<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Psychological methods*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Falkenström, F. (2021)

Using Copulas to Enable Causal Inference from Non-Experimental Data: Tutorial and Simulation Studies

Psychological methods

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-108219>

Accepted for publication in Psychological Methods 20210503

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/met0000414

Using Copulas to Enable Causal Inference from Non-Experimental Data:

Tutorial and Simulation Studies

Fredrik Falkenström

Department of Psychology, Linnaeus University, Sweden

Department of Behavioral Sciences and Learning, Linköping University, Sweden

Sungho Park

SNU Business School, Seoul National University, South Korea

Cameron N. McIntosh

Employment and Social Development Canada, Canada

Corresponding author: Fredrik Falkenström, Department of Psychology, Linnaeus University, Sweden. Email: Fredrik.Falkenstrom@lnu.se, Phone: +46 470-70 86 65

Abstract

Causal inference in psychological research is typically hampered by unobserved confounding. A copula-based method can be used to statistically control for this problem without the need for instruments or covariates, given relatively lenient distributional assumptions on independent variables and error terms. The current study aims to: (1) provide a user-friendly introduction to the copula method for psychology researchers; and (2) examine the degree of non-normality in the independent variables required for satisfactory performance. A Monte Carlo simulation study was used to assess the behavior of the copula method under various combinations of conditions (sample size, skewness of independent variables, effect size, and magnitude of confounding). In addition, an applied example from research on the effects of parental rearing on adult personality and life satisfaction was used to illustrate the method. Simulations revealed that the copula method performed better at higher levels of skewness in the independent variables, and that the impacts of lower skewness can be offset to some extent by larger sample size. When skewness and/or sample size is too small, the copula method is biased towards the uncorrected model. In the applied example, parental rejection/punishment predicted less adaptive personality and life satisfaction, with no evidence of confounding. For parental control/overprotection, there was evidence that confounding attenuated the estimated relationship with personality/life satisfaction. Copula adjustment is a promising method for handling unobserved confounding. The discussion focuses on how to proceed when assumptions are not quite met, and outlines potential avenues for future research.

Translational Abstract

Psychology researchers are accustomed to the statement “correlation does not imply causation”, given the possibility of spurious relationships due to unobserved causes. To mitigate this problem, researchers typically try and measure a wide variety of additional covariates or “control” variables that are related to both the focal independent and dependent variables, and/or employ “instrumental” variables that are assumed to be uncorrelated with all unmeasured causal influences. This paper describes an alternative method for protecting against spurious relationships that does not require any external variables: the use of a copula function to directly capture the dependency between the modeled independent variables and unobserved causes. After providing an introduction to the copula method, we use Monte Carlo simulations to assess its behavior under various combinations of conditions (i.e., sample size, skewness of independent variables, effect size, and magnitude of confounding). In addition, we apply the approach in an illustrative example from research on the effects of parental rearing on adult personality and life satisfaction. The simulations revealed that the copula method performed better at higher levels of skewness in the independent variables, and that the impacts of lower skewness can be offset to some extent by larger sample size. When skewness and/or sample size is too small, the results of the copula correction are biased in the direction of the uncorrected results. In the applied example, parental rejection/punishment predicted less adaptive personality and life satisfaction, with no evidence of confounding. For parental control/overprotection, there was evidence that confounding attenuated the estimated relationship with personality/life satisfaction. Copula adjustment is a promising method for tackling the ubiquitous problem of unobserved confounding in observational and quasi-experimental research. The discussion focuses on how to

proceed when assumptions of this method are not quite met, and outlines avenues for future research.

Keywords: Confounding, Omitted Variable Bias, Instrumental Variables, Causal Inference, Copula

Using Copulas to Enable Causal Inference from Non-Experimental Data:

Tutorial and Simulation Studies

A Brief Conceptual Overview

Scientific research is principally focused on finding causal relationships, in order to help inform the development and evaluation of policies, programs, and interventions that can push the right “levers” to bring about positive societal and individual change. Therefore, it is crucial to ensure that research design and analysis are sufficiently rigorous to allow for causal interpretation of results. Almost from its inception, the field of psychology has been highly influenced by medical research, in which randomized experiments are the norm. Although experiments are straightforward to conduct in some subdisciplines of psychology (e.g., cognitive and perceptual processing), research questions in areas such as clinical, social, personality, educational, and developmental psychology are often unethical or simply infeasible to address using random assignment. In such cases, researchers are typically limited to analyzing relationships among variables based on observational data (Hernán & Robins, 2020; Rosenbaum, 2017).

Unfortunately, as is well known, there are several problems with trying to infer causality from observational data. Traditionally, the following three conditions are required for making causal claims: 1) the presumed cause must precede its effect temporally; 2) a cause must covary with its effect beyond what can reasonably be expected by chance; and 3) a cause – effect relationship must not be explained by other causes (e.g., Antonakis et al., 2010). Of these conditions, the last is usually the one that creates the most difficulties to meet in non-experimental research. Temporality (condition 1) can be investigated using time-lagged longitudinal designs, although getting the right time-lag can be difficult (Vaisey & Miles, 2014).

Beyond-chance covariation (condition 2) can be checked with statistical testing. However, ruling out alternative explanations is always a challenge, given that the researcher may not even know all possible additional causal influences that may be confounded with the variable of interest. This is referred to as the risk of spurious findings due to unobserved confounding, that is, omitted variable bias (OVB; Imbens & Rubin, 2015; Pearl & MacKenzie, 2018).

For example, say that a researcher wants to examine the causal effect of student motivation on academic achievement. For both ethical and logistical reasons, she cannot randomize students into groups and then manipulate their levels of motivation, so she simply measures naturally-occurring individual levels of motivation using a self-report scale, links these data to administrative records on final grades, and applies multiple regression to estimate the purported motivation → achievement relationship. Whatever the results, it would not be possible to make causal attributions about the estimated relationship due to the non-experimental nature of the design. More specifically, the researcher would be unable to guarantee that other, unmeasured variables related to both motivation and achievement (i.e., general intelligence, personality traits, attitudes toward education, etc.) have been adequately controlled for. Such variables will in the present paper be referred to as omitted variables, and the bias that they introduce in analyses as Omitted Variable Bias (OVB). In an experiment, these types of variables, even though unobserved, would be balanced across groups (i.e., levels of motivation) due to random assignment, and therefore not lead to any bias in the estimation of the causal effect. However, in the observational case, one cannot rule out the possibility that those with higher levels of motivation are also more intelligent, more conscientious, and have more positive attitudes toward education than those with lower degrees of motivation, making it impossible to identify the unique causal impact of motivation on achievement. Thus, OVB is a potential

problem with all kinds of observational data (whether in the form of cross-sectional, time-series, or panel data).

It can be said that OVB is the core problem of causal inference. Of the criteria for causal inference mentioned, criterion 3) is explicitly about OVB, but criterion 2) can also be reformulated as an omitted variable problem. If variables X and Y are measured simultaneously, traditional statistical methods (e.g., regression analysis) cannot determine the direction of causality. It may be that Y causes X ($Y \rightarrow X$) rather than the other way around. In this case, the outcome variable Y itself can be said to be the omitted variable that causes OVB through its effect on the predictor. A more complex version is the simultaneity problem (Wooldridge, 2010), i.e., X and Y may simultaneously cause each other ($X \rightleftarrows Y$). Still, the same can be said, Y creates OVB through its effect on X . Likewise, if the outcome variable at an earlier timepoint (Y_{t-1}) is the true cause of both X_t and Y_t , then Y_{t-1} is the omitted variable causing OVB. Thus, problems of temporality are essentially problems of OVB.

The most commonly-used statistical approach for addressing OVB is to measure a wide variety of theoretically-relevant variables in addition to the focal causes, and include them as covariates in the statistical model in order to adjust the estimates of the causal effects (see Becker et al., 2016; Bernerth et al., 2017; VanderWeele, 2019). Another method that has recently begun making inroads in psychological research is Instrumental Variable (IV) regression (e.g. Bollmann et al., 2019; Crits-Christoph et al., 2020; Maydeu-Olivares et al., 2020). IV regression uses information from variables that are not included in the focal structural equations but are used as instruments to “purify” the endogenous regressors of their relationships with omitted confounders. This method can in principle yield correct estimates of causal effects even when the researcher neither knows what the confounders actually are nor has any data on those variables.

However, as discussed later on in the current paper, both covariate adjustment and IV regression rely on strong assumptions that are difficult to satisfy in practice.

Due to the limitations of traditional methods of controlling for OVB, methodologists have recently developed some innovative alternatives. One particularly promising approach that requires neither covariates nor instrumental variables is the copula method (Park & Gupta, 2012), which instead utilizes assumptions about the (joint) distribution of regressors and error terms in order to estimate causal effects in the presence of unobserved confounding. Park and Gupta's initial simulation studies showed that the copula method can provide unbiased estimates of causal effects, even with relatively small samples.

Although the copula method for addressing OVB is a worthy contender to approaches such as covariate adjustment and IV regression – especially given that it does not require any external information for implementation – it has, to our knowledge, not been used in psychological research studies. Thus far, application of the method seems to be limited to the areas of business and economics (Christopoulos et al., 2019; Datta et al., 2017; Hult et al., 2018; Tran & Tsionas, 2015; Yu & Jaenicke, 2020). In addition, there remain some open questions about the performance of the method under different conditions. In particular, the method assumes that the regressors are non-normally distributed, but no simulation studies have been conducted to examine the extent to which estimation and inference potentially deteriorate under deviations from this condition.

Importance in Application

Due to practical and ethical problems with conducting randomized trials, as well as threats to both the internal and external validity of these designs (e.g., Deaton & Cartwright, 2018), there is an urgent need for stringent alternative methods that enable causal inference

without experimental design. As we have seen, the problem of OVB is the core issue that needs to be adequately addressed before causal inference can be made. However, most existing methods for adjusting for OVB (e.g., IV regression) are limited due to strong assumptions that make them difficult to apply and/or sensitive to bias in most psychology research contexts. The copula method is based on relatively mild assumptions, and when its assumptions are met it handles, in principle, all potential threats to causal inference.

Purposes of the Present Study

With these issues in mind, the purpose of the present paper is twofold. First, we attempt to give a user-friendly introduction to and tutorial on the copula correction method, in order to facilitate the adoption of the approach by psychology researchers. Second, to provide additional evidence on the empirical performance of the method, we conduct a set of Monte Carlo simulation experiments to investigate the degree of non-normality in the regressors that is needed to obtain reliable estimates, in the context of varying sample sizes, causal effect sizes, and magnitudes of OVB. Before introducing the copula method and our simulation experiments, we provide an overview of the nature of the OVB problem and traditional methods for addressing it.

The Problem of Omitted Variable Bias

In this section, we will illustrate the problem of OVB using simple path diagrams and tracing rules (see Chen et al., 2018; Cinelli & Hazlett, 2020; Pearl, 2013). The path diagrams in Figure 1 depict simple instances of third-variable confounding and its consequences for the estimation of causal effects. The diagrams a), b) and c) at the top of Figure 1 under the heading “True causal model” all depict variants on a situation where a causal variable (X) and an outcome variable (Y) are both observed, but are jointly influenced by an unobserved third variable (OV), which is unknown to the researcher. All three models share the same causal

structure, but with different standardized population effect sizes for the pathways, in order to help illustrate here the different types of misinterpretations that might result from unobserved confounding.

The diagrams in the middle of Figure 1, under the heading “Unadjusted model”, show the specific statistical biases that will arise when the researcher estimates the $X \rightarrow Y$ relationship under ignorance of OV . In case a), and assuming sufficient statistical power, the impact of ignoring OV would be a Type I error. More specifically, although the true causal effect of X on Y is zero, the omission of OV renders it non-zero due to the effects of OV being “channeled” through the $X \rightarrow Y$ path, as can be shown via path tracing: $0.50 (OV \rightarrow X) \times 0.50 (OV \rightarrow Y) = 0.25^1$. Turning now to case b), it is evident that ignoring OV will lead to a Type II error when regressing Y on X : $0.25 (X \rightarrow Y) + 0.50 (OV \rightarrow X) \times -0.50 (OV \rightarrow Y) = 0.25 - 0.25 = 0$. Finally, case c) shows how OVB can even reverse the direction of a causal effect: $0.25 (X \rightarrow Y) + 0.71 (OV \rightarrow X) \times -0.71 (OV \rightarrow Y) = 0.25 - 0.50 = -0.25$.

When there is OVB, the effects of OV on Y are encapsulated in the structural error term for Y (ε_Y). For OVB to occur, OV also needs to be correlated with regressor X . This means that when there is OVB, X will invariably be correlated with ε_Y , a condition that has been termed “endogeneity” (e.g., Wooldridge, 2010). Endogeneity can occur for other reasons than OVB, such as measurement error and simultaneity (although as mentioned, these can be seen as specific instances of OVB). This means that one can re-specify the omitted variable problem as per diagrams a)-c) at the bottom of Figure 1, under the heading “Respecified causal model”, that

¹ This and the bias illustrated in the other examples can also be verified through Monte Carlo simulations, where the data are first generated by specifying standardized population effect sizes for all three causal paths, followed by fitting the misspecified model (excluding OV) to the simulated population data for X and Y . In the online supplemental materials, R code for conducting such simulations for the three examples included in this section is available. The code can easily be expanded to examine potential omitted variable biases in more complex scenarios.

is, by replacing OV with the endogeneity parameter; the correlation between X and ε_Y . When the path coefficients are standardized, as in Figure 1, the size of this correlation is the product of the $OV \rightarrow X$ and $OV \rightarrow Y$ paths, respectively. Specifically, in re-specifying diagram a), we have $0.50 \times 0.50 = 0.25$; and in diagram b), $0.71 \times -0.71 \approx -0.50$. If the endogeneity parameter could be directly estimated, that would solve the problem of OVB, since the endogeneity parameter subsumes the combined influence of any number of omitted confounders. The problem, however, is that the respecified models shown at the bottom of Figure 1 are not identified, that is, there is no unique solution for the unknown parameters.

Given the potential for serious statistical errors under OVB, and corresponding negative real-world consequences for policies, programs and treatments based on the results, it is critical to take preventative measures against this insidious problem.

