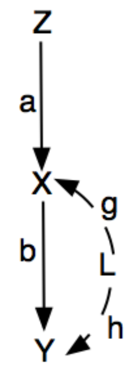
Bayesian control function linear IV: Testing for confounding and weak instruments

A linear structural causal model (SCM) as depicted in the DAG below is plagued by an omitted, correlated variable problem.



An instrumental variable strategy identifies the causal effect of interest and may effectively estimate the parameter so long as the confounding is not too severe and the instruments are not too weak. However, gauging the severity of such problems can be challenging. We propose a control function linear-IV (CF-IV) strategy and conduct simulations to assess its relative effectiveness by comparing results with Bayesian linear regression (OLS), Bayesian IV in which the first stage regression is sampled (dynamic-IV), and Bayesian IV in which the first stage regression is fixed (static-IV).

The model is as depicted in the DAG above. That is,

$$\begin{array}{cc} Z = U_z, & U_z \sim N(0,1) \\ X = aZ + gU + U_x, & U_x \sim N(0,1) \\ Y = bX + hU + U_y, & U_y \sim N(0,1) \\ a = 1, b = 1 & U \sim N(0,1) \\ g = \{1, 5, 10, 50\} & h = \{-50, -10, 10, 50\} \end{array}$$

The values of g and h are the experimental values where the weakness of the instrument is increasing in g and the degree of confounding is increasing in the absolute value of h, holding other things equal.

The OLS estimand the regression coefficient of Y onto X, r_{YX}, is $\frac{2+g^2+gh}{2+g^2}$ while the linear-IV estimand is $\frac{r_{YZ}}{r_{XZ}} = \frac{ab}{a} = b = 1$. Hence, OLS bias is determined by $\frac{gh}{2+g^2}$, for instance, if $h = 0$ the bias is zero. When h is negative, we have a Simpson's reversal paradox frame, that is, the OLS estimate is negative while the parameter of interest is $b = 1$. Also, since projecting X into Z results in loss of information, linear-IV is less efficient than OLS (produces wider credible intervals in the Bayesian frame).

1 Control function-IV strategy

An alternative linear-IV strategy derives from a control function strategy. Regress Y onto X and u_x, where u_x is the residuals from the first stage regression of X onto Z. The estimand for the causal effect of X on Y is $b = 1$. The idea is as follows. The first stage regression gives

$$u_x = M_z X = (I - P_z) X$$

where

$$P_z = Z \left(Z^T Z \right)^{-1} Z^T$$

the projection matrix in the columns of Z. The second stage regression, by double residual regression, involves recovering the residuals from conditioning X on u_x

$$\begin{aligned} res_x &= X - P_{u_x} X \\ &= X - M_z X \left(X^T M_z X \right)^{-1} X^T M_z X \\ &= X - M_z X \\ &= P_z X \end{aligned}$$

where $P_{u_x} = M_z X \left(X^T M_z X \right)^{-1} X^T M_z$. Since $P_z X$ is precisely the quantity employed by standard linear-IV, control function linear-IV produces the same estimand, $b = 1$. In addition, the coefficient on u_x is $\frac{gh}{1+g^2}$. In other words, it's an indicator of the extent of confounding.

2 Simulation

We report 20 simulations of $1,000$ posterior draws ($20,000$ draws) for each pair (g, h) for each strategy based on samples of size $n = 1,000$. Priors are diffuse (zero means and null information matrices). The simulations compare the performance of the control function IV (CF-IV) strategy with dynamic-IV (first stage sampled like the CF-IV strategy), static-IV (first stage fixed), and OLS (no instruments). In addition, to exploring the efficacy of the control function-IV in reducing bias and gauging the extent of confounding, dynamic sampling of the first stage regression may help to mitigate the bias associated with weak instruments (large values of g). Let d_{CF-IV} be the coefficient on u_x for the control function linear-IV strategy, b is the target estimand, and a is the first stage estimand (first-stage simulation results reported only for the sampled strategies, the same sample is employed for the control function and dynamic IV strategies).

First, we report results when the instrument is reasonably strong, $g = 1$, but confounding is high, $h = 10, -10$.

