# Direction Dependence Analysis: A Unified Framework to Evaluate the Direction of Dependence in Linear Models

Wolfgang Wiedermann
University of Missouri, Columbia, MO, USA – wiedermannw@missouri.edu

**Abstract**

In observational studies, at least three possible explanations exist for the association of two variables $x$ and $y$: 1) $x$ is the cause of $y$, 2) $y$ is the cause of $x$, or 3) an unmeasured confounder is present. Statistical tests which identify which of the three explanatory models fits best would be a useful adjunct to use of theory alone. The present paper introduces one such statistical method, Direction Dependence Analysis (DDA), which assesses the relative plausibility of the three explanatory models based on higher moment information of the variables (i.e., skewness and kurtosis). DDA involves the evaluation of three properties of the data: 1) The observed distributions of the variables, 2) the residual distributions of competing models, and 3) the independence properties of predictors and residuals of competing models. When observed variables are non-normally distributed, it is shown that DDA components can be used to uniquely identify each explanatory model.
**Keywords:** direction of dependence; direction of effect; non-normality; linear model

## 1. Introduction

This paper introduces a unified statistical framework to discern the direction of dependence in linear models using observational data. Existing regression-type methods allow researchers to quantify the magnitude of hypothesized effects but are of limited use when establishing the direction of effects between variables, that is, whether $x \rightarrow y$ or $y \rightarrow x$ correctly describes the causal flow between two variables $x$ and $y$. When an association between $x$ and $y$ exists, at least three possible explanations can be entertained: 1) $x$ causes $y$ ($x \rightarrow y$), 2) $y$ causes $x$ ($y \rightarrow x$), and 3) neither relation exists due to a spurious association of both variables with a third variable (sometimes termed a "confounder"). The Pearson product-moment correlation and ordinary least square (OLS) estimates do not adjudicate regarding the model which best represents the data-generating mechanism. Researchers who use regression models must therefore make their decision as to the direction of effect on the basis of a priori theory and substantive arguments. However, statistical tools often are desirable to empirically demonstrate the explanatory superiority of one theory over plausible alternatives. The present contribution introduces such a tool – Direction Dependence Analysis (DDA; Wiedermann & von Eye, 2015a). While standard regression models use only estimates of first- and second order moments (i.e., means, variances, and covariances) to assess the magnitude and statistical significance of regression weights, DDA, by contrast, uses estimates of higher order moments (i.e., skewness and kurtosis) to assess the relative plausibility to directional alternatives.

## 2. Model Definitions

We start by defining the statistical models considered. Assume that a construct $\mathcal{X}$ causes construct $\mathcal{Y}$ through mechanism $\mathcal{F}$, i.e., $\mathcal{Y} = \mathcal{F}(\mathcal{X})$. Further, let $x$ and $y$ be continuous operationalizations of $\mathcal{X}$ and $\mathcal{Y}$ and define $f$ as the statistical model to approximate $\mathcal{F}$, i.e., $y = f(x)$. The direction dependence framework provides a set of statistical tools to evaluate the directionality assumption of $y = f(x)$ implied by the causal theory $\mathcal{X} \rightarrow \mathcal{Y}$. DDA assumes that the "true" predictor of the data-generating mechanism is a *non-normal* external influence. Further, we assume and that the data-generating mechanism relating the two continuous variables is recursive in nature and can be approximated by the linear model, i.e., given that $\mathcal{X} \rightarrow \mathcal{Y}$ constitutes the "true" mechanism, $y = f(x)$ can be written as (without loss of generality the intercept is fixed at zero)

$$y = b_{yx}x + e_{yx} \tag{1}$$

where the error term is assumed to be normally distributed (with zero mean and variance $\sigma^2_{e_{yx}}$), serially independent, and independent of $x$. When $\mathcal{Y} \rightarrow \mathcal{X}$ describes the causal mechanism, the corresponding linear model is

$$x = b_{xy}y + e_{xy} \tag{2}$$

where $e_{xy}$ denotes a normally-distributed error term with zero mean and variance $\sigma^2_{e_{xy}}$ which is serially independent and independent of $y$. As a third possible explanation, we consider an unconsidered confounding construct $\mathcal{U}$ (and its continuous operationalization $u$) which has a causal effect on both, $\mathcal{X}$ and $\mathcal{Y}$. The model in (1) then changes to

$$\begin{aligned} x &= b_{xu}u + e_{xu} \\ y &= b_{yx}x + b_{yu}u + e_{yx}. \end{aligned} \tag{3}$$

where "pure" confounding is included as a special case when $b_{yx} = 0$. In model (3), $u$ is assumed to be a non-normally distributed external influence and $e_{xu}$ and $e_{yx}$ are normally distributed error terms (exhibiting zero means and variances $\sigma^2_{e_{xu}}$ and $\sigma^2_{e_{yx}}$) which are independent of $u$ and of each other. In all three models, additional covariates must be known to be on the explanatory side of the statistical model. In addition, one must ensure that a recursive causal ordering of the covariates themselves is theoretically possible and that all covariates can be expressed as linear combinations of mutually independent external influences.

