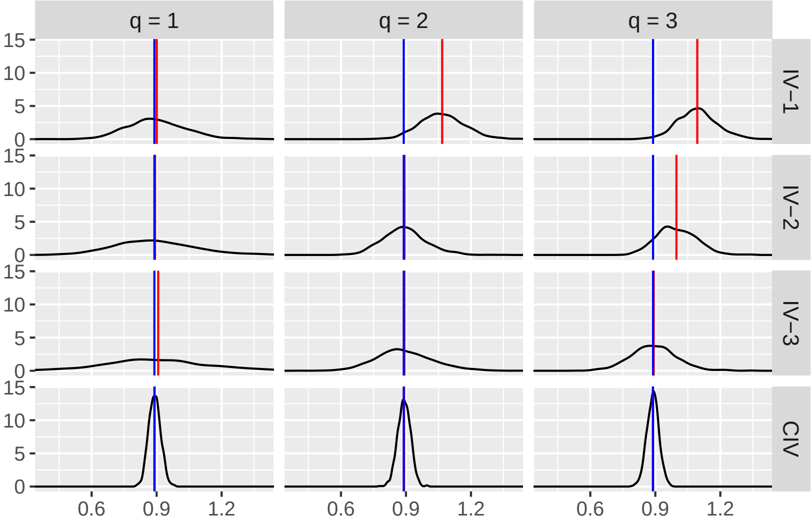


U_{t+1-k} is a valid instrument ($\varepsilon_t^Y = \sum_{i=0}^k \phi_{t-i}^Y$, $k > 1$) for a. If k is large, $U_t - E(U_t | Y_{t-1}^1)$ may be a better instrument.

Simulations



Simulations



Generalized method of moments

The generalized method of moments Sargan [1958], Hansen [1982] assumes we have an n-dimensional vector Y_t such that

$$E(g(Y_t, \theta)) = 0$$

for known g and unknown θ (often not assuming independence of Y_s and Y_t). We can write

$$E((Y - \beta X)I) = 0$$

to see that the IV methods fit into this framework.

State-space models

Instrumental variable methods have also been studied in *state-space models* [Gustafsson, 2001], here with *errors-in-variables*,

$$
\begin{aligned}
X_t &= AX_{t-1} + B\tilde{U}_{t-1} + \varepsilon_t^X \\
\tilde{Y}_t &= CX_t + D\tilde{U}_t \\
U_t &= \tilde{U}_t + \varepsilon_t^U \\
Y_t &= \tilde{Y}_t + \varepsilon_t^Y.
\end{aligned}
$$

Variables U_t and Y_t are measured. Matrices A, B, C, D are not in general identifiable, however, one can identify an 'equivalence class' of parameters.

Summary

Instrumental variable methods exploit an exogeneity assumption to obtain identification. The assumptions are often represented using graphs.

The precise conditions depend on the model class, however, examples can be found in linear SCMs, binary SCMs [Pearl, 2009], and time series models [Gustafsson, 2001, Thams et al., 2022]. Examples can also be found in continuous-time processes [Mogensen, 2023].

References I

Joshua D Angrist. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *The american economic review*, pages 313–336, 1990.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439): 1171–1176, 1997.

Tony Gustafsson. Subspace identification using instrumental variable techniques. *Automatica*, 37(12):2005–2010, 2001.

Alastair R Hall. *Generalized method of moments*. Oxford University Press, 2005.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

P Joseph, J Lewis, and J Tou. Plant identification in the presence of disturbances and application to digital adaptive systems. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 80(1):18–24, 1961.

References II

Søren Wengel Mogensen. Instrumental processes using integrated covariances. In *Proceedings of the 2nd Conference on Causal Learning and Reasoning (CLeaR)*, 2023.

Judea Pearl. *Causality*. Cambridge University Press, 2009.

Olav Reiersøl. Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9(1):1–24, 1941.

Olav Reiersøl. *Confluence analysis by means of instrumental sets of variables*. PhD thesis, Stockholms Högskola, 1945.

John D Sargan. The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415, 1958.

Nikolaj Thams, Rikke Søndergaard, Sebastian Weichwald, and Jonas Peters. Identifying causal effects using instrumental time series: Nuisance IV and correcting for the past. *arXiv:2203.06056*, 2022.

K Y Wong. *Estimation of parameters of linear systems using the instrumental variable method*. PhD thesis, University of California, Berkeley, 1966.

References III

Kwan Wong and Elijah Polak. Identification of linear discrete time systems using the instrumental variable method. *IEEE Transactions on Automatic Control*, 12 (6):707–718, 1967.

Philip G Wright. *Tariff on animal and vegetable oils*. Macmillan Company, New York, 1928.