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	5.004	0.946	0.956	0.952	4.307	0.992
std. dev.	0.473	0.443	0.446	0.442	0.173	0.067
quantile:						
1%	4.052	−0.238	−0.200	−0.210	3.905	0.848
2.5%	4.178	−0.030	0.013	0.009	3.969	0.869
5%	4.294	0.170	0.188	0.193	4.021	0.888
10%	4.437	0.370	0.380	0.381	4.081	0.910
25%	4.676	0.673	0.674	0.671	4.190	0.949
50%	4.967	0.977	0.973	0.974	4.307	0.990
75%	5.291	1.255	1.263	1.261	4.424	1.031
90%	5.624	1.487	1.505	1.498	4.530	1.075
95%	5.844	1.622	1.653	1.635	4.590	1.109
97.5%	6.057	1.729	1.786	1.756	4.644	1.144
99%	6.309	1.849	1.918	1.884	4.710	1.181

Posterior simulation for $a = 1, b = 1, d = \frac{10}{2}, g = 1, h = 10, bias = \frac{10}{3}$

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	−4.966	0.979	0.969	0.968	−2.340	0.997
std. dev.	0.492	0.442	0.441	0.436	0.181	0.061
quantile:						
1%	−6.136	0.041	0.013	0.012	−2.762	0.867
2.5%	−5.942	0.162	0.154	0.142	−2.695	0.886
5%	−5.788	0.271	0.261	0.261	−2.642	0.903
10%	−5.613	0.409	0.399	0.404	−2.575	0.923
25%	−5.304	0.665	0.655	0.658	−2.463	0.955
50%	−4.955	0.971	0.965	0.966	−2.338	0.994
75%	−4.615	1.282	1.270	1.272	−2.214	1.035
90%	−4.329	1.558	1.540	1.533	−2.109	1.077
95%	−4.178	1.713	1.707	1.689	−2.046	1.105
97.5%	−4.059	1.855	1.843	1.824	−1.991	1.132
99%	−3.919	2.040	2.014	1.973	−1.927	1.157

Posterior simulation for $a = 1, b = 1, d = \frac{-10}{2}, g = 1, h = -10, bias = \frac{-10}{3}$

With high levels of confounding but a strong instrument, these results are as expected. All three instrumental variable strategies produce credible intervals around the true parameter value, the control function IV also effectively predicts the extent of confounding and OLS is biased as predicted. Further, the credible intervals for OLS are much tighter than those for IV but the bias is sufficiently severe that the interval fails to even approach parameter b and as suggested earlier the sign is even wrong when $h = -10$.

Next, we explore efficacy of these strategies with a weaker instrument, $g = 5$

but h remains $10, -10$.

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	2.103	0.819	0.918	0.890	2.854	0.956
std. dev.	0.505	0.502	0.480	0.442	0.025	0.195
quantile:						
1%	1.320	−0.806	−0.247	−0.260	2.792	0.522
2.5%	1.399	−0.385	−0.055	−0.068	2.801	0.591
5%	1.4723	−0.097	0.093	0.098	2.810	0.643
10%	1.565	0.178	0.284	0.287	2.821	0.707
25%	1.754	0.569	0.598	0.287	2.838	0.821
50%	2.011	0.912	0.943	0.935	2.856	0.954
75%	2.355	1.167	1.250	1.213	2.872	1.088
90%	2.742	1.352	1.508	1.422	2.886	1.210
95%	3.023	1.446	1.666	1.535	2.894	1.282
97.5%	3.307	1.515	1.799	1.634	2.902	1.346
99%	3.728	1.590	1.972	1.737	2.911	1.422

Posterior simulation for $a = 1, b = 1, d = \frac{50}{26}, g = 5, h = 10, bias = \frac{50}{27}$

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	−2.063	1.1370	1.102	1.069	−0.857	0.986
std. dev.	0.594	0.592	0.629	0.506	0.025	0.236
quantile:						
1%	−4.053	0.268	0.157	0.156	−0.924	0.470
2.5%	−3.486	0.341	0.262	0.261	−0.914	0.547
5%	−3.127	0.414	0.350	0.357	−0.904	0.611
10%	−2.771	0.521	0.463	0.471	−0.890	0.689
25%	−2.328	0.741	0.674	0.692	−0.872	0.819
50%	−1.961	1.036	0.988	1.010	−0.855	0.976
75%	−1.665	1.402	1.391	1.378	−0.840	1.144
90%	−1.445	1.842	1.865	1.754	−0.827	1.300
95%	−1.341	2.195	2.238	2.005	−0.820	1.390
97.5%	−1.261	2.556	2.591	2.211	−0.813	1.460
99%	−1.195	3.135	3.196	2.456	−0.806	1.541

Posterior simulation for $a = 1, b = 1, d = \frac{-50}{26}, g = 5, h = -10, bias = \frac{-50}{27}$

Even though the instrument is weaker than the above cases, the IV strategies effectively addresses the confounding while OLS is substantially biased. Also, d_{CF-IV} from the control function-IV is an effective gauge of confounding.