Statistical Methods for Protecting against OVB

Covariate Adjustment

The most popular statistical approach to protect against OVB is covariate adjustment, where the researcher measures a wide variety of additional variables believed to be related to both the focal causes and the outcome, and includes them in the model to remove (partial out) any spuriousness from the estimated effects of interest. This approach also encompasses a wide variety of techniques commonly used to evaluate the impacts of interventions or exposures, such as inverse probability weighting (Tao & Fu, 2019; Thoemmes & Ong, 2015), propensity score matching (Oyenubi, 2020; Wu et al., 2020; Zhao, 2019), and entropy balancing (Tübbicke, 2020; Zhao & Percival, 2016), all of which use observed covariate information in different ways to ensure correct estimation of causal effects.

Covariate adjustment approaches are effective when all relevant confounders are known to the researcher, or more precisely, when she has pinpointed a core subset of these that: (1) is sufficient for blocking or intercepting the influences of unobservables (Cheng et al., 2020; Cummiskey et al., 2020; Gultchin et al., 2020; van der Zander et al., 2019; Witte et al., 2020); and (2) can be reliably measured (Lockwood & McCaffrey, 2019; Sengewald et al., 2019). Intuitively, the preferred strategy would seem to be an “everything but the kitchen sink” approach, in which as many potential confounders as possible are measured and included in the model as covariates. However, covariate adjustment must be performed judiciously – using a rigorous combination of theory and graphical methods to determine whether a given adjustment set is appropriate – or the estimated causal effects may be even farther off the mark than they would have been in the unadjusted case, a phenomenon known as *bias amplification* (Pearl, 2011; Richardson et al., 2019; Stokes et al., 2020; VanderWeele, 2019). However, even with careful covariate selection methods, adequate adjustment sets are difficult to find, as full knowledge of the true causal process (including the unobserved variables and their roles) is required, and this is typically not known to the researcher.

Instrumental Variable Regression

Another means of dealing with OVB is Instrumental Variable (IV) regression, which can be used alone or in combination with covariate adjustment to tackle unmeasured confounding. Although invented nearly a century ago (Stock & Trebbi, 2003), and currently ubiquitous within its “mother” discipline of econometrics (Wooldridge, 2010), it is mostly within the last decade that IV regression has appeared in the applied and methodological literature in psychology (e.g. Bollmann et al., 2019; Crits-Christoph et al., 2020; DeMaris, 2014; Gennetian et al., 2008; Maydeu-Olivares et al., 2020). IV estimation is typically implemented through two-stage least

squares regression, in which the endogenous regressors are first modeled as functions of the instruments, followed by using the predicted values of the regressors to model the outcome variable. The instruments need to be completely independent of any omitted confounders, an assumption that needs to be justified from theory or design. This means that the values of an endogenous regressor that have been predicted from the instrument are also independent of any omitted confounders. It is also possible to estimate the full IV model simultaneously using the maximum likelihood technique (Maydeu-Olivares et al., 2020). In contrast to covariate adjustment, IV regression can recover the causal effect even in the presence of confounders that are unmeasured, and perhaps not even known to the researcher.

The IV method is based on the following three strong assumptions: (1) *relevance* (i.e., the instruments must be highly correlated with the regressors); (2) *orthogonality* (i.e., the instruments must be completely independent of all unmeasured confounders of the regressor → outcome relationship); and (3) *exclusion* (i.e., the instruments can only affect the outcome variable through the focal regressors, akin to the perhaps more widely-understood notion of full mediation; MacKinnon et al., 2020). The latter two assumptions distinguish instruments from covariates and, when satisfied, allow for better handling of unmeasured confounding than can be achieved under covariate adjustment alone.

In practice, however, IV assumptions can be difficult to meet, particularly the orthogonality and exclusion conditions (unless the instruments themselves are randomized; Sajons, 2020), and therefore IV regression can be a rather risky option. In fact, IV regression can yield even more biased results than standard methods (e.g., ordinary least squares regression) when the orthogonality and/or exclusion conditions are violated (Murray, 2006; Muthén et al., 2016), particularly when instruments are not highly relevant (Andrews et al., 2019; Crown et al.,

2011; Harding, 2016; Kim & Pohlmeier, 2016), thereby potentially making the cure worse than the disease. To be sure, a number of techniques have been developed to help mitigate such issues, such as methods for accommodating weak instruments (Andrews et al., 2019; Han, 2020), sensitivity analyses for determining the potential impacts of violations of orthogonality and/or exclusion on inferences (van Kippersluis & Rietveld, 2018; Wang et al., 2018), and even relaxing the exclusion condition entirely (Baum & Lewbel, 2019). Further, the validity of IVs becomes more defensible when they are conditioned on covariates (Chalak & White, 2011; White & Chalak, 2010; Zander et al., 2015), but again, the adjustment variables must be carefully chosen or bias amplification could result (Deuchert & Huber, 2017).

Instrument-free Methods

Due to the difficulties in finding suitable instruments and/or appropriate covariate adjustment sets, several methodologists have tried to develop methods for adjusting for OVB without the need for any external variables. For example, the Latent IV method decomposes an endogenous regressor into a latent variable and an error term (regressor error) capturing omitted confounders (Ebbes et al., 2005; Lewbel, 1997; Rigobon, 2003). Due to unobserved confounding, this regressor error is assumed to be correlated with the error term in the main structural equation (structural error). However, the derived latent IV is assumed to be exogenous, that is, uncorrelated with both the structural error and the regressor error. Different variants on this method rely on different assumptions on the structure of the latent IVs for identification of the causal effects. Specifically, the latent IVs are assumed to be either skewed (Lewbel, 1997), heteroscedastic (Rigobon, 2003), or discrete (Ebbes et al., 2005). More recently, Dzhumashev and Tursunalieva (2019) have developed a method that constructs valid instruments from information on both the endogenous regressor and the outcome variable, given assumptions

about the directions of both: (1) the relationship between the endogenous regressor and the outcome variable; and (2) the relationship between the endogenous regressor and the error term in the structural equation.

In addition, sensitivity analysis is an option for examining the potential impacts of confounding without bringing in additional variables. As mentioned, the endogeneity correlation cannot be directly estimated, as such a model is not statistically identified. This identification problem can be circumvented by using sensitivity analysis, which involves specifying plausible values for the correlation between the endogenous regressor and the structural error, thereby allowing the researcher to examine how large OVB would need to be in order to substantially invalidate the study findings (Franks et al., 2019; Gustafson & McCandless, 2018; VanderWeele, 2015). For instance, if a researcher estimates the effect of X on Y to be 0.40 (without any adjustment for possible OVB), the size of the correlation between X and the structural error in Y would need to be 0.40 to make the true effect of X on Y equal to zero, meaning that the standardized effects of any omitted variable on both X and Y need to be about .63 (or, for instance, .80 and .50), which is quite strong. Of course, it is not necessary for the omitted variable to explain the entire $X \rightarrow Y$ relationship to overturn the conclusions, rather, it may be enough that the focal causal effect becomes so small under the chosen sensitivity parameters that it is practically insignificant (even though it may still be statistically significant), and therefore not useful for informing theory or interventions. Moreover, in practice the number of omitted variables may be many. Although as stated, the endogeneity parameter captures the combined influence of any number of omitted variables, it may be more difficult to argue for what is a reasonable effect size for the endogeneity parameter, if this is a combination of the effects of many variables – some of which may reinforce each other whereas others may cancel each other

out. In any case, the researcher needs to make a reasoned argument that her selected values for the magnitude of OVB are reasonable in the research context.

Adjusting for OVB using Copulas

Having provided the background and context on the OVB problem and an overview of standard mitigation strategies, we now move on to our focal approach, namely the novel copula correction method (Park & Gupta, 2012). The copula correction for OVB is a member of the broader class of instrument-free methods, but is unique in that it does not rely on a latent variable model, synthetic IVs, or educated guesses about the magnitude of OVB. Instead, the technique specifies the dependence structure between the endogenous regressors and the structural errors using a Gaussian (or other) copula function. The OVB problem, and the more general endogeneity problem due to regressor-error dependence, can be handled without any instrumental variables if the joint distribution of the endogenous regressor(s) and the error is known. However, as we have seen, the joint distribution of X and ε_Y is unknown. Park and Gupta's (2012) copula approach tackles this problem by constructing a joint distribution function using a Gaussian copula, and then obtaining estimates of the causal effect by maximizing the joint distribution function.

The word “copula” comes from the Latin word for “link” or “tie”, that is, something that binds two things together. In statistics, a copula is a multivariate distribution function used to model the dependence structure of variables. According to Sklar's (1959) theorem, any multivariate joint distribution can be written as a combination of univariate marginal distributions and a copula, which describes the dependence structure among the variables. In other words, the copula is a function that binds the separate and distinct univariate marginal distributions to construct the joint distribution.

It can be shown (see Appendix A) that the inverse of the cumulative normal distribution of X can be calculated from the data and used to model the Gaussian copula function between X and ε_Y . Specifically, there are two ways of applying this new variable (henceforth labelled ‘ C ’): 1) by using it in a Maximum Likelihood function to construct the $X \leftrightarrow \varepsilon_Y$ correlation ρ ; or 2) including it directly as an additional regressor in a least squares regression model. Park and Gupta (2012) showed that these two methods yield the same results. Therefore, we chose method 2) for this study due to its simplicity.

Now, we briefly explain the copula method in a simple linear regression setting. A model of interest is: $Y = \beta_0 + \beta_1 X + \varepsilon_Y$, where X and ε_Y are correlated due to omitted confounders. As described earlier, the endogeneity parameter $Corr(X, \varepsilon_Y)$ is not typically estimated as part of the regression analysis, as such a model would be underidentified. Therefore, if this correlation is non-zero, the estimate of β_1 must absorb the unobserved confounding, resulting in bias. Thus an innovative solution is to estimate $Corr(X, \varepsilon_Y)$ as an “outlet” for any unobserved confounding, allowing unbiased estimation of β_1 . Specifically, the copula method does this by assuming that X and ε_Y are linked through a Gaussian (i.e., bivariate normal) copula. Then, we can rewrite the regression model as follows:

$$Y = \beta_0 + \beta_1 X + \sigma_{\varepsilon_Y} \cdot \rho \cdot \Phi^{-1}(F_X(X)) + \sigma_{\varepsilon_Y} \cdot \sqrt{1 - \rho^2} \cdot \zeta_Y,$$

where $\Phi^{-1}(\cdot)$ is an inverse cumulative distribution function of a standard normal distribution, $F_X(\cdot)$ is the cumulative distribution function of X , ρ is a copula parameter capturing the correlation between X and ε_Y , σ_{ε_Y} is the standard deviation of ε_Y , and ζ_Y is a random shock following a standard normal distribution. Note that the assumed model structure allows us to rewrite ε_Y as $\sigma_{\varepsilon_Y} \cdot \rho \cdot \Phi^{-1}(F_X(X)) + \sigma_{\varepsilon_Y} \cdot \sqrt{1 - \rho^2} \cdot \zeta_Y$. Moreover, we can obtain $\Phi^{-1}(F_X(X))$

from the data. Consequently, by adding $\Phi^{-1}(F_X(X))$ as an additional regressor (we call this the copula correction term, ‘C’), we can circumvent the endogeneity problem.

In this way, both the correlation between X and ε_Y and the true effect of X on Y , as per Figure 1, in the respecified diagrams a), b), and c) are identified and thus jointly estimable, thereby adjusting the estimates of the $X \rightarrow Y$ effect for OVB (or other causes of endogeneity). The copula method, like ordinary least squares regression, assumes that ς follows the normal distribution (cf. Tran & Tsionas, 2019). However, a key distributional requirement for the copula method is that X should *not* follow the normal distribution. This is a condition that is required for model identification.

A perhaps more intuitive explanation of the method would be to say that the variance in Y due to the causal effect of X (β_1) is correlated with X and non-normally distributed, whereas the variance in Y captured by the copula (the endogeneity parameter) is correlated with X but normally distributed. Finally, the random error (ς_Y) is independent of X and normally distributed. This way, each of these parameters can be uniquely identified and estimated separately, enabling the estimation of the causal effect without endogeneity bias. A more detailed statistical explanation of the copula method is provided in Appendix A. The copula method can be used with continuous or ordered-categorical predictors (although not binary ones) as long as they are not normally distributed, and with any kind of data (e.g., cross-sectional, time-series, or panel data).

As stated, it is important to note that although the examples in Figure 1 consider the error term ε_Y as containing only a single omitted confounder, real applications of multiple regression in psychological research will likely involve ε_Y terms that capture a combination of numerous unmeasured influences. However, this situation does not pose a problem for the copula

correction. Even if ε_Y represents thousands of relevant omitted variables, the correlation between X and ε_Y summarizes the impact of OVB, and therefore estimating this correlation along with the $X \rightarrow Y$ effect will still yield proper results, provided that the distributional assumptions under the copula method hold. Moreover, any number of endogenous predictors can be included in the same model, with copula adjustment for each. With one additional endogenous regressor, we would calculate separate copula correction terms C_1 and C_2 for each of the regressors X_1 and X_2 , and estimate the following equation: $Y = \beta_0 + \beta_1 X_1 + \beta_2 C_1 + \beta_3 X_2 + \beta_4 C_2 + \zeta_Y$.

The initial simulation studies on the performance of the copula method conducted by Park and Gupta (2012) showed that the method was robust to misspecification of the distributional form of ζ_Y , as well as to deviations from Gaussianity in the joint distribution of the endogenous regressor and ε_Y . To our knowledge, however, there have been no studies examining of the degree of non-normality (e.g., skewness) in X that is needed for the method to work properly. In psychology, it is safe to say that perfectly normally-distributed predictors are the exception rather than the rule (Micceri, 1989), so identification through a non-normality assumption can be considered quite soft relative to those involved in standard covariate adjustment- and IV-based identification. However, this is still a matter of degree. For psychology researchers interested in using the copula method in their work, it is imperative to have some empirically based guidelines regarding how much non-normality is needed in their predictors to achieve consistent estimates of causal effects, and the point at which estimation and inference start to become unreliable.