## 3. The Direction Dependence Principle

DDA consists of three components: 1) distributional properties of observed variables, 2) distributional properties of error terms of competing models, and 3) independence properties of error terms and predictors in competing models. Unique patterns of DDA component outcomes exist for each of the three models described in Section 2 when the "true" predictor deviates from normality.

### *3.1 DDA Component I: Distributional Properties of Observed Variables*

***Absence of Confounders***. Asymmetry properties in terms of observed variable distributions emerge from the additive nature of the linear model. Adding a normal error term to a non-normal predictor will necessarily cause the response to be more normally distributed than the predictor. Dodge and Rousson (2000, 2001) as well as Dodge and Yadegari (2010) presented algebraic proofs for this relation and showed that the Pearson correlation $\rho_{xy}$ has asymmetric properties when considering higher moments of $x$ and $y$. Specifically, one obtains

$$\rho^3_{xy} = \frac{\gamma_y}{\gamma_x} \quad \text{and} \quad \rho^4_{xy} = \frac{\kappa_y}{\kappa_x} \tag{4}$$

with skewness $\gamma_x = E[(x - E[x])^3]/\sigma^3_x$) and excess-kurtosis $\kappa_x = E[(x - E[x])^4]/\sigma^4_x - 3$ ($\gamma_y$ and $\kappa_y$ are defined in a similar fashion). Because $\rho_{xy}$ is bounded on the interval [–1, 1], absolute values of skewness and excess-kurtosis of $y$ will always be smaller than absolute skewness and excess-kurtosis values of $x$. One obtains $|\gamma_y| < |\gamma_x|$ and/or $|\kappa_y| < |\kappa_x|$ under the model $x \rightarrow y$ and $|\gamma_y| > |\gamma_x|$ and/or

$|\kappa_y|>|\kappa_x|$ under $y \rightarrow x$. An auxiliary regression approach can be used to adjust for covariates $z_j$ in which $x$ and $y$ are regressed on $z_j$ and directional decisions are based on the estimated regression residuals of the two models reflecting the (unexplained) portion of variation after adjusting for the covariates $z_j$.

***Presence of Confounders***. Any continuous non-normal confounder can affect the distributions of $x$ and $y$. Directional decisions are influenced by 1) the magnitude of non-normality of $u$, 2) the connection strength of $u$ and $x$, and 3) the connection strength of $u$ and $y$ which follows from $\gamma_x = \rho_{xu}^3 \gamma_u$, $\kappa_x = \rho_{xu}^4 \kappa_u$, $\gamma_y = \rho_{yu}^3 \gamma_u$, and $\kappa_y = \rho_{yu}^4 \kappa_u$. The influence of the confounder on direction dependence decisions is given through

$$\frac{\gamma_y}{\gamma_x} = \left(\frac{\rho_{yu}}{\rho_{xu}}\right)^3 \quad \text{and} \quad \frac{\kappa_y}{\kappa_x} = \left(\frac{\rho_{yu}}{\rho_{xu}}\right)^4. \tag{5}$$

No biases (in terms of descriptively selecting the correct model) are expected when $|\rho_{yu}|<|\rho_{xu}|$ because $|\gamma_y|<|\gamma_x|$ and $|\kappa_y|<|\kappa_x|$ still hold, which suggests the model $x \rightarrow y$. In contrast, model selection biases are likely to occur when $|\rho_{yu}|>|\rho_{xu}|$ because $|\gamma_y|>|\gamma_x|$ and $|\kappa_y|>|\kappa_x|$ increase the risk of erroneously selecting the mis-specified model $y \rightarrow x$.

***Statistical Inference***. von Eye and DeShon (2012) proposed using normality tests, such as, D'Agostino's skewness and/or Anscombe and Glynn's kurtosis test, to evaluate hypotheses compatible with observed-variable based direction dependence. Directional decisions are based on separately evaluating non-normality of predictor and response. In addition, Pornprasertmanit and Little (2012) suggested nonparametric bootstrap CIs for higher order moment differences, $\Delta(\gamma) = |\gamma_x| - |\gamma_y|$ and $\Delta(\kappa) = |\kappa_x| - |\kappa_y|$.

