**Twin Research and Human Genetics**

**Supplementary material**

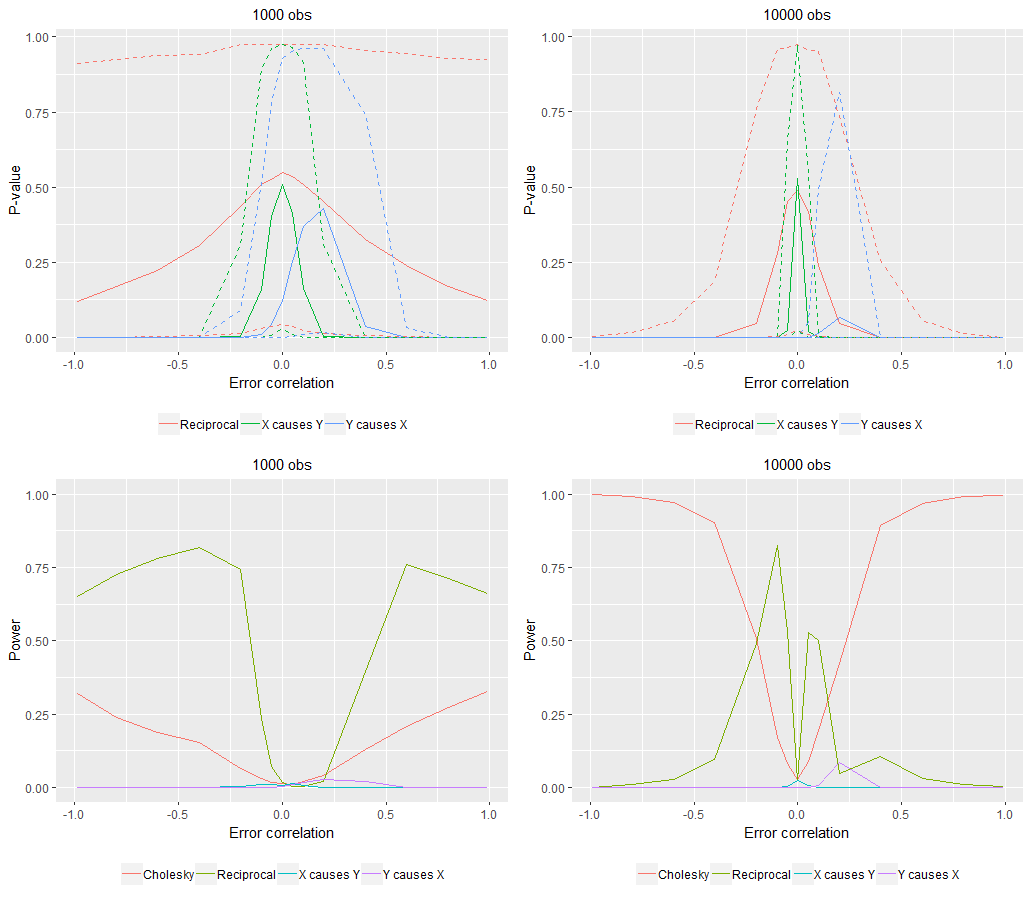
**A major limitation of the direction of causation model: non-shared environmental confounding**

Stig Hebbelstrup Rye Rasmussen, Steven Ludeke, and Jacob v.B. Hjelmborg

**Appendix A: Additional results**

**Results from univariate ACE-ACE models**

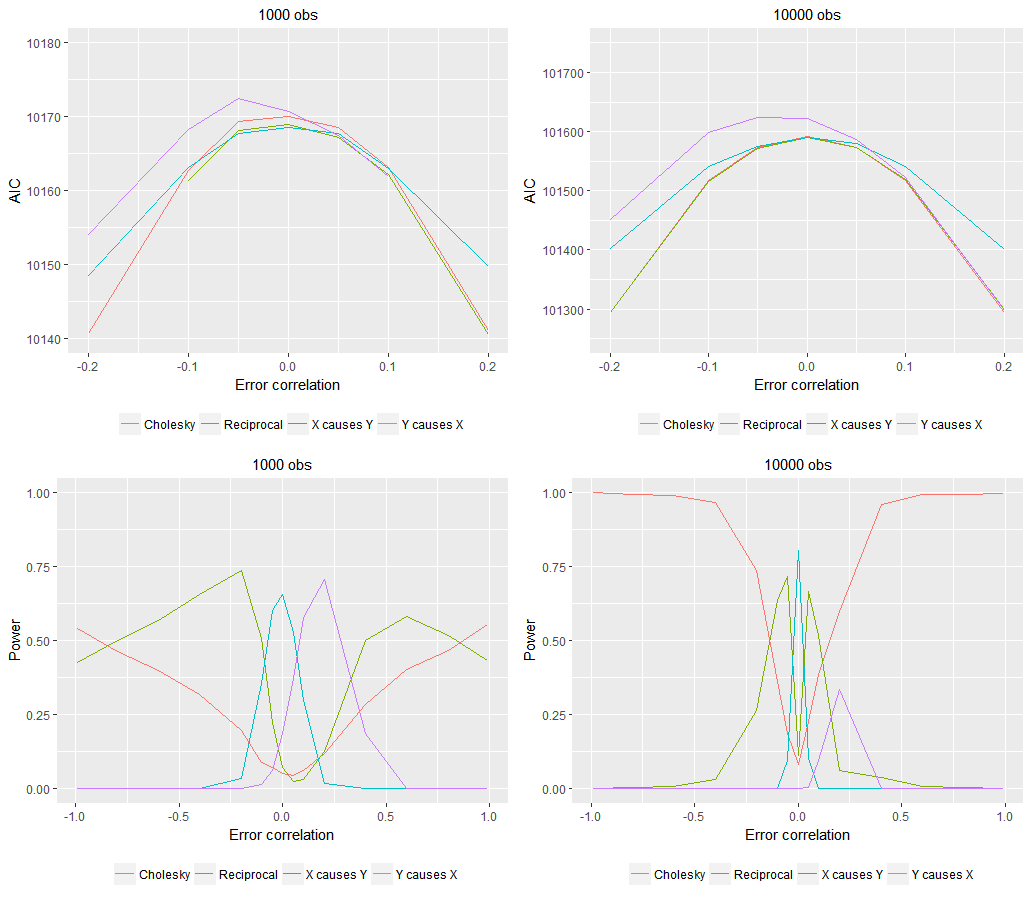
**Comparison of likelihood ratio tests in two sample size situations for ACE-ACE models**



\* The bold lines in the top two graphs represent the median, and the dashed lines represent 95 % confidence intervals.