Testing for Endogeneity

It is also important to point out that regardless of one's choice of strategy for dealing with OVB, the need for such methods should be assessed, given that there are trade-offs in terms of

efficiency of the estimates. Because adjustment methods for OVB usually come with a considerable loss of efficiency, which leads to increased standard errors and consequently reduced statistical power, we do not want to apply these adjustments unless necessary. This is known to be the case for IV regression (Young, 2020), and can be a problem for the instrument-free methods such as the copula correction as well (Park & Gupta, 2012). Covariate adjustment may not necessarily involve large decrease in efficiency, but on the other hand it requires stronger assumptions regarding the population causal structure.

The Hausman (1978) endogeneity test is used to compare an estimator that is known to be unbiased with another estimator which is efficient but may be biased. This test is essentially a chi-square test, with the null hypothesis that the efficient estimator (e.g., ordinary least squares) is unbiased. If the test is significant, this is evidence that the efficient estimator is biased and should not be used. As with any statistical test, the Hausman test depends on sample size, effect size (i.e., difference between correct and biased model, or size of $X \leftrightarrow \varepsilon_Y$ correlation in the case of copula adjustment), and alpha level for its statistical power. Given that the null hypothesis is that there is no endogeneity, a non-significant but underpowered test may result in a false sense of security that the efficient estimator can be used rather than the unbiased one. Moreover, with IV regression and covariate adjustment, the test assumes valid instruments/covariates, and these assumptions are unverifiable.

Antonakis et al. (2010) showed that the Hausman specification test is equivalent to testing whether the correlation between X and ε_Y is zero. With copula adjustment, the Hausman test focuses directly on whether the estimate of this parameter is reliably different from zero. As the copula method does not require any assumptions about instruments or covariates, the only issues to contend with are whether the distributional assumptions are satisfied, and statistical

power is adequate. We will examine endogeneity testing in the context of the copula method in this paper, as well as the statistical power of the test.

Purposes of the Present Study

With these issues in mind, the purposes of the present study are to: (1) conduct Monte Carlo simulations to investigate the amount of skewness in X required for adequate performance of the copula method in estimating the causal effect; and (2) provide a tutorial on/user-friendly introduction to the copula method for psychology researchers, given that application of the technique has been largely restricted to the business and economics fields. In addition, we will consider endogeneity testing and statistical power for the copula method. In the next section, we report on our simulations examining the effects of various degrees of skewness in X on estimation and inference under the copula method. After having discussed these results on the working conditions for the copula correction, we present a step-by-step example using real data, illustrating the considerations that are needed when using this method in applied research. The paper concludes with some practical suggestions on using the copula method in psychological studies.

Simulation Study

Methods

Aims

The aims of the simulation study were to examine the performance of the copula correction method for recovering the causal effect of X on Y in the face of unobserved confounding. Specifically, we were interested in the degrees of skewness in X required for adequate performance of the copula method. Other forms of non-normality may also be used; we chose skewness as an example of a common form of non-normality.

Data-generating Mechanism

The data generating mechanism was a simple linear regression model of the form $Y = \beta_0 + \beta_1 X + \gamma OV + \varepsilon_Y$, where Y is the outcome, X is the focal predictor, β_1 is the coefficient denoting the causal effect of X on Y , and OV is an unknown variable (or a linear combination of a number of unknown variables) related to both Y and X , thus confounding their relationship when omitted from the estimating equation, with γ being the population effect of OV on Y . The data were generated in R version 3.6.2 (R Core Team, 2019), with normal distributions for both OV and ε_Y , and skewness generated in X by the use of an exponential transformation. Thus, X was first generated with a normal distribution, and then transformed according to the following formula: $X_{transformed} = sign(\lambda) \times X + e^{\lambda \times X}$, where λ is a constant controlling the amount of skewness in the transformed version of X .

Estimands

The estimated model was the same as the data generating model, except for two important differences. First, OV was omitted from the estimation model. Second, a copula correction term (C) was calculated according to the procedures shown in Appendix A and used to replace OV in the estimated model, that is, $Y = \beta_0 + \beta_1 X + \beta_2 C + \zeta_Y$. With this specification, the regression coefficient β_1 is deconfounded. When all variables are standardized, β_2 is an estimate of the endogeneity parameter ρ .

The following experimental conditions were used in a fully factorial manner:

Skewness. Values of λ in the exponential transformation were chosen so that the skewness statistics were approximately 0.40, 0.60, 0.80 ... 2.00 (i.e., nine levels of skewness). Figure 2 provides a visualization of the distributions of X at these selected skewness values.

Sample Size. The skewness conditions were crossed with varying sample sizes, $N = 100, 200, 500, 1000, \text{ and } 3000$.

Effect Size for β_1 . The effect sizes for β_1 were chosen to represent small, medium and large effects as usually defined in psychology, i.e., standardized coefficients of .10, .30 and .50 (Cohen, 1988, p. 83). Cohen stated these rules of thumb for correlation coefficients, not standardized regression coefficients. However, in bivariate regression models, standardized regression coefficients are the same as correlations. In addition, a condition with zero effect of X on Y was also included to check the empirical Type-I (i.e., false positive) error rate.

Magnitude of Endogeneity. The magnitude of OVB is likely to affect the performance of any correction method. Because what matters is the endogeneity parameter, i.e., the correlation between X and ε_Y , denoted as ρ , which as mentioned in the case of a single omitted variable is the product of $OV \rightarrow X$ and $OV \rightarrow Y$ (assuming standardized coefficients, the particular combination of $OV \rightarrow X$ and $OV \rightarrow Y$ should not matter). For ρ , we chose the values -.30, -.20, -.10, 0, .10, .20, and .30. Thus, with a single omitted variable, the largest conditions would require effects of about .55 (or -.55) from the omitted variable to X and to Y (i.e., $\sqrt{.30} \approx .55$). This corresponds to large effects (Cohen, 1988) of OV on both X and Y . The smallest values of ρ (-.10 and .10) correspond to medium-sized effects of OV on X and Y of about .32. The size of ρ is directly related to the amount of bias in the coefficient of X that will be generated if endogeneity is ignored (on a standardized scale). Thus, when $\rho = .10$, ignoring endogeneity will result in positive bias of .10. This also means that smaller ρ than .10 will generate fairly small bias, which may be ignorable in most circumstances.

In total, the number of experimental conditions was 1260 (9 skewness values $\times$ 5 sample sizes $\times$ 4 effect sizes for $\beta_1 \times$ 7 effect sizes for ρ). For each condition, 2000 random samples

were generated and analyzed. Annotated R code for the simulations is provided in the Online Supplement, with comments showing where researchers can input their own values for skewness and sample size, in order to enable them to evaluate the potential of the copula correction method to work with their own data.

Performance Measures

Mean Coefficient Bias. This was calculated as $\overline{\hat{\theta}_t} - \theta$, i.e., the mean of estimates of θ across the 2000 replications minus its true population value. This measure was used to test whether estimates converge on their true population value over many replications.

Standard Error Bias. Standard error bias was calculated as $\overline{SE(\hat{\theta})_i} - SD(\hat{\theta})_i$, i.e. the average of analytic standard errors over all Monte Carlo replications minus the Monte Carlo standard deviation, which can be considered an approximation of the population standard error (Muthén & Muthén, 2002).

Confidence Interval Coverage. Coverage concerns the proportion of replications in which the 95% confidence interval contains the true population coefficient, $\Pr(\theta_{low} \leq \theta \leq \theta_{upp})$. This was calculated by setting a variable to ‘1’ each time θ was within the confidence interval, and to ‘0’ if not. Coverage was calculated as the proportion of 1s across all replications. This proportion should be close to 95%, but preferably at least between 91-98% (Muthén & Muthén, 2002).

Root Mean Squared Error (RMSE). The RMSE, calculated as $\sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_t - \theta)^2}{N}}$, was used to calculate the expected bias in a given replication. Smaller values are desirable, as it means that the estimate from a particular study is more likely to be close to the true value.

Statistical Power for Tests of β_1 . Statistical power for the test of the estimated causal effect of X on Y was calculated as the proportion of Monte Carlo replications for which the null hypothesis that θ is equal to zero is rejected for each parameter at the .05 level (two-tailed test with a critical value of 1.96). The statistical test is the ratio of the parameter estimate to its standard error, an approximately normally distributed quantity (z-score) in large samples. Thus, each time the test yielded a p -value $< .05$ (and the effect was in the correct direction), an indicator variable was set to '1', whereas if $p \geq .05$ the same variable was set to '0'. Power was then calculated as the proportion of 1s across all replications. Usually, in the behavioral sciences a power level of $\geq 80\%$ is aimed for (Cohen, 1992).

Statistical Power of Endogeneity Tests. To be able to test for the endogeneity of the regressors, one can either 1) test the difference in the size of the coefficient for X (β_1) between the uncorrected and corrected models, or 2) test the statistical significance of the copula correction term itself (i.e., the generated regressor, β_2). Both are essentially Hausman tests, but with different parameterizations. In the current simulations, both approaches yielded highly similar estimates, although slightly higher power was observed under the second method. Therefore, we only present the results for the statistical power of the test of the copula coefficient. Statistical power was calculated in the same way as for the test of β_1 .

Meta-Models

To explore the impact of the various experimental conditions, we used a series of meta-models (Boomsma, 2013), that is, statistical models with the estimated coefficients from each of the Monte Carlo experimental conditions used as data. Standard error bias and RMSE are naturally bounded at zero and strongly skewed. Thus, we used Poisson regression analysis with robust standard errors, a method that has been recommended as an alternative to log-transformed

linear regression (Gould, 2011; Wooldridge, 2010). Mean coefficient bias is difficult to model as it is, since the best values are close to zero with large positive *and* negative values being worse.

Thus, we chose to take the absolute value of the coefficient bias to make it bounded at zero.

Meta-modeling could then be done by Poisson regression (as with SE bias and RMSE).

Coverage and statistical power are proportions that are bounded between 0 and 1, so we used Binomial regression with a logit link and robust standard errors, as recommended by Papke and Wooldridge (1996). In a first step, only main effects were tested, followed by a model testing all one-way interactions. To keep this section as simple as possible, we did not model any higher-order interactions. The effect size measures used for the meta-modeling were exponentiated coefficients for Poisson regression (which can be interpreted as proportional change coefficients), and odds ratios for Binomial regression. To enable interpretation of results, N was centered on 100 and divided by 100 so that the coefficient for N would represent the effect of an increase in sample size by 100. ρ was 1) separated into two variables, one containing negative (“ ρ^- ”) and one containing positive values (“ ρ^+ ”), 2) ρ^- was multiplied by -1 so that its coefficient would represent the effect of increased endogeneity, 3) both ρ^+ and ρ^- were multiplied by 10, so that their coefficients represent the effect of an increase in ρ of .10. β_1 was multiplied by 10, so that its coefficient would represent the effect of an increase in β_1 of .10. Skewness was centered on 0.40 and multiplied by 10 so that its coefficient would represent the effect of an increase in skewness of .10. Thus, model intercepts would represent outcome when $N = 100$, $\rho = 0$, $\beta_1 = 0$, and skewness = 0.40.

Results

Mean Coefficient Bias

As a first step, it is of interest to check whether estimates converge on the true population values on average over many replications. Mean coefficient bias was, on average across all conditions, 0.00 (SD = 0.03) ranging between -0.19 and 0.22. Thus, mean bias was small, but there was large variation across conditions. To explore the impact of experimental conditions on mean coefficient bias, we conducted a Poisson regression analysis using the absolute of mean bias as the dependent variable and N , skewness, ρ^+ , ρ^- and β_1 as predictors. The first model, which tested only main effects of all four independent variables, showed that all conditions except the size of β_1 had significant impact on bias (Pearson Goodness of Fit (1254) = 6.84, $p = 1.00$, pseudo $R^2 = .16$). There were no significant one-way interactions, so the main effects model was used for interpretation. Results showed that an increase in ρ of .10 would increase mean bias with 70.3% ($p < .001$). ρ^+ and ρ^- had highly similar effects on absolute mean bias (but bias of the raw bias was, naturally, in opposite directions for ρ^+ and ρ^-). Increasing skewness in X by 0.10 would decrease mean bias by 16.6% ($p < .001$). Finally, increasing N by 100 reduced bias by 11.0% ($p < .001$). Therefore, as predicted, larger skewness, larger N and smaller degrees of endogeneity contributed to lower mean coefficient bias.

To further illustrate these findings, Figure 3 shows the simulation results with respect to the causal effect estimates obtained under copula correction for the different sample sizes, skewness levels of X , and ρ^+ when the true value of $\beta_1 = .30$ (although as mentioned above, the magnitude of β_1 did not matter for coefficient bias). As can be seen from Figure 3, the mean estimate of β_1 under copula correction approaches the true value of .30 as both sample size and skewness increase. Small skewness and/or sample size led to estimates being biased towards the

expected value obtained from the uncorrected model (shown in Figure 3 by the dashed horizontal reference line). This was expected, because when the skewness of X approaches zero, the correlation between X and the copula correction term will approach 1.0. In this case, the Copula correction term adds no unique information to the modeling, and therefore the estimates of the causal effect are biased toward the value from the uncorrected model.

Root Mean Square Error

To estimate the amount of bias that is expected in a given study, we also have to take account of the variance of the estimator, not just the mean. Stated differently, even if an estimator converges on the true value over many replications, the variance among individual replications may be large, meaning that bias in a particular sample may be large. To this end, the RMSE is a more appropriate measure than mean coefficient bias, because it takes into account both mean bias and variation across replications. On average across all conditions, the RMSE was 0.24 (SD = 0.17), and ranged from 0.03 to 0.84. Again, we started by estimating a Poisson regression model with RMSE as outcome, and N , skewness, ρ^+ , ρ^- and β_1 as predictors. This time, there were two statistically significant interaction effects, which were retained in the final model (Pearson Goodness of Fit (1251) = 18.25, $p = 1.00$; pseudo $r^2 = .09$). Results showed that increasing N by 100 reduced RMSE by 5% ($p < .001$) and increasing skewness by 0.10 reduced RMSE by 7%. Although there was a statistically significant interaction between N and skewness ($p = .003$), the effect of this interaction was very small. If N increased by 100, the effect of skewness on RMSE was 0.001% larger (or vice versa). Increasing β_1 by .10 reduced RMSE by about 2% ($p < .001$) when $\rho = 0$, and this effect increased (i.e., the reduction in RMSE got larger) by about 2% for each increase of .10 in ρ (interaction effect $p < .001$). Conversely, ρ had no effect on RMSE when $\beta_1 = 0$, but when β_1 increased by 0.10, the effect of increasing ρ by 0.10

was a 2% decrease in RMSE (and so on). Figure 4 illustrates the relationship between RMSE, N , skewness and ρ when $\beta_1 = .30$.