Next, we explore efficacy of these strategies with an even weaker instrument,

$g = 10$ but h remains $10, -10$.

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	1.652	0.338	143.0	0.913	1.980	1.067
std. dev.	43.92	43.92	17100.	0.717	0.008	0.449
quantile:						
1%	0.428	−6.035	−2.738	−2.070	1.962	0.027
2.5%	0.502	−2.326	−0.887	−0.914	1.965	0.180
5%	0.552	−0.884	−0.313	−0.349	1.967	0.318
10%	0.610	−0.072	0.166	0.159	1.970	0.475
25%	0.732	0.665	0.710	0.735	1.974	0.761
50%	0.942	1.047	1.100	1.090	1.979	1.078
75%	1.327	1.258	1.441	1.306	1.985	1.382
90%	2.061	1.381	1.794	1.462	1.990	1.644
95%	2.875	1.438	2.057	1.560	1.993	1.792
97.5%	4.317	1.487	2.540	1.673	1.995	1.917
99%	8.033	1.562	4.323	1.903	1.998	2.063

Posterior simulation for $a = 1, b = 1, d = \frac{100}{101}, g = 10, h = 10, bias = \frac{100}{102}$

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	−1.257	1.267	−249.7	1.118	0.020	0.983
std. dev.	27.99	27.99	32163.	0.660	0.007	0.411
quantile:						
1%	−8.103	0.428	−5.481	0.131	0.004	0.020
2.5%	−4.311	0.518	−0.070	0.330	0.006	0.171
5%	−2.942	0.579	0.358	0.442	0.008	0.309
10%	−2.101	0.645	0.481	0.548	0.011	0.457
25%	−1.397	0.784	0.664	0.725	0.015	0.704
50%	−0.999	1.010	0.932	0.978	0.020	0.987
75%	−0.773	1.409	1.402	1.329	0.025	1.262
90%	−0.634	2.113	2.314	1.806	0.029	1.508
95%	−0.569	2.952	3.435	2.314	0.032	1.653
97.5%	−0.509	4.325	5.615	2.968	0.034	1.785
99%	−0.420	8.107	15.24	3.815	0.037	1.948

Posterior simulation for $a = 1, b = 1, d = \frac{-100}{101}, g = 10, h = -10, bias = \frac{-100}{102}$

Weak instruments seems to produce highly variable and biased Dynamic-IV causal estimates. Otherwise, weakening of the instrument produces moderately poorer results. Next, we explore extreme confounding with a relatively strong instrument, $g = 1, h = -50, 50$.

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	25.96	0.178	0.239	0.250	17.69	0.975
std. dev.	2.836	2.489	2.474	2.452	1.061	0.059
quantile:						
1%	19.01	−6.396	−6.103	−6.110	15.36	0.830
2.5%	20.13	−5.214	−4.985	−4.968	15.69	0.853
5%	21.14	−4.160	−3.965	−3.977	15.97	0.874
10%	22.26	−2.984	−2.871	−2.865	16.32	0.899
25%	24.17	−1.316	−1.272	−1.246	16.93	0.938
50%	26.02	0.299	0.310	0.330	17.69	0.977
75%	27.83	1.773	1.805	1.823	18.43	1.015
90%	29.49	3.158	3.280	3.225	19.09	1.049
95%	30.56	4.133	4.187	4.137	19.46	1.070
97.5%	31.47	4.949	5.069	5.069	19.76	1.087
99%	32.64	5.998	6.064	6.038	20.09	1.106

Posterior simulation for $a = 1, b = 1, d = \frac{50}{2}, g = 1, h = 50, bias = \frac{50}{3}$

	d_{CF-IV}	b_{CF-IV}	$b_{dynamic-IV}$	$b_{static-IV}$	b_{OLS}	a
mean	−24.71	0.685	0.646	0.625	−15.93	1.002
std. dev.	2.831	2.699	2.675	2.676	1.163	0.066
quantile:						
1%	−31.12	−5.010	−5.098	−5.058	−18.71	0.854
2.5%	−30.16	−4.320	−4.409	−4.425	18.30	0.875
5%	−29.30	−3.698	−3.698	−3.709	−17.92	0.895
10%	−28.38	−2.821	−2.793	−2.849	−17.43	0.917
25%	−26.74	−1.279	−1.284	−1.330	−16.67	0.955
50%	−24.70	0.628	0.596	0.567	−15.91	1.002
75%	−22.65	2.567	2.543	2.536	−15.18	1.048
90%	−21.02	4.239	4.204	4.189	−14.43	1.088
95%	−20.14	5.168	5.071	5.054	−13.99	1.110
97.5%	−19.38	6.006	5.809	5.766	−13.61	1.129
99%	−18.60	6.921	6.595	6.605	−13.24	1.151