### *3.2 DDA Component II: Distributional Properties of Error Terms*

***Absence of Confounders***. The second DDA component focuses on the distributional shape of the error terms, $e_{yx}$ and $e_{xy}$. Wiedermann, Hagmann, and von Eye (2015) and Wiedermann (2015) showed that higher moments of the error term obtained from the mis-specified model ($e_{xy}$) can be expressed as functions of the third and fourth moments of the "true" predictor ($x$), i.e.,

$$\gamma_{e_{xy}} = (1-\rho_{xy}^2)^{3/2} \gamma_x \quad \text{and} \quad \kappa_{e_{xy}} = (1-\rho_{xy}^2)^2 \kappa_x. \tag{6}$$

Because normality of the error term is assumed in the "true" model (i.e., $\gamma_{e_{yx}} = \kappa_{e_{yx}} = 0$), differences in higher moments of $e_{yx}$ and $e_{xy}$ provide, again, information about the directional plausibility of a linear model. This DDA component can straightforwardly be extended to multiple linear regression models when adjusting for possible covariates (Wiedermann & von Eye, 2015b). The model $x \rightarrow y$ is preferred when $|\gamma_{e_{xy}}|>|\gamma_{e_{yx}}|$ and/or $|\kappa_{e_{xy}}|>|\kappa_{e_{yx}}|$. Conversely, $y \rightarrow x$ is more likely to hold when $|\gamma_{e_{xy}}|<|\gamma_{e_{yx}}|$ and/or $|\kappa_{e_{xy}}|<|\kappa_{e_{yx}}|$.

***Presence of Confounders***. When an unmeasured confounder is present, the two competing models can be written as $y = b'_{yx} x + e'_{yx}$ and $x = b'_{xy} y + e'_{xy}$ where $b'_{yx}$ and $b'_{xy}$ are biased estimates of $b_{yx}$ and

$b_{xy}$. Higher moments of $e'_{yx}$ and $e'_{xy}$ depend on the magnitude of non-normality of $u$ and the magnitudes of $b_{xu}$ and $b_{yu}$. Specifically, higher moments can be written as functions of semi-partial correlations and higher moments of $u$. That is, for $e'_{yx}$ and $e'_{xy}$ one obtains

$$\gamma_{e'_{yx}} = \rho^3_{y(u|x)}\gamma_u \quad \text{and} \quad \kappa_{e'_{yx}} = \rho^4_{y(u|x)}\kappa_u, \tag{7}$$

$$\gamma_{e'_{xy}} = \rho^3_{x(u|y)}\gamma_u \quad \text{and} \quad \kappa_{e'_{xy}} = \rho^4_{x(u|y)}\kappa_u, \tag{8}$$

with $\rho_{y(u|x)} = (\rho_{yu} - \rho_{xy}\rho_{xu})/\sqrt{1-\rho^2_{xy}}$ being the semi-partial correlation coefficient for $y$ and $u$ given $x$ and $\rho_{x(u|y)} = (\rho_{xu} - \rho_{xy}\rho_{yu})/\sqrt{1-\rho^2_{xy}}$ being the semi-partial correlation between $x$ and $u$ given $y$. The distribution of both error terms will be close to normality and no distinctive decision is possible when $u$ is close to normality and/or semi-partial correlations are close to zero. If the confounder is sufficiently non-normal, distributional properties of error terms and, thus, directional decisions depend on the magnitude of the semi-partial correlations. Unbiased directional decisions are possible when $|\rho_{y(u|x)}| < |\rho_{x(u|y)}|$ because $|\gamma_{e'_{xy}}| > |\gamma_{e'_{yx}}|$ and $|\kappa_{e'_{xy}}| > |\kappa_{e'_{yx}}|$ which implies $x \to y$. In contrast, if $|\rho_{y(u|x)}| > |\rho_{x(u|y)}|$ then erroneously selecting $y \to x$ is likely to occur because $|\gamma_{e'_{xy}}| < |\gamma_{e'_{yx}}|$ and $|\kappa_{e'_{xy}}| < |\kappa_{e'_{yx}}|$.