Standard Error Bias

Standard error bias was small, with the average of the analytic standard errors being 0.23 ($SD = 0.15$), whereas the Monte Carlo standard deviation was 0.24 ($SD = 0.16$) and the average absolute bias was 0.01 ($SD = 0.01$, range = 0.00 – 0.07). We again ran a Poisson regression analysis using absolute standard error bias as the dependent variable, and N , skewness, β_1 , ρ^+ and ρ^- as independent variables. There was a significant interaction between N and skewness ($p = .01$; Pearson Goodness of Fit (1253) = 6.14, $p = 1.00$, pseudo $R^2 = .08$). Results showed that all variables except ρ^+/ρ^- had a significant impact on bias. Increasing N with 100 reduced SE bias by 11.9% ($p < .001$) when skewness was 0.40; and increasing skewness with 0.10 reduced SE bias by 8.3% ($p < .001$) when $N = 100$. The interaction effect between N and skewness was small; for each increase in N of 100, the effect of skewness increased by 0.3% and vice versa. Finally, increasing β_1 by .10 reduced SE bias by 4.9% ($p < .001$).

Confidence Interval Coverage

On average, the confidence interval coverage rate for β_1 was 94.8% ($SD = 0.6\%$, range = 92.3 – 96.2%). Thus, coverage for β_1 was within the adequate range (91-98%) in all of the tested conditions. In addition, the proportion of significant estimates of β_1 when $\beta_1 = 0$ in the population (i.e., false positives) was very close to the nominal 5% level across all conditions ($M = 5.16\%$, $SD = 0.53$, range = 3.75 – 7.00%). Thus, there seem to be no increased risk for Type-I error when using the copula correction method.

Statistical Power for β_1

On average across all conditions in which $\beta_1 > 0$, statistical power for the effect of β_1 was 46.6% (SD = 36.2, range 1.4 – 100%). Thus, statistical power was fairly low on average, but varied widely across conditions. An initial binomial regression model including only the main effect of the four experimental conditions showed that all conditions significantly affected the statistical power for β_1 . As usual in power analysis, larger N (odds ratio = 1.12, $p < .001$) and effect size for β_1 (odds ratio = 2.87, $p < .001$) increased statistical power. In addition, larger skewness (odds ratio = 1.18, $p < .001$) and more endogeneity (ρ^+ odds ratio = 1.27, $p < .001$, and ρ^- odds ratio = 1.14, $p < .001$) increased statistical power. A second model including all first-order interactions revealed three statistically significant interaction effects, $N \times$ skewness (odds ratio = 1.01, $p < .001$), $N \times \beta_1$ (odds ratio = 1.07, $p < .001$) and skewness $\times \beta_1$ (odds ratio = 1.05, $p < .001$). Thus, all the interaction effects were in the direction of increasing the effects of the other variable.

Of particular interest here is that more endogeneity increased statistical power of β_1 even when ρ was negative. This may seem surprising, because: 1) β_1 was positive in all conditions, and thus ρ^- has the effect of diminishing the estimate of β_1 ; and 2) during the simulation, any estimate that may have turned out significant but in the wrong direction were not counted as significant. Taken together, this should make it more difficult to find a significant positive effect of β_1 . Indeed, coefficient bias was negative in all conditions with negative ρ . This suggests that the improved efficiency of the copula adjustment method with larger endogeneity outweighs the negative coefficient bias created by ρ being in the opposite direction of the causal effect.

Performance of Endogeneity Test

To be useful in practice, the performance of the endogeneity test is important to evaluate. We used the following three performance measures: 1) confidence interval coverage for the coefficient of the copula correction term (β_2), which as mentioned is an estimate of ρ when all variables are on a standardized scale (as in our simulation); 2) the proportion of false positives for β_2 ; and 3) statistical power for β_2 .

As with β_1 , confidence interval coverage for β_2 was excellent; the mean across all conditions was 94.8% (SD = 0.6%, range 92.7% to 96.4%). The proportion of false positives were very close to expected, with a mean of 5.0% (SD = 0.5, range = 3.6 – 6.9%). Across all conditions in which $\rho \neq 0$, the statistical power of the test of β_2 was 30.2% (SD = 30.8%, range = 2.5 – 100.0%). Thus, on average, power was low, although with great variation. Binomial regression analysis showed that all experimental conditions significantly influenced power. As expected, effect size of endogeneity strongly affected statistical power (ρ^+ odds ratio = 3.84, $p < .001$, ρ^- odds ratio = 3.85, $p < .001$), and so did larger N (odds ratio = 1.12, $p < .001$), skewness (odds ratio = 1.18, $p < .001$) and β_1 (odds ratio = 1.17, $p < .001$). The reason why larger β_1 increased power of β_2 is most likely because this decreases residual variance. There were also five significant interactions: $N \times$ skewness (odds ratio = 1.01, $p < .001$), $N \times \rho^+$ (odds ratio = 1.08, $p < .001$), skewness $\times \rho^+$ (odds ratio = 1.09, $p < .001$), $N \times \rho^-$ (odds ratio = 1.08, $p < .001$) and skewness $\times \rho^-$ (odds ratio = 1.09, $p < .001$). Again, all interactions indicated that increasing one variable would strengthen the effects of the other. Figure 5 displays statistical power as a function of N , skewness and ρ^+ when $\beta_1 = .30$. As the Figure shows, power increases when N , skewness and ρ^+ increase.

Summary and Conclusions from Simulation Study

The simulation study shows that the copula adjustment method recovered the true population values of β_1 across many replications, even at fairly small sample sizes and modest deviations from normal distribution of X . In a given study, however, larger samples and skewness levels are needed for reasonable confidence that estimates are closer to population values than in uncorrected models. As Figure 4 shows, with large skewness and endogeneity, sample sizes around 100-200 may be enough to warrant the use of the copula method, while at small skewness levels and endogeneity, copula adjustment is unlikely to be better than ordinary unadjusted regression.

The endogeneity test works very well, with excellent coverage of confidence intervals and the number of false positives very close to the nominal 5% level. However, statistical power tends to be low; for detecting smaller degrees of endogeneity sample sizes of several thousand are needed together with strongly skewed X . Comparing Figures 4 and 5, it appears that there are conditions that are likely to produce negative endogeneity tests but for which the copula method would still yield improved estimates compared to an uncorrected model. Stated differently, there is a risk for Type-II error for the endogeneity test (i.e., concluding that there is no endogeneity when in fact there is). Our recommendation to researchers then is that if the endogeneity test (i.e., the test of β_2) is significant, the copula adjusted coefficient should be preferred to the unadjusted one. However, if β_2 is non-significant, the researcher can consult our figures or run their own simulations using our R code (see Online Supplement) tailored to the specific conditions (i.e., N and skewness of X) of their own study, to determine whether the copula adjustment is likely to produce improved coefficient estimates. In this, the RMSE is the measure to examine closely. If examination of the RMSE shows that copula adjustment is likely to yield

better estimates than the unadjusted model, the copula adjusted estimates of β_1 should be preferred.

Empirical Example

We turn now to an application of the copula method to an applied example with psychological data, which serves to illustrate the kinds of research problems that the copula method can be used for, and the associated considerations that need to be taken into account. The data were obtained from a representative survey which was conducted in Eastern and Western Germany for the standardization of several psychological tests, including the Giessen-Test (Brähler et al., 1999), the Questionnaire for the Assessment of Recalled Parental Rearing Behaviour (Perris et al., 1980), and the Life Satisfaction Questionnaire (Fahrenberg et al., 2000). The dataset, which is publicly available for research purposes (Brähler, 2014), was obtained from the Leibniz Institute for Psychology Information (ZPID) and contains information on a total of 3047 individuals.

Several psychological theories assert that memory structures formed with primary caregivers during childhood, variously called internal working models of attachment (Bowlby, 1988; Main et al., 1985), schemas (Beck et al., 1979; Young, 1999) or object relations (Greenberg & Mitchell, 1983), are used as templates for close relationships later in life. Therefore, if early (i.e., parental) relationships are difficult, it is likely that the individual will develop schemas that are adaptive for handling the relationships with the parents, but are maladaptive in other contexts later in life. For instance, if parents are ‘rejecting’ with respect to their child’s needs for support and comfort, the child may develop schemas for avoiding close personal contact. Although adaptive at the moment, it may make it difficult for the child to develop satisfying intimate relationships later in life.

Research question

The research question studied was whether recalled memories of parental rearing behavior affect aspects of adult personality and satisfaction with life. Because there are a number of potential confounders of such a relationship such as common method bias (Podsakoff et al., 2003) due to all measures being interview-based, trait affectivity (Heller et al., 2002), and a virtually unlimited array of unmeasured personality, motivational, attitudinal variables (Sajons, 2020), this issue was deemed a good example for demonstrating the copula method.

Method***Data Collection***

Participants were randomly sampled from the German population using the random route method, whereby the target person to be surveyed within the household was also determined randomly. There was a 31.5% dropout rate, mostly due to either refusal to participate or to no-one being home despite two visits by the researchers.

Participants

Participants were between 14 and 92 years old and consisted of 55.8% women. Due to some missing data in particular measures, the total sample size available for analysis varied between $N = 2997$ and $N = 2999$, depending on particular combinations of independent and outcome variables.

Measures

The **Giessen-Test** (Brähler et al., 1999) is a personality test based on psychoanalytic theory, and the second most used test by German clinicians and has been used in over 1000 research publications (Holtforth (2001). The Giessen Test has been criticized because of low reliability and sample dependency (Holtforth, 2001), but given that the copula method also

adjusts for bias due to measurement error (which also induces correlations between regressors and errors and can be seen as a special case of OVB), this was not seen as an insurmountable problem. The test is arranged into six subscales, each consisting of six items: *Social Response* measures the person's sense of being attractive, respected, popular and self-assertive. *Dominance* goes from submissive, co-operative, and patient, to impatient, domineering, and argumentative. *Self-Control* is about issues of conscientiousness regarding e.g., money, truth, tidiness, and being able to let go. *Underlying Mood* measures self-criticism, suppressed anger, anxiousness and depression. *Permeability* is about trust and openness in close relationships. *Social Potency* measures issues about sociability, competitiveness, and the ability to love. In the present sample, internal consistency of the subscales varied from poor for the subscales Dominance (.40), Self-Control (.46), and Underlying Mood (.56) to reasonable for subscales Social Potency (.60), Permeability (.65) and Social Response (.68). All of the subscales were normally distributed, with a maximum skewness of -0.11 (Dominance), which makes the assumption of normally distributed error terms more likely to hold. Because the subscales were scored in different directions, some of the scales were reversed so that higher scores on all scales meant "good" and lower scores meant "bad."

The **Life Satisfaction Questionnaire** (Fahrenberg et al., 2000) measures life satisfaction in the following dimensions: health, job and profession, finances, leisure, spouse/partner, relationship to own children, self, sexuality, friends and relatives, and home. The questionnaire consists of 70 items, each rated from 1 (very dissatisfied) to 7 (very satisfied). For the purpose of the present study, we used the mean of all items, which had excellent internal consistency (.96). This questionnaire was not quite normally distributed (skewness = 0.68), so to make the

assumption of normally distributed errors more likely we used a log transformation of this variable. The transformation reduced skewness to -0.04.

The **Questionnaire for the Assessment of Recalled Parental Rearing Behavior** (Perris et al., 1980) is a questionnaire designed to assess recalled parenting styles. The questionnaire includes 24 items to be answered by mother and father separately. The response format is a Likert scale with the categories 1 (no, never), 2 (yes, sometimes), 3 (yes, often) and 4 (yes, always). The items are grouped into the following three subscales, each containing 8 items: *rejection/punishment*, *emotional warmth* and *control/overprotection*. The scales have shown moderate stability over a two-year period (Asselmann et al., 2015). As a requirement of the copula method is that the predictor is skewed, the subscale of emotional warmth had to be dropped, given that it was too close to a normal distribution (skewness = -0.19 for mother and 0.04 for father). For the purpose of the present study, the scales for mother and father were combined (they were strongly intercorrelated, $r = .70$ for rejection/punishment and $.77$ for control/overprotection). Thus, we were left with two subscales, each consisting of 16 items, which had good to excellent internal consistency ($.85$ for control/overprotection and $.92$ for rejection/punishment). The rejection/punishment subscale was strongly (1.45) skewed, and that for control/overprotection moderately (0.47) skewed.

Data analysis

To reduce the number of tests, we first inspected intercorrelations among the GT subscales. The correlations among subscales 1, 4, 5, and 6 were between $.33$ and $.59$ ($M = .43$), while subscales 2 and 3 had small or negative correlations with all the other subscales. Since internal consistency also was lowest for subscales 2 and 3, we decided to drop them and combine

the others by taking their average. The resulting index (henceforth called “Functioning”) had internal consistency $\alpha = .75$ and skewness = 0.15).

The data were analyzed using the following linear regression models:

$$Y = \beta_0 + \beta_1 X + \varepsilon_Y. \quad \text{Equation 1}$$

and

$$Y = \beta_0 + \beta_1 X + \beta_2 C + \zeta_Y. \quad \text{Equation 2}$$

Equation 1 represents the simple regression of each of the outcomes (Functioning and Life Satisfaction) on each of the predictors (rejection/punishment and control/overprotection), whereas Equation 2 also adds the copula correction term. Separate bivariate regression models were run for all combinations of the two outcome variables and two predictors. Although it is straightforward to include several endogenous predictors with copula adjustment for each in the same model, we wanted to keep the applied example in line with our simulations for this demonstration. In addition, as mentioned in the introduction, the issue of whether to adjust for covariates requires some elaboration of the hypothesized causal structure for the relationships between the predictors, and we did not have sufficient information for such an exercise.