There are several major differences with this model compared to the AE-CE model. First of all, it is quite difficult to distinguish between the two unidirectional models in this situation when the sample size is 1000. The success ratio is essentially zero even when there is zero confounding; primarily because both the unidirectional model (X → Y) and the reciprocal model cannot be rejected in many cases and we cannot therefore, based on likelihood ratio testing alone, determine the direction of causation. This situation is not improved when the sample size increases to 10000.. In most other situations than zero confounding the model which we cannot reject is the reciprocal model, and we would therefore, based on the logic of the DoC model, have to choose this model in many cases, although it is not the true causal model. When we have 10.000 observations we would choose the Cholesky model in the face of severe confounding.

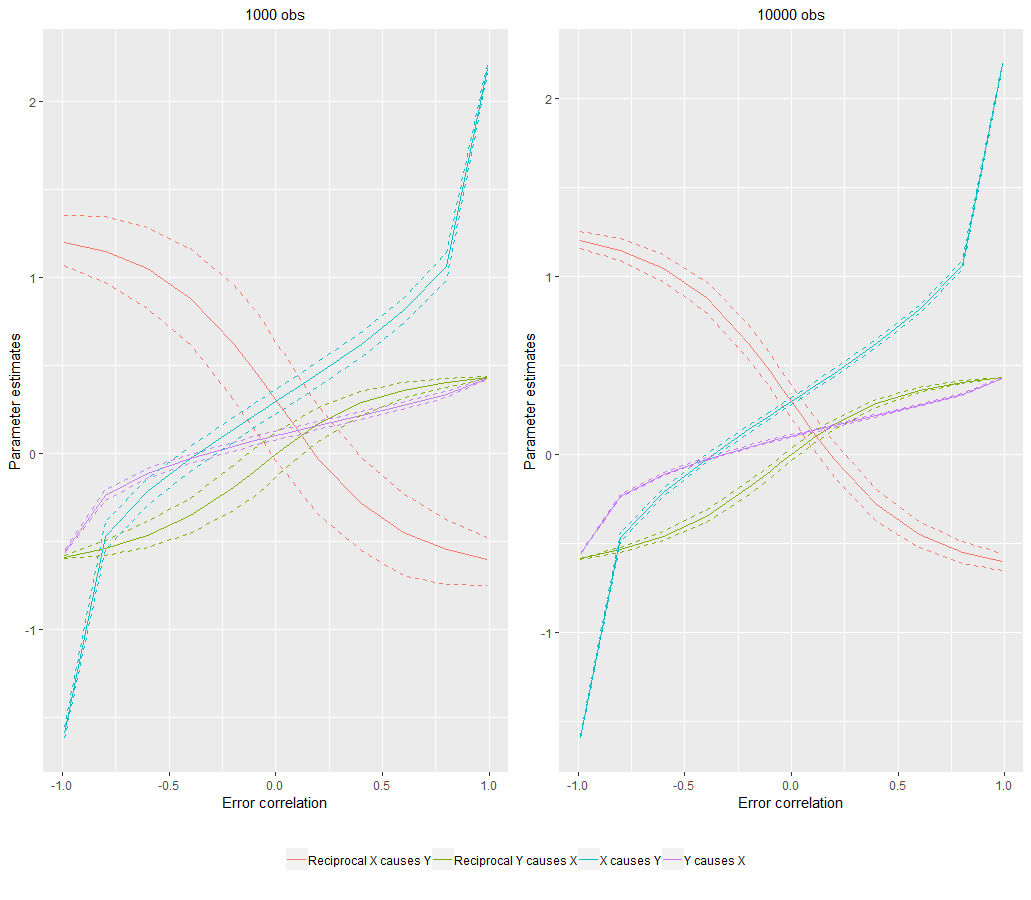
**Comparison of AIC values in two sample size situations for ACE-ACE models**



\* The bold lines in the top two graphs represent the median, and the dashed lines represent 95 % confidence intervals..

In the situation with no confounding the true causal model, i.e. the X → Y model, has the lowest AIC for both samples sizes. In almost all other situations than zero confounding it is difficult to choose the correct model; when the sample size is small both the reciprocal model or the Y → X model are sometimes chosen above the correct model and when the sample size is large either the reciprocal model or the Cholesky model is the preferred model. This differs from the AE-CE model where the Cholesky model was consistently the preferred model. Most true models are obviously mixes of these two models and our expectation would be that as the mode of inheritance becomes increasingly dissimilar the more often we would end up choosing the Cholesky model as compared to the reciprocal model even if the X → Y model is the correct model.

**Comparison of causal parameter estimates in two sample size situations for ACE-ACE models**



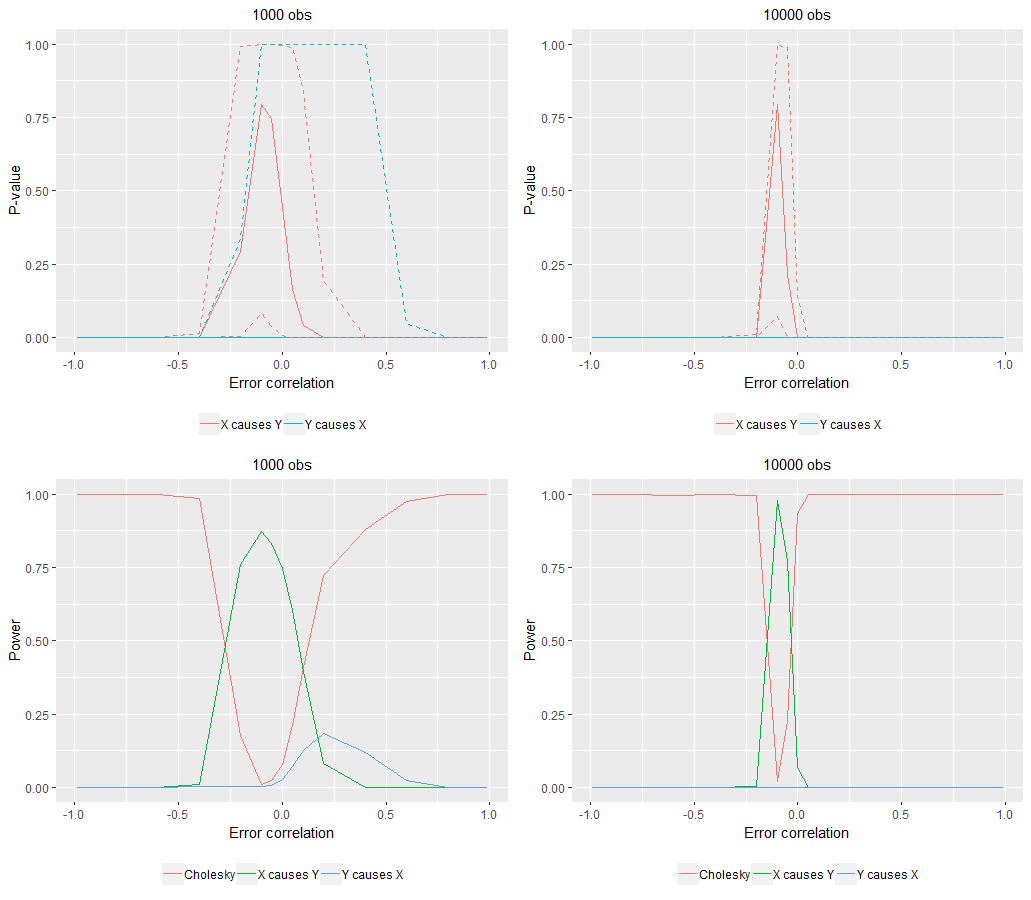
\* The bold lines represent the median causal parameter estimate, and the dashed lines represent 95 % confidence intervals