***Statistical Inference***. Again, non-normality tests can be used to separately evaluate distributional properties of model residuals (Wiedermann et al., 2015). An asymptotic significance test and bootstrap CIs for the skewness difference of residuals, $\Delta(\gamma_e) = |\gamma_{e_{xy}}| - |\gamma_{e_{yx}}|$, have been proposed by Wiedermann et al. (2015) and Wiedermann and von Eye (2015b). The asymptotic test requires normality of the "true" error term. Only error symmetry is required for the bootstrap approach. Analogous procedures for the difference in excess-kurtosis values were discussed by Wiedermann (2015).

### *3.3 DDA Component III: Independence Properties of Predictor and Error Term*

***Absence of Confounders***. The independence assumption in the linear model implies that the magnitude of the error made when fitting the response is not related in any form to the predictor(s). In OLS regression, estimated residuals will be linearly uncorrelated with the predictor(s) by definition. However, when the "true" predictor $x$ is non-normal, the error term and the predictor of the mis-specified model, $y$ and $e_{xy}$, will be *stochastically non-independent*. The error term of the mis-specified model in (2) can be expressed as $e_{xy} = x - b_{xy}y = (1-\rho^2_{xy})x - b_{xy}e_{yx}$ from which follows that the "true" predictor $x$ and the "true" error term $e_{yx}$ contribute to $e_{xy}$ and $y$ in (1). Thus, stochastic non-independence will hold when $x$ deviates from normality according to the Darmois-Skitovich theorem (Darmois, 1953; Skitovich, 1953). Because independence is assumed in the correctly specified model, direction dependence statements are possible through separately evaluating independence in competing models (Entner et al., 2012; Shimizu et al., 2011; Wiedermann & von Eye, 2015a). If $H_0: x \perp e_{yx}$ is retained and, at the same time, $H_0: y \perp e_{xy}$ is rejected, then it is more likely that the observed effect transmits from $x$ to $y$. Conversely, if $H_0: x \perp e_{yx}$ is rejected and $H_0: y \perp e_{xy}$ is retained, then the model $y \to x$ should be preferred. Covariates can straightforwardly be included in the models (1) and (2) provided that covariates fulfill the requirements described above.

***Presence of Confounders***. When confounding affects the relation between $x$ and $y$, predictor(s) and errors of both models contain information of the confounder, i.e.,

$$e'_{yx} = [b_{yu} + (b_{yx} - b'_{yx})b_{xu}]u + (b_{yx} - b'_{yx})e_{xu} + e_{yu} \tag{9}$$

$$e'_{xy} = [b_{xu} - b'_{xy}(b_{yu} + b_{yx}b_{xu})]u + (1 - b'_{xy}b_{yx})e_{xu} + b'_{xy}e_{yu} . \tag{10}$$

Thus, through re-considering the "true" model given in (3) and, again, making use of the Darmois-Skitovich theorem, one concludes that the independence assumption is likely to be violated in *both* candidate models whenever a non-normal confounder is present.

***Statistical Inference***. Significance tests to evaluate non-independence of (linearly uncorrelated) variables have extensively been discussed in signal processing (Hyvärinen, Karhunen & Oja, 2001). One possible class of tests uses the basic definition of stochastic independence, $E[g_1(\upsilon_1)g_2(\upsilon_2)] - E[g_1(\upsilon_1)]E[g_2(\upsilon_2)] = 0$ for any continuous unbounded functions $g_1$ and $g_2$. Independence tests can be constructed using correlation tests of the form $cor[g_1(x), g_2(e_{yx})]$ and $cor[g_1(y), g_2(e_{xy})]$ where at least one function is non-linear. Two non-linear functions may be of particular value in the present context: the square function, $g(\upsilon) = \upsilon^2$, and the hyperbolic tangent function, $g(\upsilon) = \tanh(\upsilon)$. The square function constitutes a powerful candidate because covariances of predictor and error in the mis-specified model contain information of higher moments of the "true" predictor (Wiedermann & von Eye, 2015, 2016). $\tanh(\upsilon)$ is the derivative of the log-density of an inverse hyperbolic cosine distribution which provides an approximation of the likelihood ratio of directionally competing models in the bivariate case. Alternatively, the Hilbert-Schmidt Independence Criterion (HSIC; Gretton et al., 2008) can be used. The HSIC evaluates the independence of functions of random variables and is provably omnibus in detecting any dependence between two random variables in the large sample limit (for a discussion of the HSIC in the context of the linear regression model see Sen & Sen 2014).