All variables were standardized before analysis ($M = 0$, $SD = 1$), in order to enable comparison with our simulation study, and because the scales of the measures aren’t readily interpretable anyway unless one is very familiar with the tests. This is, however, not a requirement for the copula method. The models were estimated using Stata 16.1 (StataCorp, 2019).

To enable further comparison with our simulation study, we drew random sub-samples of the same sizes as in our simulation study (i.e., $N = 100, 200, 500$ and 1000) from the full sample – which happened to correspond well with the largest sample size used in our simulation study

($N = 3000$). This way, we hoped to be able to convey the instability of estimates when sample size and/or skewness is too small, as well as the increasing stabilization of estimates as sample sizes increase. The results in each sub-sample as well as in the full sample were first examined for evidence of endogeneity, that is, whether the correlation between X and ε_Y was larger than what would have been expected by chance alone. As described previously, this was done by checking the statistical significance of β_2 , the coefficient for the copula correction term. In a second step, if there was evidence for endogeneity, the copula correction term was retained, and the final estimate for β_1 was taken from Equation 2. If there was no evidence for endogeneity, the estimate for β_1 was taken from Equation 1, which excludes the copula correction term.

Results

Step 1: Checking Data for Applicability of the Copula Method

The full sample size was large ($N = 2998$), which as we saw in our simulation study increases the likelihood for the copula method to work well. The skewness of the control/overprotection predictor was moderate (0.47), large for rejection/punishment (1.45), and low for the dependent variables (increasing the likelihood that the structural errors will also have low skewness). Therefore, in light of the results from our simulation study, the rejection/punishment scale should work well for copula correction, probably down to almost $N = 200$ for adjusting the coefficient (see Figure 4). Statistical power for detecting endogeneity of $\rho = .10$ would be only just above 40% even at $N = 3000$, but for $\rho = .20$ power would be $> 80\%$ somewhere between $N = 2000$ and 3000 , and for $\rho = .30$, 80% power would be reached already between $N = 500$ and 1000 (see Figure 5). The skewness of the control/overprotection variable should work for getting closer to the true population coefficient in the full sample, at least if endogeneity is fairly strong ($\rho = .20$ or above, see Figure 4). Statistical power for β_2 would be

low for anything but large endogeneity ($\rho > .30$), even in the full sample. For the smaller sample sizes, we would not expect the copula method to work well with this skewness.

Step 2: Calculating Copula Correction Term and Estimating Models

The second step is to calculate the copula correction term (C), which is done using the R code provided in the Online Supplement. This new variable is included in Equation 2. One correction term is calculated for each (potentially endogenous) predictor. After this, Equations 1 and 2 are estimated using ordinary linear regression analysis, which can be done using any statistical software. As mentioned, we chose here to estimate one endogenous predictor at a time, although it would be possible to add two or more endogenous predictor, each with its own copula correction term, within the same regression equation.

Step 3: Testing for Endogeneity of Regressors

The third step is to test for endogeneity. Table 1 shows the estimates obtained from Equations 1 and 2 for the effect of rejection/punishment on the two outcome variables. As mentioned, the significance of β_2 is the test for endogeneity of regressors. As can be seen in Table 1, no significant endogeneity could be detected in any of the regressions involving rejection/punishment, with all p -values for β_2 being larger than .05 even in the full sample. The size of β_1 was largely unaffected by copula adjustment in the full sample, but in the smaller samples it can be seen that the estimates of β_1 differed more between Equations 1 and 2.

Somewhat surprisingly given the low power of the test for β_2 , there was evidence of significant endogeneity in the control/overprotection (Table 2) models down to $N = 500$. This may be an indication of fairly strong endogeneity for these analyses. The estimates of β_2 indicate that ρ is between .20 and .30 in the full sample but get as large as .60 in some of the smaller samples. These latter figures seem unreasonably large, and may be due to the collinearity

between predictors that happen when skewness and sample size are small. There is of course the possibility that the significant β_2 estimates are false positives, but as mentioned our simulations showed no increase in false positives when using the copula adjustment method compared to ordinary linear regression (i.e., alpha levels were close to the nominal 5% level).

Because there was no evidence for endogeneity for any of the regressions involving rejection/punishment, the estimate of β_1 from Equation 1 in Table 1 should be interpreted as the causal effect. However, for the regressions involving control/overprotection (Table 2), we would need to interpret β_1 from Equation 2. Interestingly, the estimates for β_1 are *larger* in Equation 2 than in Equation 1. Therefore, in these instances, it seems that OVB has the effect of attenuating the estimated effect of control/overprotection on personality/life satisfaction, and adjustment using the copula method results in increased estimates of these effects. Moreover, this increase seems non-trivial, with up to a three-fold increase in effect size (for functioning) after copula adjustment in the full sample (as mentioned, the very large coefficients in the smaller subsamples are likely over-estimated).

Discussion

This paper makes two key contributions. First, it introduces the copula method to psychology researchers, with a view to improving the rigor of research in this field. Most psychological research is at high-risk for OVB given that randomization is typically infeasible, so an accessible method for protecting against this bias is crucial. To be sure, we are not suggesting that the copula method completely replace existing techniques for causal inference with observational data, but rather be considered as an alternative part of the toolkit for dealing

with OVB when the assumptions of conventional approaches such as covariate adjustment and instrumental variables are difficult to defend.

Second, our simulation studies confirm the results of earlier studies (Park & Gupta, 2012; Tran & Tsionas, 2015; Tran & Tsionas, 2019) showing that the copula method works well in adjusting for OVB, and also extend prior work by pinpointing the key conditions under which the copula method performs optimally. In particular, our simulations showed that the amount of skewness in X clearly influences the performance of the copula method, with larger skewness leading to better performance both in terms of smaller coefficient bias, better efficiency (leading to smaller standard errors) and higher statistical power for testing endogeneity. We also found that to some extent, lower levels of skewness can be compensated for by a larger sample size. Conversely, when skewness is large, a smaller sample may be sufficient for the method to work. Our findings also showed that when skewness and/or sample size is too small, coefficient estimates will, on average, be biased towards the parameters of the uncorrected regression model. More specifically, as the distribution of X approaches normality, collinearity between X and the copula correction term approaches 1.0, and therefore causal identification using the copula method breaks down (i.e., there is no additional information provided by the copula correction term over and above X).

Our example analyzed the relationship between memories of relationships with caregivers and personality/life satisfaction variables. It was shown that OVB may work in any direction, not just by creating spurious relationships between variables (as is often thought) but also by attenuating causal relationships (i.e., biasing them toward zero). We observed instances of this phenomenon in the relationship between recalled parental control/overprotection on the one hand and personality functioning/life satisfaction on the other hand. In these analyses, there

was evidence of OVB, but adjusting for OVB resulted in *stronger* effects than in the unadjusted analyses. Most of the results showed significant effects of recalled parenting on adaptive personality characteristics/life satisfaction, and these effects were either unaffected by OVB or were increased by adjusting for OVB. This lends some support to developmental psychology theories about the importance of internal working models of attachment and adaptive personality/satisfaction with life (e.g. Bowlby, 1988).

Recommendations for Applied Users

Because OVB is a potential problem in most fields of psychological research (aside from cases using stringent randomized experiments), the copula method can be a useful tool for assessing the degree of OVB. Our simulation results can be used by researchers seeking to determine whether the Copula method is applicable to their data, and our R code can be adjusted to run simulations tailored to the scenario at hand.

In our applied example, we have shown the considerations needed for application of the Copula method in psychological research. As emphasized in that section, we advise researchers to follow these steps:

1. Check the data for applicability of the copula method. Researchers can compare histograms of their predictor(s) with the distribution plots in Figure 2, to help determine the applicability of our simulation findings to their situations, or they can calculate the level of skewness in their data and compare to our values. The most important performance measure to check is the RMSE. The researcher can either use Figure 4 to get a rough sense for the appropriateness of the method to their data, or – if more detailed results are needed – the R code provided in the online supplement can be used to run focused simulations using parameter values (skewness and N)

- calculated from their own data together with a range of values of the unknown parameters β_1 and ρ . We suggest as rough rules of thumb that the expected RMSE should be lower than the unadjusted model's RMSE (the upper horizontal reference line in Figure 4), and preferably close to the covariate adjusted model's RMSE (the lower horizontal reference line in Figure 4). The latter can be a rather tough criterion to meet, though, requiring very large samples and highly skewed predictors.
2. If the data are deemed to be sufficient in terms of sample size and skewness of X , the next step is to calculate the copula correction term. The R code for this step is provided in the online supplement. The code can probably be easily converted to any other software (or exported from R to the software of choice). After this is done, the models can be estimated. As this is done using ordinary linear regression, it can be done using any statistical software. We recommend that researchers estimate both Equations 1 and 2 (i.e., both with and without copula adjustment), to get an impression of the difference in the estimate of β_1 between these. The smaller the difference, the less likely it is that X is endogenous.
 3. The next step is to formally test for endogeneity, that is, test the statistical significance of β_2 . If this test is statistically significant, it is clear that the estimate of β_1 from Equation 2 (i.e., the copula-adjusted equation), should be preferred. Because false positives for β_2 are unlikely, ignoring a statistically significant test of β_2 can probably never be defended. However, if β_2 is not statistically significant, we need to check the statistical power of this test. If statistical power is low, but Step 1 still indicates that the RMSE is likely to be improved with copula adjustment, there may be reason to prefer the copula-adjusted estimate of β_1 (i.e., the estimate from

Equation 2) despite the non-significance of β_2 . This leads to a problem, though, because if statistical power is low for β_2 , it is likely to also be low for β_1 if estimated from Equation 2. Thus, the most difficult situation is the one in which β_1 is statistically significant in Equation 1, both β_1 and β_2 are statistically non-significant in Equation 2, and statistical power for β_2 and copula adjusted β_1 are low. In this situation, the researcher may need to state that the test is inconclusive.

Limitations

The copula method is not without its limitations, and these are important to keep in mind. First, we want to highlight the large RMSE that our simulations showed, largely due to the collinearity between X and the copula correction term, which increases as the distribution of X gets closer to a normal distribution. Although larger skewness in X leads to better efficiency of the copula method, its standard errors are larger than if a known confounder is included as covariate. This trade-off in efficiency may be acceptable given the stronger assumptions we need to make under covariate adjustment. However, if the researcher is certain about the causal structure of the variables involved, it would be better to measure and include known confounders as covariates in the statistical model, rather than relying on the copula method as a universal cure for endogeneity.

A second limitation of the copula method is the rather low statistical power of the test for endogeneity. As Figure 5 shows, samples of several thousand observations are needed to get sufficient statistical power to detect smaller endogeneity, even if skewness is large. This is important, because if power of the endogeneity test is ignored, a researcher will easily make a Type-II error, concluding that there is no endogeneity when in fact there is. Such an error may have fatal consequences if the researcher concludes that a particular intervention has a causal

effect, when in fact it does not (or may even be harmful). In such cases, it may be wise to consider choosing the copula-adjusted estimates rather than the unadjusted ones, even if the copula coefficient is statistically non-significant (assuming a large enough sample size and skewness of X). The adjusted coefficients are more likely to be unbiased than the unadjusted ones. At the same time, however, one may need to consider the reduced statistical power of X in the copula-adjusted models. Thus, this is a situation in which the researcher may need to balance the risks for making a Type-II vs. Type-I error. Depending on the research question, Type-I/II errors may carry differential real-world risks, so that the researcher may choose to risk one more than the other.

A limitation of our study is that we did not consider alternative (non-normal) distributions of X , so it is possible that our results are specific to skewed transformations of the normal distribution. However, although the specific estimates we obtained might change with other distributions, the same principles should hold: greater deviations from normality should lead to better performance of the copula method, and sample size may to some extent compensate for smaller degrees of non-normality. Park and Gupta (2012) used a uniform distribution for X , which yielded results similar to those under a skewness level around 1.0 in the present study. In a previous version of this paper, we used a skew-normal distribution, which yielded highly similar values as the transformed ones we used in the final version. Further, we did not consider interactions between deviations from the non-normality assumption for X and other assumptions of the copula method. As mentioned in the introduction, these include normally distributed errors for Y and bivariate normality of X and ε_Y . These assumptions have been tested in prior studies (Park & Gupta, 2012; Tran & Tsionas, 2019), and the method has been found to be robust to violations. However, there will obviously be limits to how much these

assumptions can be violated, and we do not know whether violating more than one of the assumptions simultaneously lead to additional problems. This should be the topic of further study. An interesting extension of the copula method would be to explore non-parametric copula functions (Mukhopadhyay & Parzen, 2020; Wu et al., 2015). With a non-parametric copula function, we would not need to make any assumption about the form of the $X \leftrightarrow \varepsilon_Y$ dependency structure, making the method even more robust for causal inference.

Finally, we have only considered the copula method in the context of a very simple linear regression model. Park and Gupta (2012) also tested the method in random slope and logistic models, but further studies of the copula correction in conjunction with advanced statistical modeling techniques (e.g., multilevel and structural equation models with latent variables) should be conducted, in order expand the applicability of this useful method.

Finally, we focused on the canonical case of continuous X variables. In psychological research, however, many independent (and dependent) variables are discrete (e.g., binary, ordered-categorical, nominal, and count variable). Under the copula method, continuous X variables are ideal, as they provide the largest amount of identifying statistical information for the copula method. Nonetheless, Park and Gupta (2012) showed that the copula method can work reasonable with discrete independent variables, as long as they are not binary, in which case there is too little information in the distribution for reliable use of the technique. In these cases, covariate adjustment and/or instrumental variables may be preferable for addressing OVB. Future research should focus on the case of discrete X to more fully establish the limits of the copula correction in this regard.