The parameter estimates from the ACE-ACE models are quite well recovered in the case of zero confounding, both when we estimate the true model as well as when we estimate the reciprocal model, where the Y → X path is estimated at roughly zero and the X → Y path is estimated at roughly 0.3. If we on the other hand end up choosing the reciprocal model and there is even a slight amount of confounding the parameter estimates quickly diverge from the true model. For instance, when there is a negative amount of confounding of 0.2 the median X → Y estimate is estimated at 0.14 and the median effect of Y → X is estimated at 0.04.

**Results from reciprocal models**

This section outlines the simulation results for the DoC model when the population model is a reciprocal model. The AE-CE and ACE-ACE models are identical to those investigated previously except we add an effect of 0.2 from Y to X.

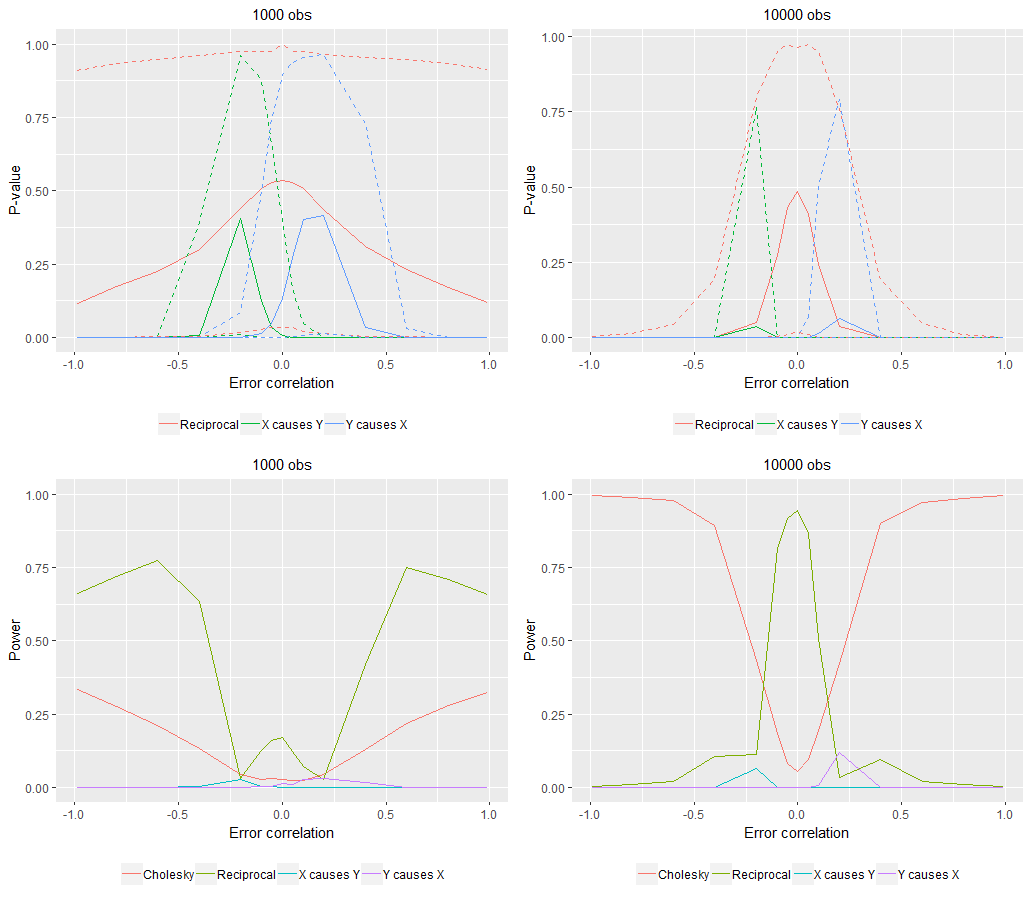
*Comparison of likelihood ratio tests in two sample size situations for AE-CE models*



\* The bold lines in the top two graphs represent the median, and the dashed lines represent 95 % confidence intervals.

As before we cannot use likelihood ratio testing for the reciprocal model since it has the same degrees of freedom as the Cholesky model. We can however compare the Cholesky model to the two unidirectional models. When we have 1000 observations and there is no confounding on average we would end up choosing the unidirectional model where X causes Y – not surprising since this is the strongest effect. In most other cases, i.e. when there is a slight amount of confounding we would choose the Cholesky model.