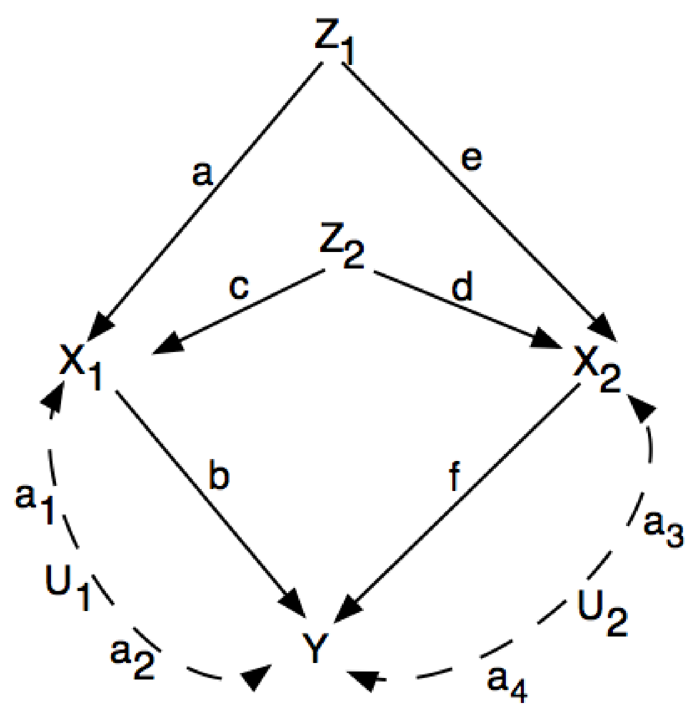
Posterior simulation for $a = 1, b = 1, d = \frac{-50}{2}, g = 1, h = -50, bias = \frac{-50}{3}$

Despite the first-stage indicating a reasonably strong instrument, with extreme levels of confounding, linear-IV does not effectively recover the causal parameter of interest (most any credible interval includes zero with mean and median biased toward zero). However, the control function-IV provides strong evidence of confounding suggesting that OLS (which is bounded substantially away from the true parameter value) is not reliable. Weaker instruments predictably produce further deterioration of results (not reported), however, the control function-IV continues to reliably indicate confounding unless the instrument is extremely weak (for $g = 50$, the first-stage indicates serious weakness and the control function-IV strategy has little diagnostic power as the credible intervals are quite wide where almost all intervals include zero).

All linear-IV strategies perform similarly with differences indistinguishable from sampling variation. This is somewhat surprising as evidence on the strength of the instrument was expected to differ between dynamic and static sampling of the first-stage regression. The first stage credible interval seems to effectively gauge weakness in the instrument. Finally, the control function-IV appears to effectively gauge the level of confounding. An added diagnostic tool for linear models.