Conclusions

Despite the above shortcomings, the present study shows that the copula method is a promising treatment for the ubiquitous problem of OVB in psychological research, when the data at hand enable its application. The statement ‘correlation does not imply causation’ is an oft-repeated mantra in introductory courses in statistics. However, under some circumstances, correlation may imply causation, and the copula method expands the toolbox for assessing whether it does. Moreover, causation may sometimes be hidden behind an apparent lack of correlation, if the relationship is attenuated by an unobserved confounder (rather than inflated by one), and copula correction has the potential to reveal such instances. Although fairly large samples may be needed for the application of the copula method, large samples of observational data may often be easier to collect than when experimental manipulation is required. In particular, panel data, in which the total sample size is the combination of individuals and repeated measures, may be ideal – especially with modern data collection techniques such as Ecological Momentary Assessment. Future research along the lines suggested above has the potential to further clarify the properties of the copula method and expand its utility for psychology researchers.

References

- Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in Instrumental Variables Regression: Theory and practice. *Annual Review of Economics*, 11(1), 727-753.
<https://doi.org/10.1146/annurev-economics-080218-025643>
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, 21, 1086-1120.
<https://doi.org/10.1016/j.leaqua.2010.10.010>
- Asselmann, E., Knappe, S., Wittchen, H.-U., Lieb, R., & Beesdo-Baum, K. (2015). Stability of recalled parental rearing behavior in a community sample of adolescents and young adults. *Journal of Research on Adolescence*, 25(4), 614-621.
<https://doi.org/10.1111/jora.12158>
- Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal*, 19(4), 757-767. <https://doi.org/10.1177/1536867X19893614>
- Beck, A. T., Rush, A., Shaw, B., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.
- Becker, T. E., Atinc, G., Breugh, J. A., Carlson, K. D., Edwards, J. R., & Spector, P. E. (2016). Statistical control in correlational studies: 10 essential recommendations for organizational researchers. *Journal of Organizational Behavior*, 37(2), 157-167.
<https://doi.org/10.1002/job.2053>
- Bernerth, J. B., Cole, M. S., Taylor, E. C., & Walker, H. J. (2017). Control variables in leadership research: A qualitative and quantitative review. *Journal of Management*, 44(1), 131-160. <https://doi.org/10.1177/0149206317690586>

- Bollmann, G., Rouzinov, S., Berchtold, A., & Rossier, J. (2019). Illustrating Instrumental Variable regressions using the career adaptability – job satisfaction relationship. *Frontiers in Psychology, 10*(1481). <https://doi.org/10.3389/fpsyg.2019.01481>
- Boomsma, A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling, 20*(3), 518-540. <https://doi.org/10.1080/10705511.2013.797839>
- Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development*. Basic Books.
- Brähler, E. (2014). *Normative data from a standardization sample from 1994 for Giessen-Test, Giessen Subjective Complaints List (Short Form), Questionnaire for the Assessment of Recalled Parental Rearing Behaviour, Life Satisfaction Questionnaire and Coping Assessment. [Translated Title] (Version 1.0.0)*. <https://doi.org/https://doi.org/10.5160/psychdata.brer94re08>
- Brähler, E., Schumacher, J., & Brähler, C. (1999). Erste gesamtdeutsche Normierung und spezifische Validitätsaspekte des Gießen-Tests. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*, 231-243.
- Chalak, K., & White, H. (2011). Viewpoint: An extended class of instrumental variables for the estimation of causal effects. *Canadian Journal of Economics/Revue canadienne d'économique, 44*(1), 1-51. <https://doi.org/10.1111/j.1540-5982.2010.01622.x>
- Chen, B., Pearl, J., & Kline, R. (2018). Graphical tools for linear path models. *Psychometrika, 4*(5).
- Cheng, D., Li, J., Liu, L., Yu, K., Lee, T. D., & Liu, J. (2020). *Towards precise causal effect estimation from data with hidden variables* [Working Paper. Repository arXiv]. <https://arxiv.org/pdf/2002.10091.pdf>

Christopoulos, D., McAdam, P., & Tzavalis, E. (2019). Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics*, 1-13.

<https://doi.org/10.1080/07350015.2019.1647213>

Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society B (Statistical Methodology)*, 82(1), 39-67.

<https://doi.org/10.1111/rssb.12348> (Journal of the Royal Statistical Society Series B)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

Crits-Christoph, P., Gallop, R., Gaines, A., Rieger, A., & Connolly Gibbons, M. B. (2020).

Instrumental variable analyses for causal inference: Application to multilevel analyses of the alliance–outcome relation. *Psychotherapy Research*, 30(1), 53-67.

<https://doi.org/10.1080/10503307.2018.1544724>

Crown, W. H., Henk, H. J., & Vanness, D. J. (2011). Some cautions on the use of Instrumental Variables estimators in outcomes research: How bias in Instrumental Variables estimators Is affected by instrument strength, instrument contamination, and sample size. *Value in Health*, 14(8), 1078-1084. <https://doi.org/10.1016/j.jval.2011.06.009>

Cummiskey, K., Adams, B., Pleuss, J., Turner, D., Clark, N., & Watts, K. (2020). Causal inference in introductory statistics courses. *Journal of Statistics Education*, 28(1), 2-8.

<https://doi.org/10.1080/10691898.2020.1713936>

Datta, H., Ailawadi, K. L., & van Heerde, H. J. (2017). How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response? *Journal of Marketing*, 81(3), 1-20. <https://doi.org/10.1509/jm.15.0340>

- Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.
<https://doi.org/https://doi.org/10.1016/j.socscimed.2017.12.005>
- DeMaris, A. (2014). Combating unmeasured confounding in cross-sectional studies: evaluating instrumental-variable and Heckman selection models. *Psychological Methods*, 19(3), 380-397. <https://doi.org/10.1037/a0037416>
- Deuchert, E., & Huber, M. (2017). A cautionary tale about control variables in IV estimation. *Oxford Bulletin of Economics and Statistics*, 79(3), 411-425.
<https://doi.org/10.1111/obes.12177>
- Dzhumashev, R., & Tursunaliyeva, A. (2019). *Synthetic Instrumental Variables* [Working Paper].
<https://ssrn.com/abstract=3370143>
- Ebbes, P., Wedel, M., Böckenholt, U., & Steerneman, T. (2005). Solving and testing for regressor-error (in)dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quantitative Marketing and Economics*, 3(4), 365-392. <https://doi.org/10.1007/s11129-005-1177-6>
- Fahrenberg, J., Myrtek, M., Schumacher, J., & Brähler, E. (2000). *FLZ - Fragebogen zur Lebenszufriedenheit*. Hogrefe.
- Franks, A., D'Amour, A., & Feller, A. (2019). Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 1-33. <https://doi.org/10.1080/01621459.2019.1604369>
- Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques

- coupled with experimental data. *Developmental Psychology*, 44(2), 381-394.
<https://doi.org/10.1037/0012-1649.44.2.381>
- Gould, W. (2011, 01/27). Use poisson rather than regress; tell a friend.
<https://blog.stata.com/2011/08/22/use-poisson-rather-than-regress-tell-a-friend/>
- Greenberg, J. R., & Mitchell, S. A. (1983). *Object relations in psychoanalytic theory*. Press, Harvard University.
- Gultchin, L., Kusner, M. J., Kanade, V., & Silva, R. (2020). *Differentiable causal backdoor discovery* Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), Palermo, Italy. <https://arxiv.org/pdf/2003.01461.pdf>
- Gustafson, P., & McCandless, L. C. (2018). When Is a sensitivity parameter exactly that? *Statistical Science*, 33(1), 86-95. <https://doi.org/10.1214/17-STS632>
- Han, S. (2020). Nonparametric estimation of triangular simultaneous equations models under weak identification. *Quantitative Economics*, 11(1), 161-202.
<https://doi.org/10.3982/QE975>
- Harding, M. (2016). Finite sample bias corrected IV estimation for weak and many instruments. In J. Hausman (Ed.), *Essays in Honor of Aman Ullah* (Vol. 36, pp. 245-273). Emerald Group Publishing Limited. <https://doi.org/10.1108/S0731-905320160000036016>
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.
<https://doi.org/10.2307/1913827>
- Heller, D., Judge, T. A., & Watson, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal of Organizational Behavior*, 23(7), 815-835. <https://doi.org/10.1002/job.168>
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.

Holtforth, M. G. (2001). Multidimensional Scalogram Analysis of the Giessen-Test (GT).

Psychological Reports(1), 397-422.

Hult, G. T. M., Hair, J. F., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. M. (2018).

Addressing endogeneity in international marketing applications of Partial Least Squares

Structural Equation Modeling. *Journal of International Marketing*, 26(3), 1-21.

<https://doi.org/10.1509/jim.17.0151>

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences*. Cambridge University Press.

Kim, N., & Pohlmeier, W. (2016). A note on the regularized approach to biased 2SLS estimation with weak instruments. *Oxford Bulletin of Economics and Statistics*, 78(6), 915-924.

<https://doi.org/10.1111/obes.12144>

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica*, 65(5), 1201-1213. <https://doi.org/10.2307/2171884>

Lockwood, J. R., & McCaffrey, D. F. (2019). Impact evaluation using Analysis of Covariance with error-prone covariates that violate surrogacy. *Evaluation Review*, 43(6), 335-369.

<https://doi.org/10.1177/0193841X19877969>

MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: the link is the mediator by treatment interaction.

Prevention Science, 21(2), 147-157. <https://doi.org/10.1007/s11121-019-01076-4>

Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development*, 50, 66-104.

- Maydeu-Olivares, A., Shi, D., & Fairchild, A. J. (2020). Estimating causal effects in linear regression models with observational data: The instrumental variables regression model. *Psychological Methods*, 25(2), 243-258. <https://doi.org/10.1037/met0000226>
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156-166. <https://doi.org/10.1037/0033-2909.105.1.156>
- Mukhopadhyay, S., & Parzen, E. (2020). Nonparametric universal copula modeling [<https://doi.org/10.1002/asmb.2503>]. *Applied Stochastic Models in Business and Industry*, 36(1), 77-94. <https://doi.org/https://doi.org/10.1002/asmb.2503>
- Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *The Journal of Economic Perspectives*, 20(4), 111-132.
- Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.
- Oyenubi, A. (2020). A note on covariate balancing propensity score and instrument-like variables. *Economics Bulletin*, 40(1), 202-209.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619-632. [https://doi.org/10.1002/\(SICI\)1099-1255\(199611\)11:6<619::AID-JAE418>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1)
- Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using Copulas. *Marketing Science*, 31(4), 567-586. <https://doi.org/https://doi.org/10.1287/mksc.1120.0718>

- Pearl, J. (2011). Invited Commentary: Understanding Bias Amplification. *American Journal of Epidemiology*, 174(11), 1223-1227. <https://doi.org/10.1093/aje/kwr352>
- Pearl, J. (2013). Linear Models: A useful “microscope” for causal analysis. *Journal of Causal Inference*, 1(1), 155-170.
- Pearl, J., & MacKenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.
- Perris, C., Jacobsson, L., Lindström, H., von Knorring, L., & Perris, H. (1980). Development of a new inventory assessing memories of parental rearing behaviour. *Acta Psychiatrica Scandinavica*, 61(4), 265-274. <https://doi.org/10.1111/j.1600-0447.1980.tb00581.x>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of applied psychology*, 88, 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- R Core Team. (2019). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing.
- Richardson, T. G., Davey Smith, G., & Munafò, M. R. (2019). Conditioning on a collider may induce spurious associations: do the results of Gale et al. (2017) support a health-protective effect of neuroticism in population subgroups? *Psychological Science*, 30(4), 629-632. <https://doi.org/10.1177/0956797618774532>
- Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics*, 85(4), 777-792.
- Rosenbaum, P. R. (2017). *Observation and experiment: an introduction to causal inference*. Harvard University Press.

- Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. *The Leadership Quarterly*, 101348. <https://doi.org/https://doi.org/10.1016/j.leaqua.2019.101348>
- Sengewald, M.-A., Steiner, P. M., & Pohl, S. (2019). When does measurement error in covariates impact causal effect estimates? Analytic derivations of different scenarios and an empirical illustration. *British Journal of Mathematical and Statistical Psychology*, 72(2), 244-270. <https://doi.org/10.1111/bmsp.12146>
- Sklar, M. (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8. <https://books.google.se/books?id=nreSmAEACAAJ>
- StataCorp. (2019). *Stata Statistical Software: Release 16*. In StataCorp LLC.
- Stock, J. H., & Trebbi, F. (2003). Retrospectives: Who invented Instrumental Variable Regression? *Journal of Economic Perspectives*, 17(3), 177-194. <https://doi.org/10.1257/089533003769204416>
- Stokes, T., Steele, R., & Shrier, I. (2020). *Causal simulation experiments: Lessons from bias amplification*. [Working Paper. Repository arXiv]. <https://arxiv.org/pdf/2003.08449.pdf>
- Tao, Y., & Fu, H. (2019). Doubly robust estimation of the weighted average treatment effect for a target population. *Statistics in Medicine*, 38(3), 315-325. <https://doi.org/10.1002/sim.7980>
- Thoemmes, F., & Ong, A. D. (2015). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4(1), 40-59. <https://doi.org/10.1177/2167696815621645>

- Tran, K. C., & Tsionas, E. G. (2015). Endogeneity in stochastic frontier models: Copula approach without external instruments. *Economics Letters*, 133, 85-88.
<https://doi.org/https://doi.org/10.1016/j.econlet.2015.05.026>
- Tran, K. C., & Tsionas, M. G. (2019). *To instrument or not to instrument: Efficient semiparametric copula estimation of regression models with endogeneity* [Preprint].
https://www.researchgate.net/publication/333560941_To_Instrument_or_Not_to_Instrument_Efficient_Semiparametric_Copula_Estimation_of_Regression_Models_with_Endogeneity
- Tübbicke, S. (2020). *Entropy balancing for continuous treatments* [Working paper. Repository arXiv]. <https://arxiv.org/pdf/2001.06281.pdf>
- Vaisey, S., & Miles, A. (2014). What you can—and can't—do with three-wave panel data. *Sociological Methods & Research*, 46(1), 44-67.
<https://doi.org/10.1177/0049124114547769>
- van der Zander, B., Liśkiewicz, M., & Textor, J. (2019). Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence*, 270, 1-40. <https://doi.org/https://doi.org/10.1016/j.artint.2018.12.006>
- van Kippersluis, H., & Rietveld, C. A. (2018). Beyond plausibly exogenous. *The Econometrics Journal*, 21(3), 316-331. <https://doi.org/10.1111/ectj.12113>
- VanderWeele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction*. Oxford University Press.
- VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, 34(3), 211-219. <https://doi.org/10.1007/s10654-019-00494-6>