**Comparison of likelihood ratio tests in two sample size situations for ACE-ACE models**

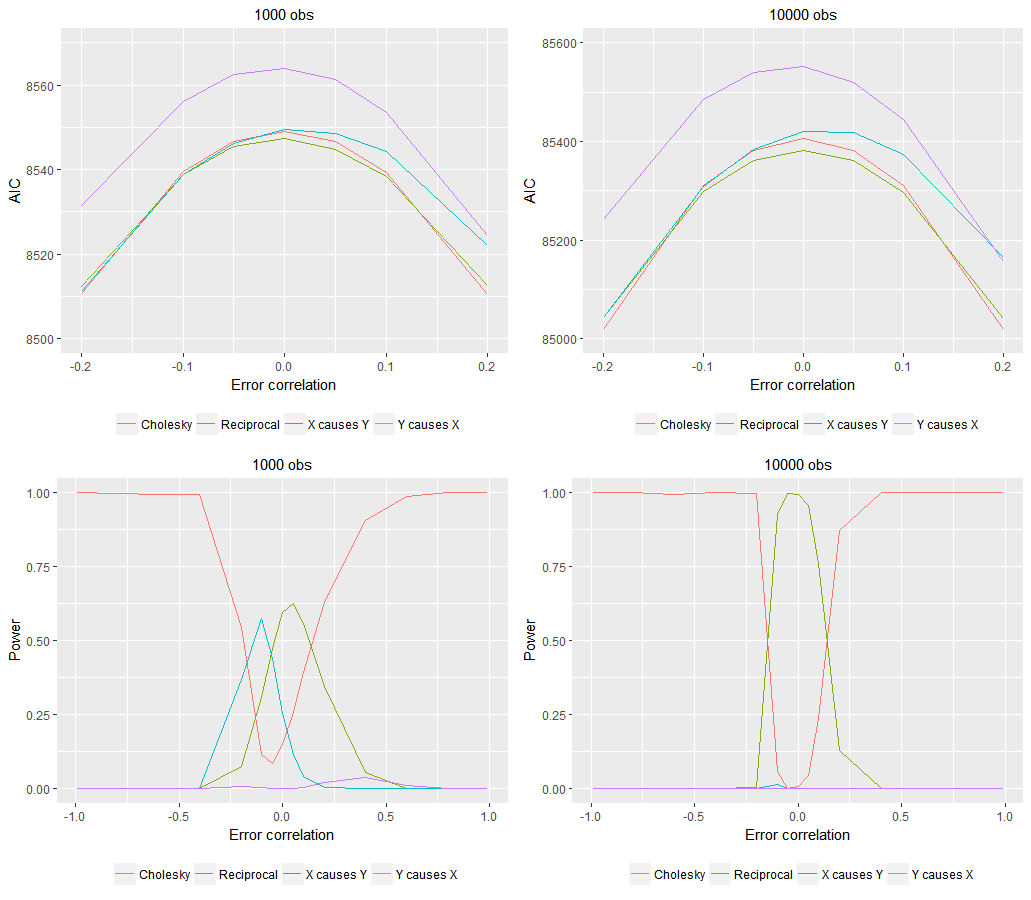


\* The bold lines in the top two graphs represent the median, and the dashed lines represent 95 % confidence intervals.

The results from the ACE-ACE are different than when the unidirectional model was the true population model: Only when we have 10,000 observations do we consistently choose X → Y model; when the sample size is only 1000 it is difficult to choose between the two unidirectional models.

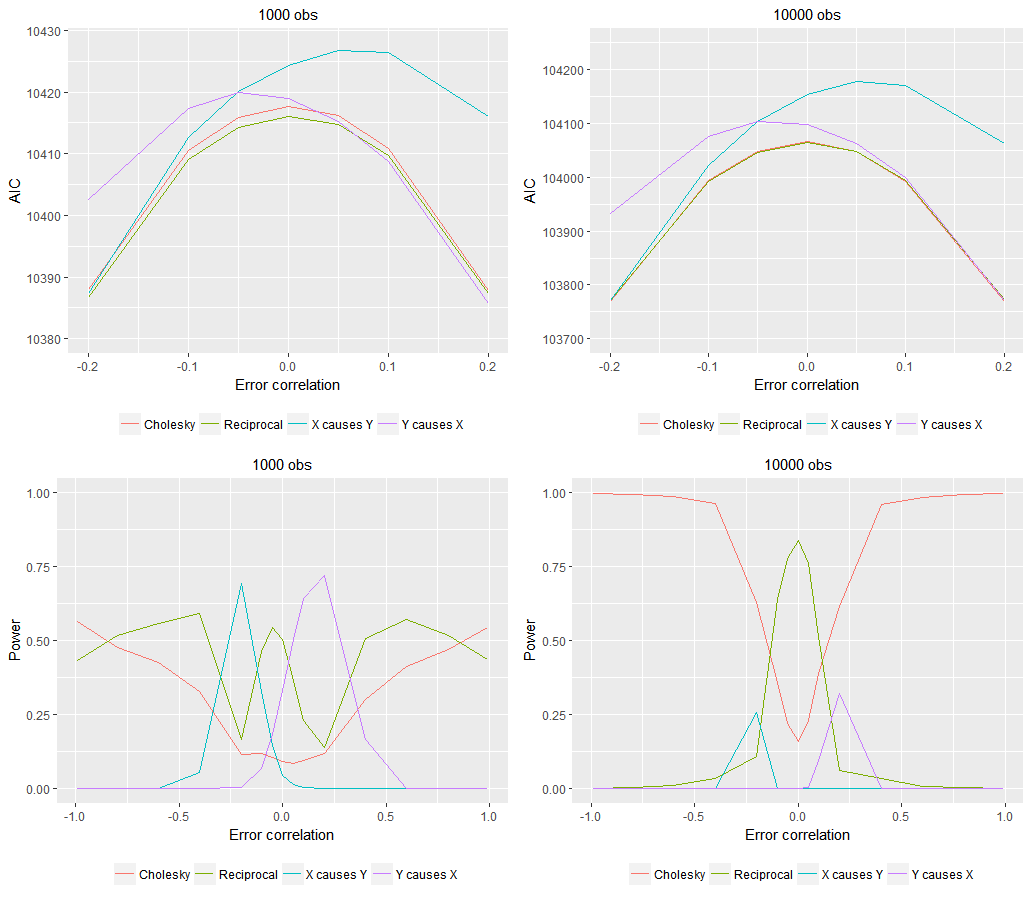
In terms of the AIC for the AE-CE model we end up choosing the correct model on balance, when there is no confounding, as evidenced by the power graphs. When the error correlation is about .2 the success ratio is very close to zero for both sample sizes.

**Comparison of AIC values in two sample size situations for AE-CE models**



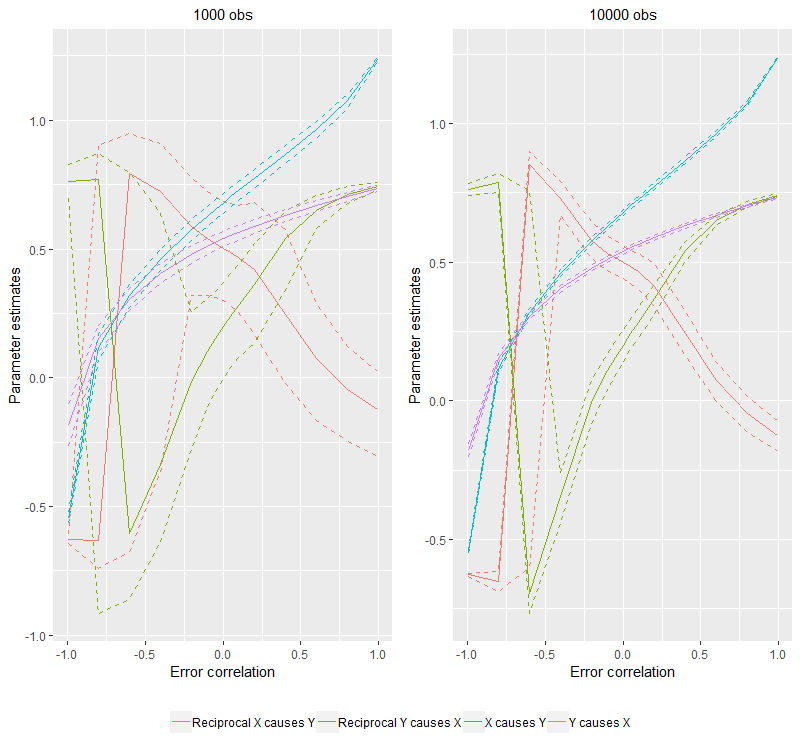
In terms of the ACE-ACE models the results for the AIC are similar to those obtained when using likelihood ratio testing for *N* = 10,000: The correct model is chosen when the confounding is zero but decreases rapidly as the error correlation increases.