## 4. Model Selection

Reconsidering possible outcomes of the three DDA components, it becomes evident that each model in Section 2 can be uniquely identified through specific DDA-component patterns. In general, DDA model selection requires the specification of a *target* and an *alternative* model. While the selection of whether $x \to y$ or $y \to x$ serves as the target model is arbitrary in terms of model comparison, we suggest that the target model reflects the substantive causal theory of interest and that the alternative model reflects the contradicting theory. The target model, e.g., $x \to y$, finds support when 1) the distribution of the response $y$ is closer to normality than the distribution of $x$, 2) the residual distribution of $x \to y$ is closer to normality than the residuals of $y \to x$, and 3) the independence assumption of residuals and predictor(s) holds for $x \to y$ and is, at the same time, violated for model $y \to x$. Independence *must hold* for $x \to y$ and the independence assumption *must be violated* for $y \to x$ to conclude that an effect is transmitted from $x$ to $y$. Otherwise, one has to conclude that unmeasured confounders are present whenever the independence assumption is either violated or satisfied in both models (the latter emerges from the fact that confounders can decrease skewness/excess-kurtosis of $x$ and $y$ to a degree that renders non-independence no longer detectable).

## 5. Conclusions

DDA allows to test hypotheses compatible with the directional relation between pairs of variables while adjusting for covariates that possibly contribute to the causal process. This empirical falsification approach is based on the translation of a substantive causal theory into a linear target model which is then compared with the corresponding alternative model. DDA component patterns can then be used to either retain the target model, retain the directionally competing model, or conclude that unmeasured confounders are present. Here, it is important to re-iterate that directional conclusions derived from DDA component patterns are based on the operationalization of latent constructs $\mathcal{X}$ and $\mathcal{Y}$ using the linear model as an approximation of an unknown "true" functional relation $\mathcal{F}$. Trustworthiness of DDA, thus, ultimately depends on both, the quality of operationalization and the validity of the linear model for the description of the causal mechanism. Although both requirements essentially apply to any linear modeling approach, they deserve particular attention in the context of DDA.

**References**

Darmois, G. (1953). Analyse générale des liaisons stochastique. *Review of the International Statistical Institute*, *21*, 2-8.

Dodge, Y., & Rousson, V. (2000). Direction dependence in a regression line. *Communications in Statistics: Theory and Methods, 29*, 1957-1972.

Dodge, Y., & Rousson, V. (2001). On asymmetric properties of the correlation coefficient in the regression setting. *The American Statistician, 55*, 51-54.

Dodge, Y., & Yadegari, I. (2010). On direction of dependence. *Metrika*, *72*, 139-150.

Entner, D., Hoyer, P. O., & Spirtes, P. (2012). Statistical test for consistent estimation of causal effects in linear non-Gaussian models. *Journal of Machine Learning Research: Workshop and Conference Proceedings*, *22*, 364-372.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., & Smola, A. J. (2008). A kernel statistical test of independence. *Advances in Neural Information Processing Systems, 20*, 585-592.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent components analysis*. New York, NY: Wiley & Sons.

Pornprasertmanit, S., & Little, T. D. (2012). Determining directional dependency in causal associations. *International Journal of Behavioral Development*, *36*, 313-322.

Sen, A., & Sen, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika, 101*, 927-942.

Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., & Bollen, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, *12*, 1225-1248.

Skitovich, W. P. (1953). On a property of the normal distribution. *Doklady Akademii Nauk SSSR*, *89*, 217-219.

von Eye, A., & DeShon, R. P. (2012). Directional dependence in developmental research. *International Journal of Behavioral Development*, 36, 303-312.

Wiedermann, W. (2015). Decisions concerning the direction of effects in linear regression models using the fourth central moment. In M. Stemmler, A. von Eye & W. Wiedermann (Eds.), *Dependent data in social sciences research: Forms, issues, and methods of analysis* (pp. 149-169). New York: Springer.

Wiedermann, W., & von Eye, A. (2015a). Direction-dependence analysis: A confirmatory approach for testing directional theories. *International Journal of Behavioral Development*, *39*, 570-580.

Wiedermann, W., & von Eye, A. (2015b). Direction of effects in multiple linear regression model. *Multivariate Behavioral Research*, *50*, 23-40.

Wiedermann, W., Hagmann, M., & Eye, A. (2015). Significance tests to determine the direction of effects in linear regression models. *British Journal of Mathematical and Statistical Psychology*, *68*, 116-141.

Wiedermann, W., & von Eye, A. (2016). Directionality of effects in causal mediation analysis. In W. Wiedermann and A. von Eye (Eds.), *Statistics and causality: Methods for applied empirical research* (pp. 63-106). Hoboken, NJ: Wiley and Sons.