- Wang, X., Jiang, Y., Zhang, N. R., & Small, D. S. (2018). Sensitivity analysis and power for instrumental variable studies. *Biometrics*, 74(4), 1150-1160.
<https://doi.org/10.1111/biom.12873>
- White, H., & Chalak, K. (2010). Testing a conditional form of exogeneity. *Economics Letters*, 109(2), 88-90. <https://doi.org/10.1016/j.econlet.2010.08.018>
- Witte, J., Henckel, L., Maathuis, M. H., & Didelez, V. (2020). On efficient adjustment in causal graphs. *Journal of Machine Learning Research*, 21, 1-45.
<https://arxiv.org/pdf/2002.06825.pdf>
- Wooldridge, J. M. (2010). *Econometric analysis of cross-sectional and panel data* (2nd ed.). The MIT Press.
- Wu, J., Wang, X., & Walker, S. G. (2015). Bayesian nonparametric estimation of a copula. *Journal of Statistical Computation and Simulation*, 85(1), 103-116.
<https://doi.org/10.1080/00949655.2013.806508>
- Wu, X., Mealli, F., Kioumourtzoglou, M.-A., Dominici, F., & Braun, D. (2020). *Matching on generalized propensity scores with continuous exposures*.
<https://arxiv.org/pdf/1812.06575.pdf>
- Young, A. (2020). *Consistency without Inference: Instrumental Variables in Practical Application (Working Paper)*. <https://uh-ir.tdl.org/handle/10657/4338>
- Young, J. E. (1999). *Cognitive therapy for personality disorders: A schema-focused approach, 3rd ed.* Professional Resource Press/Professional Resource Exchange.
- Yu, Y., & Jaenicke, E. C. (2020). Estimating food waste as household production inefficiency. *American Journal of Agricultural Economics*, 102(2), 525-547.
<https://doi.org/10.1002/ajae.12036>

- Zander, B. V. D., Textor, J., & Liskiewicz, M. (2015). *Efficiently finding conditional instruments for causal inference* Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina.
- Zhao, Q. (2019). Covariate balancing propensity score by tailored loss functions. *Annals of Statistics*, 47(2), 965-993. <https://doi.org/10.1214/18-AOS1698>
- Zhao, Q., & Percival, D. (2016). Entropy balancing is doubly robust. *Journal of Causal Inference*, 5(1), 20160010. <https://doi.org/https://doi.org/10.1515/jci-2016-0010>

Appendix A. Mathematical explanation of the Copula Method

Omitted Variable Bias

True Model: $Y_i = \beta_0 + \beta_1 X_i + \gamma OV_i + \varepsilon_i, \text{Corr}(OV_i, X_i) \neq 0$

Omitted Variable: OV_i

Estimation Model: $Y_i = \beta_0 + \beta_1 X_i + \xi_i$, where $\xi_i = \gamma OV_i + \varepsilon_i$

Now, X_i and ξ_i are correlated and we encounter an endogeneity problem (i.e., endogenous regressor or regressor-error dependence problem).

Copula approach

The copula approach tackles the endogeneity problem due to regressor-error correlations by assuming a flexible correlation structure between the two based on a Gaussian copula. OLS is biased because X_i and ξ_i are correlated. The copula approach adds the following two assumptions to specify the dependence structure between X_i and ξ_i :

$$1) \xi_i \sim N(0, \sigma_\xi^2)$$

$$2) \begin{bmatrix} \Phi^{-1}(F_\xi(\xi_i)) \\ \Phi^{-1}(F_X(X_i)) \end{bmatrix} \sim BVN \begin{pmatrix} 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$$

where $F_\xi(\cdot)$ is the CDF of ξ_i , $F_X(\cdot)$ is the CDF of X_i , $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal distribution and ρ is the endogeneity parameter. Note that a specific relationship assumed between X_i and ξ_i in 2) is referred to as a Gaussian copula. Previous studies have found that the Gaussian copula is parsimonious yet flexible in capturing a wide range of dependence structures and is proven to be robust in many empirical applications.

From 2), we can re-write $\Phi^{-1}(F_\xi(\xi_i)) = \rho \Phi^{-1}(F_X(X_i)) + \sqrt{1 - \rho^2} \varsigma_i$ where ς_i is exogenous and follows $N(0, 1)$. Moreover, since $F_\xi(\cdot)$ is the CDF of the normal distribution with

mean of zero and variance of σ_ξ^2 , $\Phi^{-1}\left(F_\xi(\xi_i)\right) = \frac{\xi_i}{\sigma_\xi}$. Using these, we obtain $\xi_i =$

$\sigma_\xi \rho \Phi^{-1}\left(F_X(X_i)\right) + \sigma_\xi \sqrt{1 - \rho^2} \varsigma_i$. This result implies that ξ_i can be decomposed into two

parts: shocks that are correlated with X_i and thus result in the endogeneity problem

$(\sigma_\xi \rho \Phi^{-1}\left(F_X(X_i)\right))$ and the remaining exogenous shocks $(\sigma_\xi \sqrt{1 - \rho^2} \varsigma_i)$. Now, we can rewrite the model as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \sigma_\xi \rho \Phi^{-1}\left(F_X(X_i)\right) + \sigma_\xi \sqrt{1 - \rho^2} \varsigma_i = \beta_1 X_i + \sigma_\xi \rho \Phi^{-1}\left(F_X(X_i)\right) + v_i,$$

where $v_i = \sigma_\xi \sqrt{1 - \rho^2} \varsigma_i$. Since we observe X_i , we can consistently estimate

$\Phi^{-1}\left(F_X(X_i)\right)$ and include it as an additional regressor in empirical analyses. Moreover, v_i is unrelated with any other terms in the right-hand side of the regression equation and thus the model does not suffer from the endogeneity problem anymore.

When we implement the method, we need to estimate $F_X(X_i)$ from the data and $\Phi^{-1}\left(\widehat{F_X(X_i)}\right)$ is used instead of $\Phi^{-1}\left(F_X(X_i)\right)$. The simplest way to obtain $\widehat{F_X(X_i)}$ is to use the empirical distribution function. In the literature, this type of regressor is referred to as a “generated regressor”.

For the identification of the model, we add one condition regarding the distribution of X_i :

3) X_i is not normally distributed. To understand the identification condition 3), let us assume that X_i follows the normal distribution. Then, $\Phi^{-1}\left(F_X(X_i)\right)$ is a linear transformation of X_i and both follow normal distributions. Consequently, we cannot separately identify α and $\sigma_\xi \rho$.

Table A1. Omitted Variable Bias and Copula Approach

<i>Model Setting</i>	<i>Assumption among variables</i>
Dependent Variable Y_i	
Exogeneous regressor X_i	
Omitted Variable OV_i	$Corr(OV_i, X_i) \neq 0$
True Model $Y_i = \beta_1 X_i + \gamma OV_i + \varepsilon_i$	
Estimation Model $Y_i = \beta_1 X_i + \xi_i, \quad \xi_i = \gamma OV_i + \varepsilon_i$	
$Corr(\xi_i, X_i) \neq 0$	
<i>Copula Approach Assumptions</i>	<i>Notes</i>
$\xi_i \sim N(0, \sigma_\xi^2)$	Common
distributional assumption for random shocks	
$\begin{bmatrix} \Phi^{-1}(F_\xi(\xi_i)) \\ \Phi^{-1}(F_X(X_i)) \end{bmatrix} \sim BVN \begin{pmatrix} 0 & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$	
Gaussian copula, known to be flexible and robust	
$F_\xi(\cdot)$: CDF of ξ_i , $F_X(\cdot)$: CDF of X_i ,	
$\Phi^{-1}(\cdot)$: inverse CDF of $N(0, 1)$	
X_i is not normally distributed	Required for the
model identification	
<i>Estimation Model after applying the Copula Approach</i>	<i>Note</i>
$Y_i = \beta_1 X_i + \sigma_\xi \cdot \rho \cdot \Phi^{-1}(F_X(X_i)) + v_i$,	X_i and $\Phi^{-1}(F_X(X_i))^a$
are included as regressors	
$v_i = \sigma_\xi \cdot \sqrt{1 - \rho^2} \cdot \varsigma_i$	v_i is uncorrelated with
X_i and $\Phi^{-1}(F_X(X_i))$	

Note. ^a $\Phi^{-1}(F_X(X_i))$ is denoted as ‘C’ in the manuscript equations.

Table 1.

Regression results with recalled parental rejection/punishment as predictor.

<i>N</i> = 2998	Equation 1				Equation 2							
Skewness = 1.45	β_1	se	<i>p</i>	95% CI	β_1	se	<i>p</i>	95% CI	β_2	se	<i>p</i>	95% CI
Functioning	-0.26	0.02	<.001	-0.30, -0.23	-0.23	0.04	<.001	-0.32, -0.15	-0.03	0.04	.47	-0.12, 0.05
Life Satisfaction	-0.30	0.02	<.001	-0.33, -0.26	-0.30	0.02	<.001	-0.34, -0.26	0.01	0.02	.78	-0.04, 0.05
<i>N</i> = 1000, skewness = 1.44												
Functioning	-0.32	0.03	<.001	-0.38, -0.26	-0.32	0.07	<.001	-0.46, -0.18	0.00	0.08	.99	-0.15, 0.15
Life Satisfaction	-0.34	0.03	<.001	-0.40, -0.28	-0.24	0.07	<.001	-0.38, -0.10	-0.12	0.07	.12	-0.26, 0.03
<i>N</i> = 500, skewness = 1.60												
Functioning	-0.31	0.04	<.001	-0.39, -0.22	-0.37	0.10	<.001	-0.56, -0.18	0.07	0.10	.51	-0.13, 0.27
Life Satisfaction	-0.41	0.04	<.001	-0.50, -0.32	-0.46	0.10	<.001	-0.65, -0.26	0.05	0.11	.61	-0.15, 0.26
<i>N</i> = 200, skewness = 1.47												
Functioning	-0.18	0.07	.01	-0.32, -0.04	-0.33	0.17	.05	-0.67, 0.00	0.18	0.18	.32	-0.17, 0.53
Life Satisfaction	-0.24	0.07	<.001	-0.38, -0.10	-0.03	0.17	.85	-0.37, 0.31	-0.24	0.18	.18	-0.59, 0.11
<i>N</i> = 100, skewness = 1.34												
Functioning	-0.21	0.10	.04	-0.41, -0.01	-0.40	0.24	.10	-0.88, 0.07	0.23	0.25	.38	-0.28, 0.73
Life Satisfaction	-0.11	0.09	.24	-0.29, 0.07	-0.00	0.22	.99	-0.44, 0.43	-0.12	0.23	.61	-0.58, 0.34

Table 2.

Regression results with recalled parental control/overprotection as predictor.

<i>N</i> = 2998	Equation 1				Equation 2							
Skewness = 0.47	β_1	se	<i>p</i>	95% CI	β_1	se	<i>p</i>	95% CI	β_2	se	<i>p</i>	95% CI
Functioning	-0.13	0.02	<.001	-0.16, -0.09	-0.39	0.10	<.001	-0.59, -0.20	0.28	0.10	.01	0.08, 0.49
Life Satisfaction	-0.16	0.02	<.001	-0.19, -0.12	-0.37	0.10	<.001	-0.56, -0.17	0.22	0.10	.03	0.02, 0.42
<i>N</i> = 1000, skewness = 0.42												
Functioning	-0.16	0.03	<.001	-0.22, -0.09	-0.72	0.18	<.001	-1.07, -0.36	0.60	0.19	<.001	0.23, 0.97
Life Satisfaction	-0.17	0.03	<.001	-0.23, -0.10	-0.63	0.18	<.001	-0.98, -0.28	0.50	0.19	.01	0.13, 0.86
<i>N</i> = 500, skewness = 0.60												
Functioning	-0.14	0.05	<.001	-0.22, -0.05	-0.51	0.21	.02	-0.92, -0.10	0.40	0.22	.07	-0.03, 0.84
Life Satisfaction	-0.18	0.05	<.001	-0.27, -0.09	-0.75	0.22	<.001	-1.18, -0.31	0.61	0.23	.01	0.15, 1.07
<i>N</i> = 200, skewness = 0.45												
Functioning	-0.07	0.07	.34	-0.21, 0.07	-0.16	0.40	.70	-0.94, 0.63	0.09	0.41	.82	-0.72, 0.91
Life Satisfaction	-0.07	0.07	.37	-0.21, 0.08	-0.28	0.40	.49	-1.08, 0.52	0.22	0.42	.59	-0.60, 1.05
<i>N</i> = 100, skewness = 0.09												
Functioning	-0.03	0.10	.75	-0.24, 0.17	-0.26	0.73	.72	-1.70, 1.18	0.24	0.77	.75	-1.28, 1.76
Life Satisfaction	-0.14	0.09	.14	-0.32, 0.04	-0.79	0.64	.22	-2.06, 0.48	0.70	0.68	.30	-0.64, 2.04

Figure 1. Path diagrams showing three instances of omitted variable bias, the result that a misspecified model without the omitted variable yields, and the possibility of recovering the true effect by estimating the correlation between X and the residual of Y .

Figure 2. Graphs of skewness of distributions at the levels of skewness in X used in the simulations.

Figure 3. Average estimated coefficients from simulations of Copula correction for different values of skewness in X and sample sizes. In all graphs, .30 is the population value for the causal effect of X on Y . The thick horizontal reference line shows the expected coefficient value for the unadjusted linear model.

Figure 4. Root Mean Squared Error from simulations of copula correction for different values of skewness in X , sample sizes, and endogeneity parameter ρ with effect size for X set at $\beta = .30$. The thick solid line shows RMSE for unadjusted OLS, while the thick long-dashed line shows RMSE for OLS with confounding variable known and included as covariate.

Figure 5. Statistical power of test of the copula coefficient as a function of N , skewness and size of endogeneity parameter ρ . Effect size for β_1 is set at .30 in all three graphs.