**Comparison of AIC values in two sample size situations for ACE-ACE models**



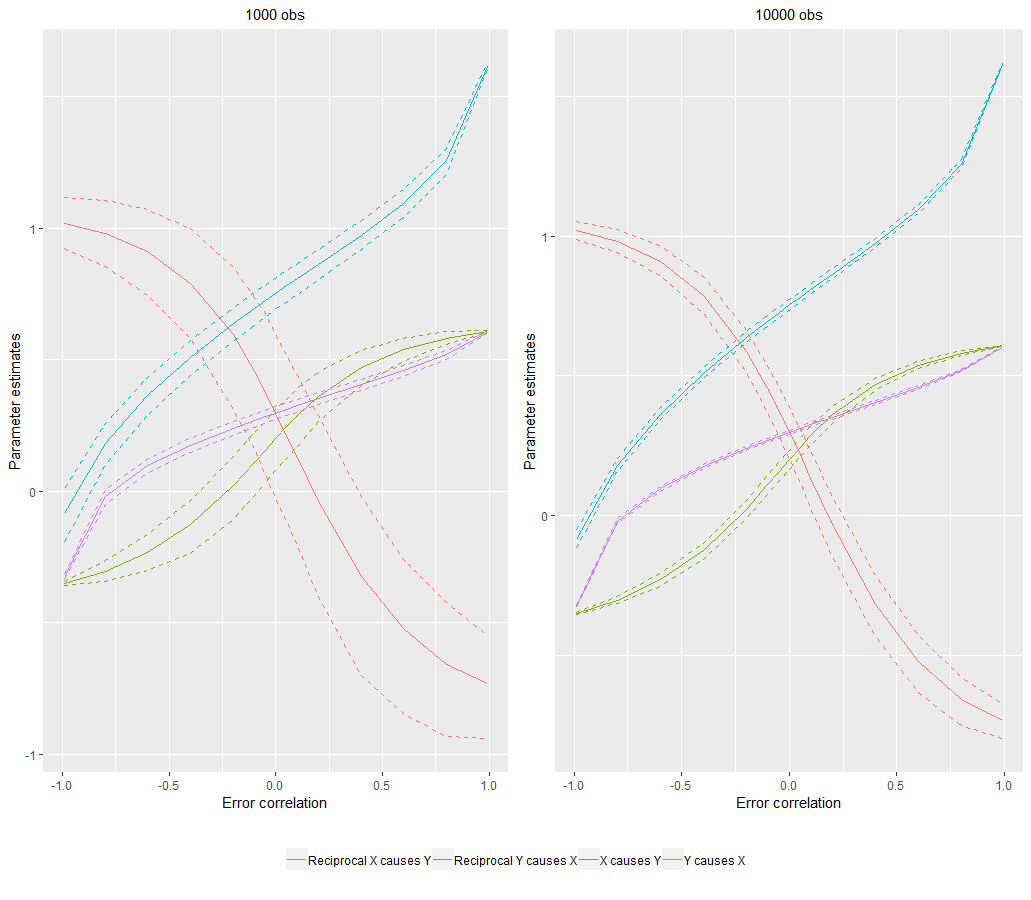
In terms of the parameter estimates for the reciprocal models in the figures blow they follow a somewhat similar pattern,. In all cases we obtain the correct parameter estimates, as evidenced by the median values for the different values, when there is zero confounding. However, in all models the parameter values change with only small changes in the amount of confounding for none of the models does it take a lot for the parameter values to diverge.

**Comparison of causal parameter estimates in two sample size situations for AE-CE models**



\* The bold lines represent the median causal parameter estimate, and the dashed lines represent 95% confidence intervals.

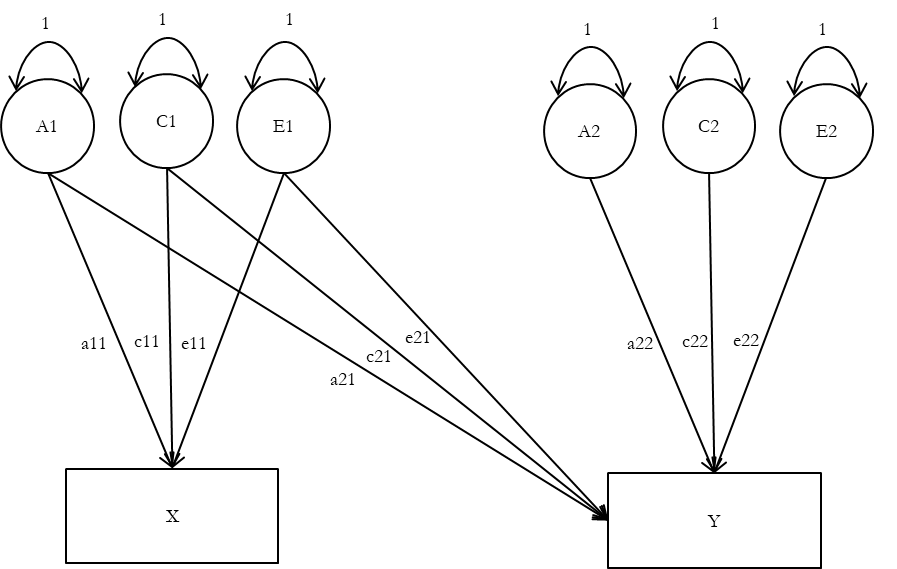
**Comparison of causal parameter estimates in two sample size situations for ACE-ACE models**



\* The bold lines represent the median causal parameter estimate, and the dashed lines represent 95% confidence intervals.