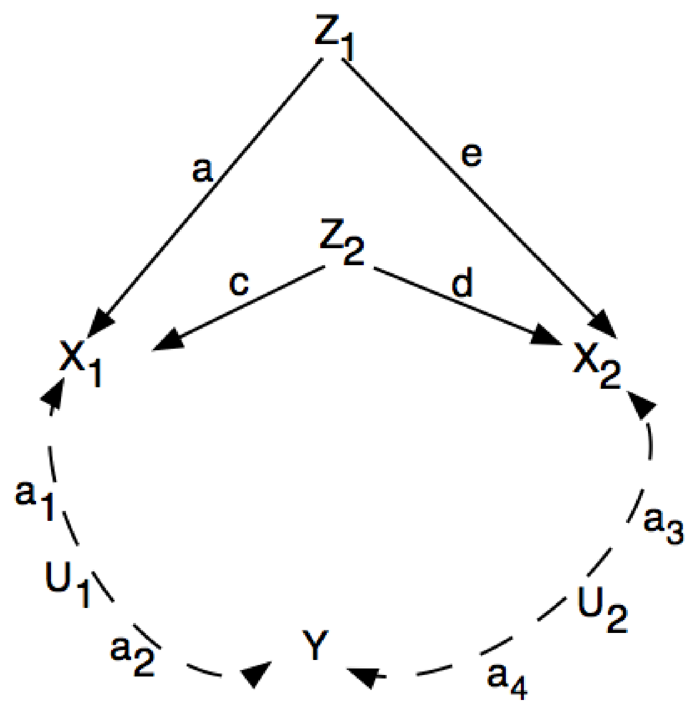
3 Multiple confounded effects

If we wish to explore multiple confounded effects, the analysis is similar. We illustrate with two confounded effects which requires a minimum of two instruments. The DAG G below depicts the situation.



DAG G

Suppose the causal effect of interest is $\Pr\left(Y \mid do\left(X_1 = x_1\right), do\left(X_2 = x_2\right)\right)$. The back-doors into causal variables X_1 and X_2 cannot be blocked due to the unobservable variables U_1 and U_2. However, Z_1 and Z_2 satisfy the conditions for instruments. They are related to X_1 and X_2 but independent of Y in the graph removing the direct paths from X_1 to Y and X_2 to Y, G_α.



DAG G_α

Since X_1 and X_2 are colliders in DAG G_α, Y is independent of Z_1 and Z_2.

Linear-IV, by components, involves $r_{YZ_1\cdot Z_2} = ab + ef, r_{YZ_2\cdot Z_1} = bc + df, r_{X_1Z_1\cdot Z_2} = a, r_{X_1Z_2\cdot Z_1} = c, r_{X_2Z_1\cdot Z_2} = e, r_{X_2Z_2\cdot Z_1} = d$. The causal estimands b and f can be recovered from a system of linear equations. For instance,

$$\begin{aligned} \frac{r_{YZ_1\cdot Z_2}}{r_{X_1Z_1\cdot Z_2}} &= \frac{ab+ef}{a} = b + \frac{e}{a}f \\ \frac{r_{YZ_2\cdot Z_1}}{r_{X_1Z_2\cdot Z_1}} &= \frac{bc+df}{c} = b + \frac{d}{c}f \end{aligned}$$

or

$$A \begin{bmatrix} b \\ f \end{bmatrix} = \begin{bmatrix} \frac{r_{YZ_1\cdot Z_2}}{r_{X_1Z_1\cdot Z_2}} \\ \frac{r_{YZ_2\cdot Z_1}}{r_{X_1Z_2\cdot Z_1}} \end{bmatrix}$$

where

$$A = \begin{bmatrix} 1 & \frac{r_{X_2Z_1\cdot Z_2}}{r_{X_1Z_1\cdot Z_2}} \\ 1 & \frac{r_{X_2Z_2\cdot Z_1}}{r_{X_1Z_2\cdot Z_1}} \end{bmatrix}$$

It is noteworthy, that A is full rank and we have exhausted the information in the instrumental variable strategy in the above steps.

Of course, standard linear-IV solves this system of equations in the conventional two-step process and the same applies to the control function-IV strategy. Let

$$P_Z = Z\left(Z^T Z\right)^{-1} Z^T$$

where $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$ and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$. Then

$$\begin{bmatrix} b \\ f \end{bmatrix} = \left(X^T P_Z X\right)^{-1} X^T P_Z Y$$

Also, let $X_u = \begin{bmatrix} X_1 & X_2 & u_1 & u_2 \end{bmatrix}$ where $u_1 = (I - P_Z) X_1$ and $u_2 = (I - P_Z) X_2$ or $X_u = \begin{bmatrix} X & (I - P_Z) X \end{bmatrix}$. Then, the control function-IV strategy recovers the causal estimands as follows.

$$\begin{bmatrix} b \\ f \\ \frac{a_1 a_2 Var[U_1]}{a_1^2 Var[U_1] + Var\left[U_{X_1}\right]} \\ \frac{a_3 a_4 Var[U_2]}{a_3^2 Var[U_2] + Var\left[U_{X_2}\right]} \end{bmatrix} = \left(X_u^T X_u\right)^{-1} X_u^T Y$$

Again, the estimands on u_1 and u_2 indicate the level of confounding.[1]

[1] Standard errors based on classical OLS estimates are misleading for the control function-IV strategy as the estimate of $Var[\varepsilon]$ is downward biased. However, Bayesian credible intervals avoid the problem.