Online Supplement for “Using Copulas to Handle Unobserved Confounding in Psychological Research: Tutorial and Simulation Studies”

R code for generating simulated data consistent with Figure 1a.

```

nobs <- 1000 #Insert number of observations here
niter <- 1000 #Insert preferred number of iterations
YonX <- 0 #Coefficient for the causal effect of X -> Y
YonOV <- -0.50 #Coefficient for OV -> Y
XonOV <- 0.50 #Coefficient for OV -> X
out <- matrix(1:niter) #Variable for saving estimation results

for (i in 1:niter) {
  #Generate Omitted Variable (OV)
  OV <- rnorm(nobs, mean = 0, sd = 1)

  #Generate error terms for X and Y so that total variance = 1.0
  ResX <- rnorm(nobs, mean = 0, sd = (sqrt(1-(XonOV^2))))
  ResY <- rnorm(nobs, mean = 0, sd = (sqrt(1-(YonOV^2+YonX^2))))

  #Generate X and Y
  X <- XonOV*OV + ResX
  Y <- YonOV*OV +YonX*X + ResY

  #Estimate linear regression ignoring OV
  sum <- lm(Y~X)
  out[i,1]<-summary(sum)$coefficients[2, 1] #save coefficient for Y on X
}
mean(out) #Show average estimated effect of Y on X (should in this example be close to
0.25)

```

R code for generating simulated data consistent with Figure 1b.

```

nobs <- 1000 #Insert number of observations here
niter <- 1000 #Insert preferred number of iterations
YonX <- 0.25 #Coefficient for causal effect of X -> Y
YonOV <- -0.50 #Coefficient for OV -> Y
XonOV <- 0.50 #Coefficient for OV -> X
out <- matrix(1:niter) #Variable for saving estimation results

for (i in 1:niter) {
#Generate Omitted Variable (OV)
OV <- rnorm(nobs, mean = 0, sd = 1)

#Generate error terms for X and Y so that total variance = 1.0
ResX <- rnorm(nobs, mean = 0, sd = (sqrt(1-(XonOV^2))))
ResY <- rnorm(nobs, mean = 0, sd = (sqrt(1-(YonOV^2+YonX^2))))

#Generate X and Y
X <- XonOV*OV + ResX
Y <- YonOV*OV +YonX*X + ResY

#Estimate linear regression ignoring OV
sum <- lm(Y~X)
out[i,1]<-summary(sum)$coefficients[2, 1] #save coefficient for Y on X
}
mean(out) #Show average estimated effect of Y on X (should in this example be close to 0)

```


R code for generating simulated data consistent with Figure 1c.

```

nobs <- 1000 #Insert number of observations here
niter <- 1000 #Insert preferred number of iterations
YonX <- 0.25 #Coefficient for causal effect of X -> Y
YonOV <- -0.71 #Coefficient for OV -> Y
XonOV <- 0.71 #Coefficient for OV -> X
out <- matrix(1:niter) #Variable for saving estimation results

for (i in 1:niter) {
#Generate Omitted Variable (OV)
OV <- rnorm(nobs, mean = 0, sd = 1)

#Generate error terms for X and Y so that total variance = 1.0
ResX <- rnorm(nobs, mean = 0, sd = (sqrt(1-(XonOV^2))))
ResY <- rnorm(nobs, mean = 0, sd = (sqrt(1-(YonOV^2+YonX^2))))

#Generate X and Y
X <- XonOV*OV + ResX
Y <- YonOV*OV +YonX*X + ResY

#Estimate linear regression ignoring OV
sum <- lm(Y~X)
out[i,1]<-summary(sum)$coefficients[2, 1] #save coefficient for Y on X
}
mean(out) #Show average estimated effect of Y on X (should in this example be close to -
0.25)

```

R code for simulation

#The following packages need to be installed before running the below code: doParallel, #HelpersMG, doRNG, lmttest (use “install.packages(‘package name’))

```
library(doParallel)
library(HelpersMG)
library(doRNG)
library(lmttest)
```

niter <- 2000 #This is the number of simulated datasets per condition

#Below are the sample sizes we used, they can be changed according to the researcher’s #needs

```
nobs <- matrix(data = NA, nrow = 5, ncol = 1)
nobs[1,1] <- 100
nobs[2,1] <- 200
nobs[3,1] <- 500
nobs[4,1] <- 1000
nobs[5,1] <- 3000
```

#Below are the effect sizes for β_1 , these can be changed according to the researcher’s #needs

```
B1 <- matrix(data = NA, nrow = 4, ncol = 1)
B1[1,1] <- 0
B1[2,1] <- .10
B1[3,1] <- .30
B1[4,1] <- .50
```

#Below are the magnitudes for ρ , these can be changed according to the researcher’s #needs

```
Rho <- matrix(data = NA, nrow = 7, ncol = 1)
Rho[1,1] <- -1 * sqrt(.30)
Rho[2,1] <- -1 * sqrt(.20)
Rho[3,1] <- -1 * sqrt(.10)
Rho[4,1] <- 0
Rho[5,1] <- sqrt(.10)
Rho[6,1] <- sqrt(.20)
Rho[7,1] <- sqrt(.30)
```

#Below are the skewness parameters. These change slightly depending on sample size #so if sample sizes are changed researchers need to use our simulator for calculating #skewness parameters to insert the values needed

```
skewpar <- matrix(data = NA, nrow = 9, ncol = 5)
skewpar[1,] <- c(.433, .42, .415, .414, .412) # skewness ~ 0.40
skewpar[2,] <- c(.535, .52, .51, .507, .506) # skewness ~ 0.60
skewpar[3,] <- c(.62, .60, .585, .583, .58) # skewness ~ 0.80
skewpar[4,] <- c(.695, .665, .65, .645, .644) # skewness ~ 1.00
skewpar[5,] <- c(.757, .725, .705, .698, .694) # skewness ~ 1.20
skewpar[6,] <- c(.817, .78, .753, .745, .739) # skewness ~ 1.40
skewpar[7,] <- c(.873, .825, .797, .787, .779) # skewness ~ 1.60
skewpar[8,] <- c(.926, .87, .836, .823, .814) # skewness ~ 1.80
```

```
skewpar[9,] <- c(.978, .915, .873, .858, .847) # skewness ~ 2.00
```

#This is just for getting the correct skewness values into the table:

```
skew <- matrix(data = NA, nrow = 9, ncol = 1)
```

```
skew[1] <- .40 # skewness ~ 0.40
```

```
skew[2] <- .60 # skewness ~ 0.60
```

```
skew[3] <- .80 # skewness ~ 0.80
```

```
skew[4] <- 1.00 # skewness ~ 1.00
```

```
skew[5] <- 1.20 # skewness ~ 1.20
```

```
skew[6] <- 1.40 # skewness ~ 1.40
```

```
skew[7] <- 1.60 # skewness ~ 1.60
```

```
skew[8] <- 1.80 # skewness ~ 1.80
```

```
skew[9] <- 2.00 # skewness ~ 2.00
```

#Below is the simulation function. It only needs to be run once per R session

```
sim <- function(i) {
```

#Generate data

#First the omitted variable, here called z (which has a normal distribution)

```
z <- rnorm(nobs[N,1], mean = 0, sd = 1)
```

#Next, predictor variable x is generated, first with a normal distribution

```
x <- abs(Rho[t,1])*z + rnorm(nobs[N,1], mean = 0, sd = sqrt(1-OV[t,1]^2))
```

#x is then transformed to be skewed using an exponential transformation

```
x <- sign(skewpar[q,N])*x + exp(skewpar[q,N]*x)
```

#The transformation changes the scale of x, so it needs to be scaled back (M = 0, SD = 1)

```
x <- scale(x, center = TRUE, scale = TRUE)
```

#Generate residual variance variable for y

```
u<-rnorm(nobs[N,1], mean = 0, sd = 1)
```

#Generate final equation for y

```
y<- YonX[p,1]*x + OV[t,1]*z + sqrt(resY)*u
```

#Generate the Copula Coefficient

#This code can be used on its own to generate C for applied data analyses.

#variable "x" is the endogenous predictor

#Generate ID variable

```
id<-seq(nobs[N,1])
```

#Put into data frame together with x

```
temp1 <- data.frame(id,x)
```

#Create a temporary variable with the data sorted on x

```
temp2 <- temp1[order(x),]
```

```
id<-temp2[,1]
```

#Calculate the values of the empirical distribution function

```
xx_p<-seq((1/nobs[N,1])-0.5*(1/nobs[N,1]),1-0.5*(1/nobs[N,1]),by=(1/nobs[N,1]))
```

#Sort back

```
temp3<-data.frame(id,xx_p)
```

```
temp4<-temp3[order(id),]
```

#Obtain copula correction term using the inverse of the standard normal cumulative

#distribution function

```
temp5<-temp4[,2]
```

```
C <- qnorm(temp5,0,1)
```

#Save parameter estimates

#Save OLS coefficients

OLS <- lm(y~x)

CfOLS <- as.numeric(coef(OLS)[2])

#Extract copula adjustment model coefficients

Corr <- lm(y~x+C)

CfX <- summary(Corr)\$coefficients[2, 1] **#Save adjusted coefficient**

CfC <- summary(Corr)\$coefficients[3, 1] **#Save Copula coefficient**

#Standard errors and confidence intervals

SeX <- summary(Corr)\$coefficients[2, 2] **#Standard error adjusted coefficient**

SeC <- summary(Corr)\$coefficients[3, 2] **#Standard error Copula coefficient**

CLlo_X <- coefci(Corr)[2,1]

CLhi_X <- coefci(Corr)[2,2]

CLlo_C <- coefci(Corr)[3,1]

CLhi_C <- coefci(Corr)[3,2]

#Calculate coverage for X

if ((CLlo_X < YonX[p,1]) & (CLhi_X > YonX[p,1])) {

CovX <- 1

} else {

CovX <- 0

}

#Calculate coverage for C

if ((CLlo_C < (sign(OV[t,1]) * OV[t,1]^2) & (CLhi_C > (sign(OV[t,1]) * OV[t,1]^2)))) {

CovC <- 1

} else {

CovC <- 0

}

Calculate statistical power for X

if (YonX[p,1] != 0) {

if (summary(Corr)\$coefficients[2, 4] < 0.05 & sign(CfX)==sign(YonX[p,1])) {

PowX <- 1

} else {

PowX <- 0

}}

else {

if (summary(Corr)\$coefficients[2, 4] < 0.05) {

PowX <- 1

} else {

PowX <- 0

}}

#Calculate statistical power for Copula coefficient

if (OV[t,1] != 0) {

if (summary(Corr)\$coefficients[3, 4] < 0.05 & sign(CfC)==sign(OV[t,1])) {

```

PowC <- 1
} else {
PowC <- 0
}}
else {
if (summary(Corr)$coefficients[3, 4] < 0.05) {
PowC <- 1
} else {
PowC <- 0
}}

```

#Estimate true model

```

true <- lm(y~x+z)
Cftrue <- as.numeric(coef(true)[2])

```

#Calculate RMSE

```

RM <- CfX - YonX[p,1] #bias for X in Copula corrected model
RMOLS <- CfOLS - YonX[p,1] #bias for X in uncorrected OLS
RM <- RM^2
RMOLS <- RMOLS^2

```

```

RMtrue <- Cftrue - YonX[p,1]
RMtrue <- RMtrue^2

```

```

if (RM < RMOLS) {
CO <- 1
} else {
CO <- 0
}

```

#Return estimated parameters

```

return(c(CfX, CfC, RM, CfOLS, RMOLS, CO, Cftrue, RMtrue, SeX, SeC, CovX, CovC,
PowX, PowC))
} #End of simulation function

```

#Perform simulations, assign number of cores for parallel computing

```

registerDoParallel(cores = 2) #Here you should set the number of cores that your
#computer has. This can be checked with the detectCores() function in package parallel
registerDoRNG(seed = 0)

```

#The number of rows in the table is equal to the number of simulation conditions, e.g.,

```

#N × q × p × t
table <- matrix(data = NA, nrow = 1260, ncol = 21)
tx <- matrix(data = NA, nrow = niter, ncol = 14)
pb <- txtProgressBar(min = 1, max = 1261, style = 3)
cl <- 1

```

**#If a researcher wants to run simulations at specific parameter values instead of looping
#across all possibilities, then change the below for-loops into assignments (e.g. N <- 4**

```

#q <- 3, etc.)
for (N in 1:5) #Five sample sizes
{
  for (q in 1:9) #Nine skewness values
  {
    for (p in 1:4) #Four effect sizes of X
    {
      for (t in 1:7) #Seven effect sizes of Rho
      {

#Calculate residual variance
resY <- 1 - (YonX[p,1]^2 + OV[t,1]^2 + 2*OV[t,1]^2*YonX[p,1])

#Call the simulation function
tx <- foreach(i = 1:niter) %dopar% sim(i)
tx <- matrix(unlist(tx), ncol = 14, byrow = TRUE)

#Save results in table
table[c1,] <- c(nobs[N,1], skew[q,1], OV[t,1], YonX[p,1], mean(tx[,1]), sd(tx[,1]),
mean(tx[,9]), sqrt(mean(tx[,3])), mean(tx[,11]), mean(tx[,13]), mean(tx[,2]), sd(tx[,2]),
mean(tx[,10]), mean(tx[,12]), mean(tx[,14]), mean(tx[,4]), sd(tx[,4]), sqrt(mean(tx[,5])),
mean(tx[,6]), mean(tx[,7]), sqrt(mean(tx[,8])))

c1 <- c1 + 1
setTxtProgressBar(pb, c1)
}
}
}
}

#Name columns in table
colnames(table) <- c("N", "Skewness", "OV", "X", "Mean X", "sd X", "se X", "RMSE X",
"Cov X", "Pow X", "Mean C", "sd C", "se C", "Cov C", "Pow C", "M X OLS", "sd X OLS",
"RMSE OLS", "Copula > OLS", "True M", "RMSE true")

head(table)

#Write to file (working directory needs to be set to where the file should be saved)
write.csv(table, file=paste("Simulation_Results.csv", sep=","))

```