**Appendix B: Additional figure**

**Figure A1: Cholesky decomposition**

****

**Appendix C: Additional results**

**Figure 1: AIC values for AE-CE model with 1000 observations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cholesky** | | | **Reciprocal model** | | | **X causes Y** | | | **Y causes X** | | |
| **Error correlation** | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile |
| *-.99* | 5509.151 | 5334.521 | 5674.659 | 5660.394 | 5487.849 | 5830.676 | 6014.655 | 5831.754 | 6200.789 | 6891.166 | 6688.298 | 7089.671 |
| *-.8* | 7254.957 | 7080.766 | 7419.365 | 7339.575 | 7168.351 | 7508.914 | 7462.563 | 7281.553 | 7646.489 | 7480.674 | 7298.817 | 7664.945 |
| *-.6* | 7720.596 | 7545.605 | 7884.121 | 7769.707 | 7592.504 | 7939.687 | 7812.708 | 7637.107 | 7985.974 | 7809.377 | 7634.523 | 7980.314 |
| *-.4* | 7962.667 | 7787.707 | 8126.022 | 7980.089 | 7805.653 | 8146.745 | 7998.74 | 7823.212 | 8163.541 | 8004.175 | 7831.321 | 8170.979 |
| *-.2* | 8087.09 | 7912.327 | 8250.334 | 8091.596 | 7916.487 | 8257.457 | 8094.94 | 7921.311 | 8256.771 | 8110.494 | 7935.476 | 8272.773 |
| *-.1* | 8116.179 | 7941.606 | 8279.522 | 8117.084 | 7943.197 | 8280.277 | 8116.988 | 7941.291 | 8278.723 | 8134.397 | 7958.201 | 8296.372 |
| *-.05* | 8123.301 | 7948.833 | 8286.712 | 8124.082 | 7949.955 | 8286.578 | 8122.837 | 7948.681 | 8285.469 | 8140.031 | 7963.143 | 8301.79 |
| *0* | 8125.613 | 7951.253 | 8289.098 | 8126.239 | 7952.188 | 8289.274 | 8124.625 | 7950.2 | 8287.893 | 8141.45 | 7964.759 | 8302.879 |
| *.05* | 8123.14 | 7948.892 | 8286.709 | 8123.58 | 7951.105 | 8286.747 | 8122.468 | 7949.265 | 8284.777 | 8138.208 | 7961.764 | 8299.624 |
| *.1* | 8115.826 | 7941.723 | 8279.517 | 8117.083 | 7944.758 | 8280.622 | 8116.782 | 7943.111 | 8279.81 | 8130.649 | 7954.885 | 8292.331 |
| *.2* | 8086.342 | 7912.567 | 8250.322 | 8090.793 | 7915.647 | 8254.006 | 8092.626 | 7916.802 | 8257.244 | 8102.727 | 7927.924 | 8263.173 |
| *.4* | 7961.727 | 7788.279 | 8125.997 | 7980.929 | 7801.748 | 8142.187 | 7996.877 | 7815.15 | 8158.145 | 7983.479 | 7809.053 | 8143.855 |
| *.6* | 7719.537 | 7546.536 | 7884.276 | 7758.549 | 7583.821 | 7922.226 | 7812.811 | 7631.04 | 7980.614 | 7757.342 | 7582.895 | 7920.605 |
| *.8* | 7254.711 | 7081.175 | 7419.677 | 7346.61 | 7174.23 | 7511.68 | 7463.555 | 7278.371 | 7634.409 | 7345.295 | 7172.706 | 7509.782 |
| *.99* | 5509.322 | 5335.197 | 5674.357 | 6485.116 | 6292.581 | 6663.958 | 6014.944 | 5821.776 | 6204.559 | 6486.827 | 6294.598 | 6665.982 |

**Figure 2: AIC values for AE-CE model with 10,000 observations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cholesky** | | | **Reciprocal model** | | | **X causes Y** | | | **Y causes X** | | |
| **Error correlation** | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile |
| *-.99* | 55010.8 | 54469.9 | 55560.5 | 56525.5 | 55978.4 | 57064.8 | 60047.0 | 59469.8 | 60657.7 | 68846.6 | 68220.4 | 69456.7 |
| *-.8* | 72464.3 | 71926.9 | 73015.4 | 73317.9 | 72770.6 | 73864.1 | 74542.6 | 74007.1 | 75096.7 | 74713.5 | 74158.6 | 75259.4 |
| *-.6* | 77115.6 | 76577.4 | 77666.1 | 77705.5 | 77162.4 | 78264.1 | 78053.2 | 77521.5 | 78593.9 | 78017.2 | 77486.4 | 78565.7 |
| *-.4* | 79536.0 | 78997.5 | 80086.4 | 79713.5 | 79160.3 | 80267.7 | 79902.3 | 79372.0 | 80453.7 | 79973.1 | 79434.4 | 80520.3 |
| *-.2* | 80781.0 | 80242.3 | 81331.3 | 80830.0 | 80284.3 | 81370.7 | 80870.1 | 80325.1 | 81414.2 | 81024.4 | 80471.4 | 81571.7 |
| *-.1* | 81073.3 | 80534.6 | 81623.6 | 81082.5 | 80542.2 | 81634.4 | 81090.8 | 80549.2 | 81646.8 | 81269.7 | 80713.7 | 81817.6 |
| *-.05* | 81145.4 | 80606.5 | 81695.6 | 81147.0 | 80607.9 | 81700.5 | 81148.6 | 80608.3 | 81703.3 | 81325.4 | 80771.0 | 81875.9 |
| *0* | 81169.3 | 80630.4 | 81719.5 | 81168.0 | 80629.6 | 81722.0 | 81169.1 | 80628.9 | 81720.6 | 81336.7 | 80784.6 | 81887.3 |
| *.05* | 81145.4 | 80606.5 | 81695.6 | 81147.0 | 80611.2 | 81697.7 | 81150.2 | 80610.6 | 81697.0 | 81304.5 | 80754.2 | 81856.1 |
| *.1* | 81073.4 | 80534.5 | 81623.5 | 81083.2 | 80551.4 | 81633.0 | 81094.5 | 80556.0 | 81642.9 | 81229.8 | 80676.4 | 81780.8 |
| *.2* | 80781.1 | 80242.2 | 81331.2 | 80825.9 | 80296.5 | 81381.8 | 80866.8 | 80329.0 | 81421.6 | 80940.0 | 80392.5 | 81492.0 |
| *.4* | 79536.2 | 78997.2 | 80086.2 | 79705.1 | 79175.2 | 80259.4 | 79908.5 | 79366.5 | 80462.0 | 79749.2 | 79206.6 | 80308.2 |
| *.6* | 77115.9 | 76576.9 | 77665.9 | 77483.7 | 76954.9 | 78043.2 | 78060.5 | 77502.4 | 78617.5 | 77487.4 | 76955.9 | 78046.2 |
| *.8* | 72464.4 | 71926.2 | 73015.3 | 73379.0 | 72842.4 | 73937.6 | 74553.2 | 73979.3 | 75112.2 | 73380.6 | 72844.9 | 73938.2 |
| *.99* | 55010.5 | 54469.9 | 55560.5 | 64794.2 | 64197.3 | 65372.4 | 60066.2 | 59435.5 | 60675.0 | 64815.0 | 64212.0 | 65391.8 |

**Figure 3: AIC values for ACE-ACE model with 1000 observations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cholesky** | | | **Reciprocal model** | | | **X causes Y** | | | **Y causes X** | | |
| **Error correlation** | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile |
| *-.99* | 7948.2 | 7774.6 | 8114.7 | 7949.4 | 7774.0 | 8114.8 | 8930.6 | 8725.4 | 9132.7 | 8523.4 | 8342.1 | 8701.3 |
| *-.8* | 9498.2 | 9324.1 | 9664.1 | 9498.3 | 9324.7 | 9662.5 | 9787.2 | 9606.4 | 9973.8 | 9712.2 | 9535.7 | 9891.0 |
| *-.6* | 9859.4 | 9686.7 | 10024.8 | 9860.0 | 9687.7 | 10024.6 | 9980.5 | 9805.3 | 10159.6 | 9956.1 | 9780.5 | 10132.0 |
| *-.4* | 10046.4 | 9874.2 | 10211.7 | 10045.8 | 9873.7 | 10210.1 | 10091.7 | 9917.0 | 10260.6 | 10088.7 | 9914.9 | 10258.5 |
| *-.2* | 10140.8 | 9969.1 | 10307.1 | 10139.5 | 9967.6 | 10305.3 | 10148.6 | 9975.7 | 10315.6 | 10154.1 | 9980.4 | 10322.2 |
| *-.1* | 10162.7 | 9989.6 | 10329.4 | 10161.3 | 9988.3 | 10327.8 | 10163.2 | 9988.9 | 10330.0 | 10168.2 | 9994.1 | 10336.4 |
| *-.05* | 10169.4 | 9996.7 | 10335.2 | 10168.2 | 9994.9 | 10333.5 | 10167.7 | 9995.5 | 10333.4 | 10172.5 | 9998.0 | 10339.9 |
| *0* | 10170.1 | 9996.9 | 10336.8 | 10169.0 | 9994.9 | 10335.2 | 10168.6 | 9993.1 | 10333.7 | 10170.8 | 9996.8 | 10338.4 |
| *.05* | 10168.6 | 9995.2 | 10334.9 | 10167.3 | 9993.2 | 10333.6 | 10167.7 | 9991.4 | 10333.0 | 10167.4 | 9993.8 | 10334.3 |
| *.1* | 10163.1 | 9989.7 | 10329.3 | 10162.2 | 9988.1 | 10328.2 | 10163.0 | 9987.3 | 10328.5 | 10162.0 | 9987.4 | 10327.5 |
| *.2* | 10141.1 | 9967.5 | 10306.9 | 10140.6 | 9967.0 | 10305.9 | 10149.8 | 9970.3 | 10313.2 | 10139.7 | 9965.1 | 10304.5 |
| *.4* | 10045.4 | 9874.2 | 10211.9 | 10045.4 | 9874.4 | 10210.7 | 10089.6 | 9909.0 | 10254.0 | 10048.6 | 9876.1 | 10213.8 |
| *.6* | 9860.0 | 9687.6 | 10026.1 | 9860.2 | 9688.1 | 10025.2 | 9981.1 | 9797.2 | 10150.3 | 9877.0 | 9701.5 | 10041.7 |
| *.8* | 9497.6 | 9324.1 | 9663.8 | 9497.8 | 9325.6 | 9663.6 | 9790.1 | 9601.6 | 9961.8 | 9542.2 | 9365.3 | 9703.8 |
| *.99* | 7948.1 | 7774.9 | 8114.9 | 7949.3 | 7775.7 | 8114.7 | 8937.0 | 8733.6 | 9129.2 | 8048.4 | 7868.0 | 8207.6 |

**Figure 4: AIC values for ACE-ACE model with 10,000 observations**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cholesky** | | | **Reciprocal model** | | | **X causes Y** | | | **Y causes X** | | |
| **Error correlation** | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile | Median | 2.5th percentile | 97.5th percentile |
| *-.99* | 79368.4 | 78829.6 | 79916.9 | 79390.0 | 78855.4 | 79942.1 | 89246.7 | 88593.8 | 89867.5 | 85162.0 | 84550.2 | 85775.7 |
| *-.8* | 94861.7 | 94322.5 | 95410.1 | 94879.6 | 94345.6 | 95431.9 | 97790.1 | 97223.5 | 98362.4 | 97046.4 | 96494.0 | 97610.1 |
| *-.6* | 98487.5 | 97947.7 | 99035.7 | 98500.4 | 97963.0 | 99052.4 | 99710.3 | 99153.2 | 100268.1 | 99467.1 | 98919.0 | 100019.9 |
| *-.4* | 100342.4 | 99802.9 | 100885.9 | 100348.8 | 99811.7 | 100898.1 | 100809.4 | 100255.4 | 101357.7 | 100777.6 | 100222.6 | 101320.8 |
| *-.2* | 101294.9 | 100754.5 | 101843.5 | 101294.4 | 100757.2 | 101850.2 | 101402.7 | 100849.1 | 101945.8 | 101451.8 | 100899.0 | 102000.7 |
| *-.1* | 101517.4 | 100977.9 | 102067.2 | 101516.4 | 100977.5 | 102068.6 | 101540.8 | 100996.8 | 102088.5 | 101598.2 | 101043.6 | 102147.8 |
| *-.05* | 101572.4 | 101033.0 | 102122.3 | 101571.4 | 101031.5 | 102121.7 | 101574.3 | 101035.9 | 102125.1 | 101623.1 | 101073.0 | 102176.5 |
| *0* | 101591.6 | 101051.2 | 102140.7 | 101590.5 | 101049.5 | 102139.0 | 101590.7 | 101050.2 | 102137.4 | 101622.3 | 101073.6 | 102174.4 |
| *.05* | 101573.3 | 101033.0 | 102122.4 | 101572.5 | 101033.0 | 102120.4 | 101579.4 | 101038.4 | 102126.0 | 101587.4 | 101043.7 | 102140.2 |
| *.1* | 101518.2 | 100978.0 | 102067.4 | 101519.3 | 100980.4 | 102066.2 | 101541.7 | 101006.2 | 102091.4 | 101522.4 | 100982.3 | 102074.9 |
| *.2* | 101294.8 | 100754.6 | 101844.2 | 101298.8 | 100761.9 | 101846.0 | 101400.7 | 100865.1 | 101949.5 | 101297.6 | 100760.7 | 101845.0 |
| *.4* | 100342.8 | 99803.0 | 100892.8 | 100353.2 | 99816.0 | 100899.4 | 100811.0 | 100267.6 | 101359.0 | 100394.2 | 99865.6 | 100940.8 |
| *.6* | 98487.4 | 97947.8 | 99037.7 | 98502.4 | 97965.4 | 99050.1 | 99712.7 | 99163.0 | 100267.1 | 98675.4 | 98149.9 | 99224.5 |
| *.8* | 94861.7 | 94322.6 | 95412.6 | 94879.1 | 94343.7 | 95428.9 | 97788.2 | 97199.1 | 98349.9 | 95320.9 | 94782.7 | 95865.2 |
| *.99* | 79367.6 | 78829.7 | 79919.5 | 79390.1 | 78860.8 | 79942.8 | 89257.8 | 88606.8 | 89864.6 | 80375.0 | 79823.7 | 80931.3 |

**Appendix D: Brief discussion of other important assumptions in the DoC model**

In the main paper we have primarily focused on the problem of non-shared confounding but there are obviously other important assumptions that the DoC rests as discussed previously in the literature. Two of the most important ones are measurement error and power.

***Measurement error***

In a linear regression framework it is well known that measurement error, other things being equal, tends to attenuate regression coefficients for the predictor variables. In a multiple regression the introduction of measurement error can affect all other parameters . It is also well known that we assume, when estimating the unidirectional ACE twin model, that the A and C components are not affected by measurement error but is absorbed in the E component; and in the bivariate case measurement error also affects the estimation of the shared E variance .

In the DoC model measurement error affects the estimation of all other parameters and in this sense thus most closely resembles the effect measurement error has in the classic multiple regression case. Several simulations studies have demonstrated that measurement error can have a large impact on which of the DoC models (1)-(3) are to be retained . For example, Neale and Cardon (1992) used a simulation study to demonstrate that in situations of unidirectional causation, unequal reliability tends to attenuate the parameter estimate of the construct with the lowest reliability, whereas in situations of reciprocal causation the bias can run in both directions. In principle the challenge of (random) measurement error can be addressed by including a measurement model in a correctly specified structural equation model with multiple indicators for each of the constructs ; in a common factor model only the shared variance between the indicators is retained in the estimation of the measurement model for the (latent) constructs whereas error variance and item specific variance is not . Verhulst and Estabrook e.g. created measurement models for both personality traits and attitudes using several indicators to define the latent construct psychoticism, a type of personality trait, and sex attitudes. When estimating reciprocal and unidirectional DoC models they then used these latent constructs to account for random measurement error which can, in a correctly specified model alleviate the problems of random measurement error.

The problem of measurement error is also relevant for the discussion about the relationship between personality traits and politics. For instance, Verhulst and colleagues (2012) found, to the surprise of many, that the direction of causation in some cases runs from political attitudes to personality traits: the personality trait Psychoticism might be caused by differences in social ideology but not the other way round. This counterintuitive result may reflect the low reliability of Psychoticism. In their study Psychoticism has a composite reliability of 0.679, whereas the composite reliability of the social ideology construct is 0.842.[[1]](#footnote-1) The fact that the measures are differentially reliable is a cause for concern regarding the robustness of this particular finding, since Verhulst and colleagues (2012) use factor scores, which like sum scores contain measurement error . Since many constructs in social science are measured with error care needs to be taken when the DoC is applied.

***Power***

The DoC model requires a very large sample size to not be underpowered; in fact, usual sample sizes in surveys of 1000-3000 respondents are likely not enough for most applications of the DoC model. To make things a bit more concrete let us consider two extreme examples from Heath and colleagues (1993) simulation studies. In the first case trait X and trait Y, which are both standardized to unit variance, are measured without error, the mode of inheritance for trait X is an AE model (e2=0.25, a2=0.75) and the mode of inheritance is trait Y is a CE model (c2=0.75, e2=0.25), and the causal effect of trait X on trait Y (X → Y) is 0.5 i.e. large. This would only require a sample size of 106 twin pairs assuming an equal number of MZ and DZ twins to reject the causal hypothesis Y → X with 80 % power. If on the other hand there is 20 % measurement error in both traits, the mode of inheritance is an ACE model for trait X (a2=0.5, c2=0.25, e2=0.25), and also an ACE model for trait Y (a2=0.25, c2=0.5,e2=0.25), and a more modest effect of X → Y of 0.2 a sample size of 16019 twin pairs, i.e. a sample size of 32038 individual respondents, would be needed to reject the false model of Y → X with 80 % power .

In general, the power to resolve alternative causal hypotheses in terms of the DoC depends on (1) the sample size, (2) the amount of measurement error in the two traits, (3) the causal effect of X on Y, and (4) the similarity of the mode of inheritance . To this we can add the finding from this study, namely that sometimes it is in fact *easier* to reject the true model in the face of non-shared confounding when sample sizes are large. In a sense this should come as no surprise: We are trying to reproduce an observed covariance matrix by estimating the wrong model – i.e. without taking into account non-shared confounding.

**References**

Kenneth A. Bollen, 1989. *Structural Equations with Latent Variables*, New York: John Wiley and Sons.

Timothy A. Brown, 2006. *Confirmatory Factor Analysis for Applied Research*, New York: Guilford Publications.

David L. Duffy and Nicholas G. Martin, 1994. Inferring the Direction of Causation in Cross-Sectional Twin Data: Theoretical and Empirical Considerations, *Genetic Epidemiology* 11.

R. Estabrook and M. Neale, 2013. A Comparison of Factor Score Estimation Methods in the Presence of Missing Data: Reliability and an Application to Nicotine Dependence, *Multivariate Behavioral Research* 48.

Andrew C Heath, Ronald C Kessler, Michael C Neale, John K Hewitt, Lindon J Eaves, and Kenneth S Kendler, 1993. Testing Hypotheses About Direction of Causation Using Cross-Sectional Family Data, *Behavior Genetics* 23.

Michael Neale and Lon Cardon, 1992. *Methodology for Genetic Studies of Twins and Families*: Springer.

T. Raykov, 1997. Estimation of Composite Reliability for Congeneric Measures, *Applied Psychological Measurement* 21.

B. Verhulst, L. J. Eaves, and P. K. Hatemi, 2012. Correlation Not Causation: The Relationship between Personality Traits and Political Ideologies, *American Journal of Political Science* 56.

Brad Verhulst and Ryne Estabrook, 2012. Using Genetic Information to Test Causal Relationships in Cross-Sectional Data, *Journal of Theoretical Politics* 24.

1. The composite reliability is calculated using the standardized factor loadings in Verhulst and Hatemi’s online appendix 2012. Correlation Not Causation: The Relationship between Personality Traits and Political Ideologies, *American Journal of Political Science* 56.. [↑](#footnote-ref-1)